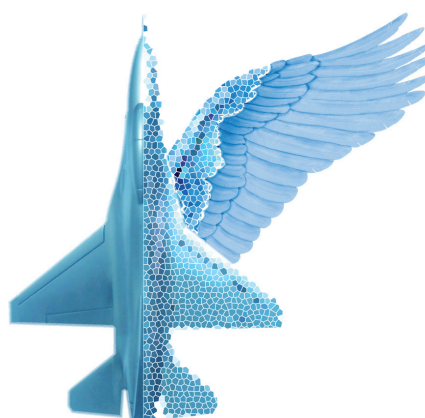


**ОНТОЛОГИЯ**  
**ПРОЕКТИРОВАНИЯ**

Научный журнал

Том 8

№ 1(27)



## EDITORIAL BOARD – РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Anatoly I. Belousov	Белоусов Анатолий Иванович, д.т.н., профессор, Самарский университет, Самара
Nikolay M. Borgest	Боргест Николай Михайлович, к.т.н., профессор, Самарский университет, член ИАОА, Самара
Stanislav N. Vasiliev	Васильев Станислав Николаевич, д.т.н., профессор, академик РАН, ИПУ РАН, Москва
Vladimir G. Gainutdinov	Гайнутдинов Владимир Григорьевич, д.т.н., профессор, КНИТУ-КАИ, Казань
Vladimir V. Golentkov	Голенков Владимир Васильевич, д.т.н., профессор, БГУИР, Минск
Vladimir I. Gorodetsky	Городецкий Владимир Иванович, д.т.н., профессор, СПИИРАН, Санкт-Петербург
Yury A. Zagorulko	Загорюлько Юрий Алексеевич, к.т.н., с.н.с., ИСИ СО РАН, Новосибирск
Anton V. Ivashchenko	Иващенко Антон Владимирович, д.т.н., профессор, Самарский университет, Самара
Alexander S. Kleshchev	Клещёв Александр Сергеевич, д.ф.-м.н., профессор, ИАПУ ДВО РАН, Владивосток
Valery A. Komarov	Комаров Валерий Андреевич, д.т.н., профессор, Самарский университет, Самара
Sergey M. Krylov	Крылов Сергей Михайлович, д.т.н., профессор, СамГТУ, Самара
Victor M. Kureichik	Курейчик Виктор Михайлович, д.т.н., профессор, Технологический институт ЮФУ, Таганрог
Dmitry V. Lande	Ландэ Дмитрий Владимирович, д.т.н., с.н.с., ИПРИ НАН Украины, Киев
Lyudmila V. Massel	Массель Людмила Васильевна, д.т.н., профессор, ИСЭМ СО РАН, Иркутск
Aleksandr Yu. Nesterov	Нестеров Александр Юрьевич, д.филос.н., профессор, Самарский университет, Самара
Dmitry A. Novikov	Новиков Дмитрий Александрович, д.т.н., проф., член-корреспондент РАН, ИПУ РАН, Москва
Alexander V. Palagin	Палагин Александр Васильевич, д.т.н., проф., академик НАН Украины, Ин-т кибернетики, Киев
Semyon A. Piyavsky	Пиявский Семён Авраамович, д.т.н., профессор, СамГТУ, Самара
Yury M. Reznik	Резник Юрий Михайлович, д.филос.н., профессор, Институт философии РАН, Москва
George Rzevski	Ржевский Георгий, профессор, Открытый университет, Лондон
Peter O. Skobelev	Скобелев Петр Олегович, д.т.н., НПК «Разумные решения», Самара
Sergey V. Smirnov	Смирнов Сергей Викторович, д.т.н., ИПУСС РАН, Самара
Peter I. Sosnin	Соснин Петр Иванович, д.т.н., профессор, УлГТУ, Ульяновск
Dzhavdet S. Suleymanov	Сулейманов Джавдет Шевкетович, д.т.н., профессор, академик, вице-президент АН РТ, Казань
Boris E. Fedunov	Федунов Борис Евгеньевич, д.т.н., профессор, ГосНИИ авиационных систем, Москва
Altynbek Sharipbay	Шарипбай Алтынбек, д.т.н., профессор, Институт искусственного интеллекта, Астана
Boris Ya. Shvedin	Шведин Борис Яковлевич, к.психол.н., ООО «Дан Роуз», член ИАОА, Ростов-на-Дону

## Executive Editorial Board - Исполнительная редакция

Chief Editor Skobelev P.O.	Главный редактор	Скобелев П.О.	в.н.с. ИПУСС РАН
Chief Editor Smirnov S.V.	Зам. главного редактора	Смирнов С.В.	зам. директора ИПУСС РАН
Executive Editor Borgest N.M.	Выпускающий редактор	Боргест Н.М.	директор изд-ва «Новая техника»
Editor Kozlov D.M.	Редактор	Козлов Д.М.	доцент Самарского университета
Technical Editor Simonova A.U.	Технический редактор	Симонова А.Ю.	редактор изд-ва «Новая техника»
Translation Editor Korovin M.D.	Редактор перевода	Коровин М.Д.	инженер Самарского университета

## CONTACTS FOUNDERS – КОНТАКТЫ УЧРЕДИТЕЛЕЙ

### ИПУСС РАН

443020, Самара, ул. Садовая, 61  
тел./факс.: +7 (846) 333 27 70

Смирнов С.В.  
smirnov@iccs.ru

### Самарский университет

443086, Самара, Московское шоссе 34, корп. 10, кафедра КиПЛА  
тел.: +7 (846) 267 46 47, факс.: +7 (846) 267 46 46

Боргест Н.М.  
borgest@yandex.ru

### ООО «Новая техника» (издательство)

Адрес редакции: 443010, Самара, ул. Фрунзе, 145, тел.: +7 (846) 332 67 84, факс.: +7 (846) 332 67 81

The journal has entered into an electronic licensing relationship with EBSCO Publishing, the world's leading aggregator of full text journals, magazines and eBooks. The full text of JOURNAL can be found in the EBSCOhost™ databases.



Журнал размещен в коллекции «Издания по естественным наукам» на платформе EastView.

The journal has been successfully evaluated in the evaluation procedure for the ICI Journals Master List 2014, 2015 and journal received the ICV (Index Copernicus Value) of 67.46 points (2014), 67.64 (2015), 77.98 (2016).

Журнал включён в перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание учёной степени кандидата наук, на соискание учёной степени доктора наук (Письмо Департамента аттестации научных и научно-педагогических работников Минобрнауки РФ от 01.12.2015 № 13-6518) по группам научных специальностей 05.13.00 и 05.07.00.

Журнал включён в Российский индекс научного цитирования. Пятилетний ИФ РИНЦ 0.839 (2013), 0.754 (2014), 1.268 (2015), 0.945 (2016).

Журнал зарегистрирован в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство ПИ № ФС 77-70157 от 16.06.2017 г. (ранее выданное свидетельство ПИ № ФС 77-46447 от 07.09.2011 г.)

[http://agora.guru.ru/scientific\\_journal/](http://agora.guru.ru/scientific_journal/)



Отпечатано в ООО «Новая техника», г. Самара, пр.К.Маркса, 24-76.  
Дата выхода 30.03.2018. Тираж 300 экз. Свободная цена. (6+).

© Все права принадлежат авторам публикуемых статей  
© ООО «Новая техника» - «New Engineering» Ltd., 2011-2018  
© Самарский университет - Samara University, 2015-2018  
© ИПУСС РАН - ICCS RAS, 2015-2018

## СОДЕРЖАНИЕ

<b>ОТ РЕДАКЦИИ</b>	
ВПЕРЁД, В БУДУЩЕЕ!	5-7
<b>ОБЩИЕ ВОПРОСЫ ФОРМАЛИЗАЦИИ ПРОЕКТИРОВАНИЯ: ОНТОЛОГИЧЕСКИЕ АСПЕКТЫ</b>	
<b>Ю.Р. Валькман, В.Б. Тарасов</b>	8-34
ОТ ОНТОЛОГИЙ ПРОЕКТИРОВАНИЯ К КОГНИТИВНОЙ СЕМИОТИКЕ	
<b>С.В. Микони</b>	35-48
ФОРМАЛИЗАЦИЯ ПОЗНАВАТЕЛЬНОГО ПРОЦЕССА НА ОСНОВЕ БАЗИСА МОДЕЛЕЙ	
<b>Ю.В. Сложеникина, А.В. Растягаев, И.Ю. Кухно</b>	49-57
АВТОРСКИЙ ТЕРМИН: К ОПРЕДЕЛЕНИЮ ПОНЯТИЯ	
<b>ПРИКЛАДНЫЕ ОНТОЛОГИИ ПРОЕКТИРОВАНИЯ</b>	
<b>В.В. Грибова, М.В. Петряева, Д.Б. Окунь, Е.А. Шалфеева</b>	58-73
ОНТОЛОГИЯ МЕДИЦИНСКОЙ ДИАГНОСТИКИ ДЛЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ	
<b>Н.В. Лукашевич, Б.В. Добров, А.М. Павлов, С.В. Штернов</b>	74-95
ОНТОЛОГИЧЕСКИЕ РЕСУРСЫ И ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА В ПРЕДМЕТНОЙ ОБЛАСТИ «БЕЗОПАСНОСТЬ»	
<b>М.А. Навроцкий, Н.А. Жукова, Д.И. Муромцев</b>	96-109
ОНТОЛОГИЯ ПРОЕКТИРОВАНИЯ, ПРИМЕНЕНИЯ И СОПРОВОЖДЕНИЯ ПОРТАЛОВ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ	
<b>Р.В. Гирин, С.П. Орлов</b>	110-123
ОБЪЕКТНО-ОРИЕНТИРОВАННАЯ ДЕКОМПОЗИЦИЯ ПРОГРАММНОЙ ЛОГИКИ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ	
<b>О.А. Егорова</b>	124-133
ОНТОЛОГИЯ ТЕРМИНА «ПРИНЦИП СОВМЕСТИМОСТИ» В РЕСТАВРАЦИОННОМ ПРОЕКТИРОВАНИИ	
<b>ИНЖИНИРИНГ ОНТОЛОГИЙ</b>	
<b>Е.А. Сидорова</b>	134-151
ПОДХОД К МОДЕЛИРОВАНИЮ ПРОЦЕССА ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ ТЕКСТА НА ОСНОВЕ ОНТОЛОГИИ	
<b>МЕТОДЫ И ТЕХНОЛОГИИ ПРИНЯТИЯ РЕШЕНИЙ</b>	
<b>В.Е. Гвоздев, Л.Р. Черняховская, Д.В. Блинова</b>	152-166
ЭВЕРГЕТИКА КАК МЕТОДОЛОГИЧЕСКАЯ ОСНОВА УПРАВЛЕНИЯ ВЫЯВЛЕНИЕМ ДЕФЕКТОВ НА ПРЕДПРОЕКТНОЙ СТАДИИ ЖИЗНЕННОГО ЦИКЛА СИСТЕМ ОБРАБОТКИ ДАННЫХ	
<b>НАУЧНЫЕ КОНФЕРЕНЦИИ</b>	167-168

## CONTENT

<b>FROM THE EDITORS</b> «COME ON!»	5-7
<b>GENERAL ISSUES OF FORMALIZATION IN THE DESIGNING: ONTOLOGICAL ASPECTS</b>	
<b>Yu.R. Valkman, V.B. Tarassov</b> FROM ONTOLOGIES OF DESIGNING TO COGNITIVE SEMIOTICS	8-34
<b>S.V. Mikoni</b> FORMALIZATION OF THE COGNITIVE PROCESS USING THE BASIS OF MODELS	35-48
<b>Yu.V. Slozhenikina, A.V. Rastyagaev, I.Yu. Kuhno</b> AUTHOR'S TERM: TO DEFINITION OF THE NOTION	49-57
<b>APPLIED ONTOLOGY OF DESIGNING</b>	
<b>V.V. Gribova, M.V. Petryaeva, D.B. Okun, E.A. Shalfeeva</b> MEDICAL DIAGNOSIS ONTOLOGY FOR INTELLIGENT DECISION SUPPORT SYSTEMS	58-73
<b>N.V. Loukachevitch, B.V. Dobrov, A.M. Pavlov, S.V. Shternov</b> ONTOLOGICAL RESOURCES AND INFORMATION-ANALYTICAL SYSTEM IN SECURITY DOMAIN	74-95
<b>M.A. Navrotskiy, N.A. Zhukova, D.I. Mouromtsev</b> ONTOLOGY FOR DESIGN, APPLICATION AND SUPPORT OF SCIENTIFIC AND ENGINEERING DATA PORTALS	96-109
<b>R.V. Girin, S.P. Orlov</b> OBJECT-ORIENTED DECOMPOSITION OF ARTIFICIAL NEURAL NETWORKS LOGIC	110-123
<b>O.A. Egorova</b> ONTOLOGY OF THE TERM «PRINCIPLE OF COMPATIBILITY» IN RESTORATION DESIGNING	124-133
<b>ONTOLOGY ENGINEERING</b>	
<b>E.A. Sidorova</b> ONTOLOGY-BASED APPROACH TO MODELING THE PROCESS OF EXTRACTING INFORMATION FROM TEXT	134-151
<b>METHODS AND TECHNOLOGIES OF DECISION MAKING</b>	
<b>V.E. Gvozdev, L.R. Chernyakhovskaya, D.V. Blinova</b> EVERGETICS AS A METHODOLOGICAL BASIS FOR MANAGING OF DEFECTS IDENTIFICATION AT THE PRE-DESIGN STAGE OF DATA PROCESSING SYSTEMS	152-166
<b>SCIENTIFIC CONFERENCES</b>	167-168



## «COME ON!» ВПЕРЕД, В БУДУЩЕЕ!

## ОТ РЕДАКЦИИ

*Две вещи поражают моё воображение: звёздное небо над головой и нравственный закон внутри нас.*

И. Кант

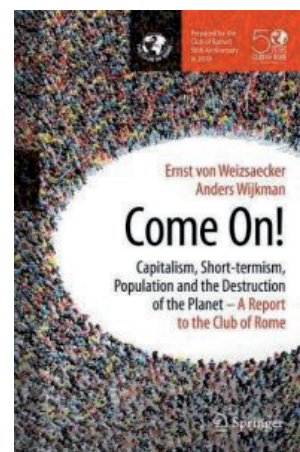
*Стоит помнить, что, чем глубже мы погружаемся в океан знаний о мире, тем опаснее становится это путешествие и тем **больше ответственность** за «звёздное небо над головой и нравственный закон внутри нас».*

Т.В. Черниговская

### Дорогой наш читатель, уважаемые авторы и члены редакционной коллегии!

Ведущие эксперты Римского Клуба пришли к выводу о неизбежности смены парадигмы развития нашей цивилизации. Жёсткая критика капитализма, отказ от «вульгарного» материализма и упрощённого понимания мира, призыв к «новому Просвещению», духовно-нравственному мировоззрению, единой гармоничной цивилизации – такова повестка будущего развития, предлагаемая Римским Клубом.

В конце прошлого года Римский Клуб представил новый доклад «*Come On! Капитализм, близорукость, население и разрушение планеты*», приуроченный к своему полувековому юбилею. Это один из важнейших документов нашего времени, т.к. именно Римский Клуб является «ориентиром» для мировой элиты. «Come On!» - второй за пятьдесят лет доклад, выражающий *консолидированную позицию* Клуба. Доклад написан президентами Клуба Эрнстом Вайцзеккером и Андерсом Вийкманом, при участии тридцати четырёх других членов<sup>1</sup>. Авторы разбирают истоки и патологии современного мировоззрения, описывают альтернативную философию «нового Просвещения».



Клуб видит задачу **образования** в формировании у молодёжи «грамотности в отношении будущего» (*futures literacy*). Образование, способное делать это, должно:

- основываться на *связанности* «вызывать интерес, активно задействовать способности каждого студента учиться для самого себя и помогать учиться другим»;
- носить ценностный характер, опираться на уважение к культурным различиям, где *ценности* воплощаются в акценте на благополучии всех живых существ и мира в целом;
- фокусироваться на знаниях, касающихся экологии, взаимосвязанности систем и *устойчивости* развития;
- культивировать *интегральное* мышление, которое способно «воспринимать, организовывать, согласовывать и воссоединять отдельные фрагменты и достигать подлинного понимания основополагающей реальности» и отличается от системного мышления, также как интеграция отличается от агрегации<sup>2</sup>;
- исходить из *плюрализма* содержания, т.к. сегодняшние студенты нуждаются в инклюзивном образовании, в котором одни формы знания дополняли бы другие.

<sup>1</sup> von Weizsäcker, E., Wijkman, A. Come On! Capitalism, Short-termism, Population and the Destruction of the Planet. — Springer, 2018. — 220 p. - <https://www.clubofrome.org/2017/10/25/new-report-to-the-club-of-rome-come-on/>.

<sup>2</sup> И здесь редакции журнала видится, что именно **онтологическое моделирование** во многом отвечает на этот вызов!

Доклад обращает внимание на наличие у цифровой экономики тёмной стороны. Так, например, сервисы, наподобие *Uber*, и их пользователи не разделяют общих расходов и в нынешнем виде не соответствуют критериям устойчивости. Есть реальная опасность неконтролируемого развития и неэтичного использования технологий.

Особое неприятие авторов доклада вызывает макроэкономический *показатель* - валовой внутренний продукт (ВВП), который отражает лишь скорость движения денег в экономике, а не уровень благополучия общества. Парадоксальные случаи, когда разлив нефти, загрязняющий среду обитания, увеличивает ВВП из-за связанных с ним расходов на ликвидацию аварии, также как болезни, бедствия и несчастные случаи уменьшают благополучие при росте ВВП. Ещё Платон в своей модели общественного блага ставил богатство фактически на последнее место<sup>3</sup>.

Свою лекцию на Гайдаровском форуме в январе 2018 года «Искусственный интеллект – вызов для человечества» профессор Черниговская Т.В.<sup>4</sup> начала с анализа именно этого доклада Римского клуба. Специалист в области естественного интеллекта выразила опасение в том, «что может вытворить технологическое и бездуховное общество, в котором мы теперь живём на этой планете». В образовании, по мнению одного из ведущих популяризаторов науки в России, «преподаватель, который приходит чтобы излагать знания, не нужен!.. Нужно: заразить, заинтересовать студента... должны быть люди, которые объясняют, комментируя - создают «провокацию», вызывают протест» и стремление к познанию.



В январе-марте 2018 года наш журнал принял участие в ряде международных мероприятий, среди которых:

- Гайдаровский форум (Москва, Россия)<sup>5</sup>;
- международная конференция «SCIENCE ONLINE XXI: электронные информационные ресурсы для науки и образования» (Бад-Гаштайн, Австрия)<sup>6</sup>;
- международная конференция Open Semantic Technologies for Intelligent Systems (OSTIS-2018, Минск, Белоруссия)<sup>7</sup>;
- международная научная конференция «Daemons in the Machine. Предвосхищая искусственный интеллект» (Москва, Россия)<sup>8</sup>;
- всероссийская научная конференция с международным участием памяти Станислава Лема «Четвертые Лемовские чтения» (Самара, Россия)<sup>9</sup>.

На конференции *OSTIS-2018* в Минске уже восьмой год обсуждаются вопросы семантического представления знаний и их унификация, исследуются средства и методы, ориентиро-

<sup>3</sup> «Первое из божественных благ - это разумение; второе - сопутствующее разуму здоровое состояние души; из их смешения с мужеством возникает третье благо - справедливость; четвертое благо - мужество. Меньшие блага - это те, во главе которых стоит здоровье, затем идёт красота, на третьем месте - сила... на четвертом - богатство...» - Платон. Законы. М., 1998.

<sup>4</sup> <http://gaidarforum.ru/program/media/otkrytyy-dialog-iskusstvennyy-intellekt-vyzov-dlya-chelovechestva/>.

<sup>5</sup> <http://ssau.ru/news/15014/>.

<sup>6</sup> <https://elibrary.ru/projects/conference/austria2018/program.asp>.

<sup>7</sup> <http://conf.ostis.net/index.php?title=OSTIS-2018>

<sup>8</sup> [http://newlaboratoria.ru/www/rus/scientific\\_advice\\_2\\_61/](http://newlaboratoria.ru/www/rus/scientific_advice_2_61/).

<sup>9</sup> <http://ssau.ru/events/show/796-IV-Lemovskie-cheniya>.

ванные на проектирование различных компонентов интеллектуальных систем (ИС), а также сами прикладные ИС, основанные на семантическом представлении используемых ими знаний, и уже давно не стоит вопрос о том, стоит ли заниматься разработкой ИС и их использованием. Но для специалистов творческих слабоформализованных профессий (например, художников, дизайнеров и др.) вопрос о внедрении систем искусственного интеллекта (ИИ) в их деятельность только начинает пробивать дорогу.

17 марта 2018 года в Британской высшей школе дизайна на конференции «*Daemons in the Machine. Предвосхищая искусственный интеллект*», посвящённой изучению и применению передовых технологий машинного обучения и ИИ в современном искусстве, нейробиологами и робототехниками, художниками и культурологами, искусствоведами и медиа-философами обсуждались такие важные для науки и жизни вопросы как:

- Где граница между искусством и научным исследованием?
- Существует ли искусство за пределами человеческой культурной практики?
- Что нам остаётся и чего не сможет ИИ?
- Как выглядит будущее музея, художника, технологической компании в присутствии ИИ?

Интересные события для читателей нашего журнала происходили в марте в Самарском университете. Так, с 22 по 24 марта при содействии Научного совета по методологии искусственного интеллекта РАН, Самарского научного центра РАН, Стратегического общественного движения «Россия 2045» и Ассоциации исследователей фантастики прошла научная конференция «*IV Лемовские чтения*», посвящённая научной фантастике. Среди тем для обсуждения были: теоретические основания научной фантастики; модели будущего в научной фантастике; научная фантастика и научно-технологический прогресс и др.<sup>10</sup> Организатор Лемовских чтений, профессор *Александр Нестеров* подчеркнул факт наступления эры технологической сингулярности, когда созданное человеком техническое окружение, среда, начинает развиваться быстрее, нежели сам человек. Научная фантастика помогает предвидеть вектор развития цивилизации, и появление массовой телефонии, Интернета, робототехники и искусственного интеллекта - тому подтверждение.

22 марта в Самарском университете состоялась открытая лекция на тему: «Создание, интеграция и эволюция нового дизайна в автомобилестроении», на которой директор по дизайну LADA *Стив Маттин* рассказал о методах разработки принципиально новых форм визуальных «ДНК автомобилей», о том, как воплотить новые формы в серийных моделях<sup>11</sup>.



В планах редакции журнала на этот год посетить десять подобных мероприятий, часть из которых представлена в разделе «Научные конференции» этого номера.

*Уважаемый автор!*

Мы по-прежнему ждём *новых результатов* в области формализации знаний, практического использования семантических моделей для создания лучшего будущего!  
*Ontologists and designers of all countries and subject areas, join us!*

<sup>10</sup> <http://ssau.ru/events/show/796-IV-Lemovskie-chteniya>

<sup>11</sup> [https://vk.com/samara\\_university?z=video-69812\\_456239098%2F14f78f9df734ae43c5%2Fpl\\_wall\\_-69812](https://vk.com/samara_university?z=video-69812_456239098%2F14f78f9df734ae43c5%2Fpl_wall_-69812)

УДК 001.892: 004.8: 681.5.015

## ОТ ОНТОЛОГИЙ ПРОЕКТИРОВАНИЯ К КОГНИТИВНОЙ СЕМИОТИКЕ

Ю.Р. Валькман<sup>1</sup>, В.Б. Тарасов<sup>2</sup>

<sup>1</sup>Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, Киев, Украина

<sup>2</sup>Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия  
Vbulbov@yahoo.com

### Аннотация

Изложены основы концепции жизненных циклов сложных технических систем. Отмечены тенденции расширения и детализации жизненных циклов, приводящие к появлению нетрадиционных стратегий проектирования и даже новых проектных этапов на стыке исследования и проектирования, проектирования и испытаний. В русле реализации онтологического подхода к проектированию рассмотрены этапы, уровни, аспекты, задачи проектирования. Особое внимание уделено исследовательскому проектированию как этапу «аналитического синтеза» сложных технических систем. Показана центральная роль НЕ-факторов в исследовательском проектировании, введено трёхмерное пространство моделирования НЕ-факторов и предложены варианты их классификации. На начальных этапах проектирования важное место занимает модельный подход, связанный с оперированием моделями и параметрами. В статье развиваются оригинальные методики представления проектных знаний в модельно-параметрическом пространстве и построения на основе биполярных шкал многомерного семантико-прагматического пространства методов и языков проектирования. На базе семантической шкалы языков моделирования сформулированы принципы соответствия и взаимодополняемости. С одной стороны, в контексте поддержки образного мышления проектанта рассмотрены методы иллюстративной и когнитивной графики. С другой стороны, проанализированы два известных варианта общей теории проектирования – формальной теории, стремящейся описать проектирование в терминах множеств, топологических пространств, исчислений, в частности, исчисления предикатов высокого порядка. Проведён анализ недостатков формальных систем и классической теории рассуждений с точки зрения принятия реальных проектных решений. В результате сделан вывод о необходимости привлечения в науку проектирования семиотических моделей. В заключительной части работы приведены основные определения семиотики и варианты классификации знаков, обсуждаются особенности перехода от семиотического моделирования к прикладной семиотике и когнитивной семиотике. Одним из основателей когнитивной семиотики как «науки-перекрёстка», предметом которой является как получение знаний из знаковых структур, так и синтез знаковых конструкций в русле инженерии знаний, был профессор Ю.Р. Валькман.

**Ключевые слова:** онтология проектирования, сложная техническая система, жизненный цикл, исследовательское проектирование, НЕ-факторы, модельно-параметрическое пространство, общая теория проектирования, семиотика, когнитивная семиотика.

**Цитирование:** Валькман, Ю.Р. От онтологий проектирования к когнитивной семиотике / Ю.Р. Валькман, В.Б. Тарасов // Онтология проектирования. – 2018. – Т. 8, №1(27). - С.8-34. DOI:10.18287/2223-9537-2018-8-1-8-34.

«Знание может быть лишь у того, у кого есть вопросы»

Х. Гадамер

### Введение

Название и замысел этой статьи выражают стремление показать творческий путь в науке известного советского и украинского учёного в области информатики и автоматизированно-



го проектирования, одного из ведущих специалистов Украины в сфере интеллектуальных систем и технологий, заведующего Отделом распределенных интеллектуальных систем Международного научно-учебного центра информационных технологий и систем НАН и МОН Украины, профессора Киевского политехнического института имени Игоря Сикорского, члена редколлегии научного журнала «Онтология проектирования», д.т.н. Юрия Роландовича Валькмана.

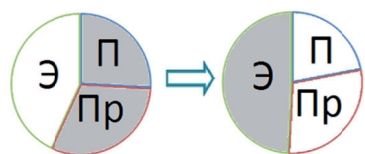
## 1 От моделирования жизненных циклов к исследовательскому проектированию

Любой цикл характеризуется *завершённостью* и *повторяемостью* взаимосвязанных этапов на определённом промежутке времени; его ключевыми характеристиками являются *временные показатели*, а именно, длительности – как отдельных этапов, так и цикла в целом. Понятие жизненного цикла (ЖЦ) системы определяется интервалом времени от начала её создания до конца эксплуатации; при этом за начало ЖЦ принято считать формирование потребности в системе (зарождение идеи системы), а за конец – снятие её с эксплуатации (утилизацию). Онтологическое моделирование ЖЦ может начинаться с *мереологии* – отношения ЖЦ как целого к его частям. Наиболее крупными частями ЖЦ являются его стадии, например, разработки, производства, эксплуатации. Разные стадии состоят из различных этапов.

Концепция ЖЦ есть основной вариант реализации системного подхода к сложным техническим объектам, таким как самолёты и корабли, направленный на отображение изменений состояния этих объектов в период их существования. Она связана с интеграцией процессов проектирования, производства и эксплуатации сложных технических систем (СТС) в рамках единой метамодели и предполагает выделение ряда стадий (этапов) ЖЦ и изучение взаимосвязей между ними. Общая идея оптимизации ЖЦ заключается в минимизации временных затрат на стадиях проектирования и производства СТС и максимизации сроков её эксплуатации (рисунок 1). ЖЦ СТС представляет собой надсистему для определения общей методологии и структуры исследований на конкретных стадиях существования и развития СТС, например, на этапе её эскизного проектирования, этапе технологической подготовки производства, этапе технического обслуживания и т.п. Онтология всего ЖЦ является онтологией верхнего уровня по отношению к онтологиям проектирования.

В связи с непрерывным совершенствованием СТС, развитием средств, методов, технологий их создания и эксплуатации, происходят, с одной стороны, всё большая *детализация* стадий и этапов ЖЦ, в частности, стадии проектирования, а с другой стороны, *расширение* ЖЦ, учёт всё больших объёмов данных, информации, знаний о проектируемых объектах, включая знания из сферы производства и эксплуатации СТС, на начальных этапах проектирования. Первый фактор эволюции ЖЦ привёл к выделению в процессе разработки СТС операций *исследовательского проектирования* (ИП) и формированию на их основе отдельного этапа ЖЦ [1, 2]. Речь идёт о самом начальном этапе проектирования, когда сформировалась потребность в новой СТС, необходимо обосновать целесообразность и эффективность её создания, а также разработать техническое задание (ТЗ). Второй фактор эволюции ЖЦ привёл к возникновению таких концепций как *совмещённая разработка* (Concurrent Engineering) [3] и *«проектирование для X»* (Design for X) [4]. Под совмещённой разработкой [1, 3, 5] обычно понимается пересечение и частичное запараллеливание отдельных этапов ЖЦ в интересах улучшения информационного взаимодействия различных подразделений предприятия и сокращения времени разработки (рисунок 2). Ещё одна трактовка совмещённой разработки – одновременное проектирование СТС и процесса её изготовления. Достаточно близкой (но, в некотором смысле, обратной) стратегией является «проектирование для

Х», предполагающее предварительную проработку на стадии проектирования важнейших проблем производства и эксплуатации СТС. Здесь под Х понимается любой этап ЖЦ после стадии проектирования. Так, существуют стратегии «проектирование для производства» (Design for Manufacturing), «проектирование для сборки» (Design for Assembly), «проектирование для технического обслуживания» (Design for Maintenance), и т.д.



П – проектирование; Пр – производство;  
Э – эксплуатация.

Рисунок 1 – Круговое представление ЖЦ системы: иллюстрация принципа сокращения сроков проектирования и производства, а также продления периода эксплуатации СТС

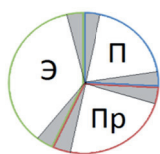


Рисунок 2 – Круговое представление ЖЦ системы на базе покрытия: наличие зон пересечения между всеми стадиями

В СССР разработка проблематики моделирования и программирования ЖЦ была начата в Институте кибернетики им. В.М. Глушкова в 1970-е годы в рамках проекта «Программируемые ЖЦ объектов новой технологии» [6]. В частности, была построена система моделирования, оптимизации и расчёта технико-экономических показателей ЖЦ эксплуатации СТС [7].

Классические модели ЖЦ являются линейными или круговыми. Линейная модель ЖЦ выражает такие свойства времени как течение, направленность, необратимость, тогда как круговая модель характеризует итеративность и ритмичность процессов на протяжении ЖЦ. В 1980-е годы в МГТУ им. Н.Э. Баумана по инициативе Л.А. Кашубы были разработаны спиральные модели ЖЦ СТС [8,9], сочетающие достоинства и компенсирующие недостатки линейных и круговых моделей ЖЦ.

В конце XX-го века понятие ЖЦ СТС ещё более расширилось: в него стали включать стадию рекуперации (Recycling), которая лежит в основе концепции «обратного ЖЦ» [10]. В монографии [1] рассмотрены соотношения «жизненные циклы – технологии – процессы», а в статье [11] введена трёхмерная система ЖЦ «ЖЦ СТС – ЖЦ предприятия – ЖЦ процесса» и предложены варианты согласованного управления ими.

В [1] на основе анализа построенных ЖЦ СТС были сделаны следующие выводы:

- разделение ЖЦ на этапы и стадии зависит от принятой точки зрения;
- стадии ЖЦ могут быть как достаточно общими, так и весьма детальными;
- некоторые стадии или этапы могут декомпозироваться намного подробнее, чем другие;
- моменты начала и конца ЖЦ зависят от контекста рассмотрения;
- можно строить иерархические структуры ЖЦ различных степеней декомпозиции;
- линейная структура и упорядоченность во времени не определяют полностью последовательность этапов ЖЦ;
- существенным аспектом является качество реализации этапов ЖЦ и возможность повторного выполнения стадий и этапов;
- сложные объекты могут иметь множество ЖЦ на разных уровнях декомпозиции со сложными взаимоотношениями.

Проектирование представляет собой начальную и важнейшую стадию ЖЦ СТС, где цена ошибки особенно велика (принятие неверных проектных решений может легко привести к недопустимому увеличению сроков производства или быстрому выводу СТС из эксплуатации). В общем смысле, проектирование служит для достижения полезного результата (создания проектно-конструкторской документации, необходимой для изготовления, материализации СТС) путём замены существующего состояния разрабатываемого объекта другим состоянием. По сути, проектирование – это информационная подготовка некоторого изменения.

Оно соответствует информационному процессу, в котором осуществляется преобразование входной информации о проектируемом объекте, состоянии знаний в рассматриваемой области, предыдущем опыте проектирования, в выходную информацию в виде проектных документов, выполненных в требуемой форме и содержащих проектные решения или результаты проектирования.

Онтологический подход к проектированию предполагает выделение и исследование этапов, уровней, методов, пространства проектирования. Как и весь ЖЦ СТС, процесс проектирования по времени делится на стадии и этапы, а по содержанию – на уровни проектирования. Стадии, этапы, уровни и модели проектирования образуют *пространство* проектирования. Согласно ГОСТ 2103-2013, основными *стадиями разработки* являются предварительное проектирование, начинающееся с изучения и анализа ТЗ и направленное на разработку технического предложения, эскизное проектирование, техническое проектирование, разработка конструкторской документации для изготовления и испытаний опытных образцов изделий. Таким образом, проектирование есть открытая система: с одной стороны, к нему относят ТЗ, формируемое на стыке областей эксплуатации и проектирования на базе научных исследований, а с другой стороны, проектирование оказывается связанным с разработкой документации для создания опытного образца, а также изготовления и эксплуатации пробной серии. По результатам этой эксплуатации может быть принято решение о коррекции ТЗ и перепроектировании.

Важнейшими аспектами проектирования являются функциональный, конструкторский и технологический. Функциональный аспект отражает назначение и функции СТС. Конструкторский аспект характеризует структуру, расположение в пространстве и форму составных частей этой системы. Технологический аспект определяет возможность и способы изготовления СТС, её технологичность и ремонтпригодность.

При использовании блочно-иерархического подхода к проектированию представление о проектируемой системе разбивают по степени «абстракции – детализации» на иерархические уровни [12]. На верхнем уровне формируется наиболее абстрактное представление, отражающее только самые общие черты и особенности проектируемой СТС, а на следующих уровнях степень подробности описания СТС возрастает вплоть до достижения уровня конкретизации, достаточного для её создания. Иными словами, процесс проектирования носит итерационный характер и завершается на таком уровне детализации, когда становится известной вся информация, требуемая при составлении рабочей документации для изготовления СТС.

В этом контексте можно говорить о концептуальном проектировании, осуществляемом проектантами изделия в целом, структурном проектировании (в рамках принятой концепции) на уровне конструкторов его систем, параметрическом проектировании (для определённой структуры) на уровне расчётчиков. Уровни проектирования в основном соответствуют его стадиям: так концептуальное проектирование относится к стадии предварительного проектирования, когда проводится разработка общей концепции проектируемого объекта, составление концептуальных моделей его систем и элементов, формирование технико-экономического обоснования, разработка технического предложения. Здесь термин «концепция» применяется для описания принципа действия и общего облика СТС.

Онтология задач проектирования стала разрабатываться ещё в 1960-е годы, задолго до появления самих терминов «онтология» (в информатике и искусственном интеллекте) и «онтологическое моделирование». На любой стадии проектирования СТС решаются задачи четырёх основных типов [13]:

- 1) поисковое проектирование (или поисковое конструирование по А.И. Половинкину [14]);
- 2) инженерный анализ, в основном связанный с проведением проектных или поверочных расчётов, а также с геометрическим моделированием;

- 3) выполнение проектных чертежей;
- 4) многокритериальная оценка и оптимизация полученных проектных решений.

Поисковое проектирование нацелено на синтез альтернатив, создание проектных решений с элементами новизны. Инженерный анализ – это изучение свойств уже готовых или известных технических решений. При этом не формируются новые объекты, а исследуются характеристики уже существующих конструкций с помощью физических и математических моделей, установления функциональных зависимостей. Классическими примерами графических конструкторских документов являются чертёж общего вида, сборочный чертёж, чертёж детали<sup>1</sup>. Принятие проектных решений включает их информационную подготовку, прогнозирование, выбор и оценку решения. Выбор технических решений происходит в условиях неопределённости, многокритериальности, противоречивости рассматриваемых критериев.

В проектно-конструкторских работах противоположные задачи синтеза и анализа технических решений неотделимы друг от друга: порождение проектной альтернативы является предпосылкой для её анализа, который, в свою очередь, определяет и ограничивает область творческого поиска.

Термин ИП очень близок к понятию «предварительное проектирование». Так, целью ИП является построение облика сложного объекта [1, 15]. Здесь принципиальным моментом является то, что речь идёт об артефактах – искусственных объектах, а также то, что в настоящее время этот артефакт не существует, т.е. «проектируют то, чего нет» [16]. Это кардинальным образом отличает объект проектирования от объекта управления.

Главной особенностью процессов ИП является принятие проектных решений на основе единства синтеза и анализа моделей. Обычно исследование соотносится с анализом, а проектирование – с синтезом. Задачей исследования СТС является получение и накопление знаний о различных свойствах и характеристиках СТС. Задача проектирования заключается в использовании накопленных знаний для создания новых СТС при наличии ряда ограничений. Сам термин ИП подчёркивает тот факт, что, по сути, рассматриваются процессы «аналитического синтеза» СТС.

Любая СТС на уровне ИП должна представлять собой целостную структуру взаимосвязанных моделей. Сегодня технологические аспекты ИП во всё большей степени определяются компьютерным моделированием. По Ю.Р. Валькману, в ИП следует выделять три основные компоненты: область исследований; совокупность знаний об этой области; методологию (совокупность согласованных методов) накопления знаний об этой области и их использования для решения практических задач.

Предметом исследования в процессах ИП является система требований к функциям и морфологии СТС, которую называют ТЗ. Как правило, ТЗ составляют на основе анализа результатов предварительных исследований, расчётов и моделирования. В реальной жизни ТЗ формируется на «стыке» стадий эксплуатации и ИП и служит инструментом коммуникации между заказчиком и исполнителем. Всё это объясняет существование в ИП таких факторов как многозначность, мультиконтекстность, мультимодельность, мультимедийность, много-связность [17].

Характерными чертами процессов ИП являются:

- значительный уровень неопределённости, источниками которой могут быть неполнота или обобщённость исходных данных, внутренняя противоречивость, неоднозначность и нечёткость ТЗ на проектирование, содержащего приближённые оценки характеристик объекта в виде числовых или даже словесно-числовых диапазонов, а также лингвистическое задание целей, ограничений и условий чисто качественного характера;

---

<sup>1</sup> В современном проектировании – это совокупность электронных моделей, содержащих полное описание СТС. *Прим.ред.*

- принципиальные ограничения по точности определения количественных параметров и, особенно, качественных факторов;
- отсутствие аналитических зависимостей, уравнений, связывающих исходные проектные критерии и параметры, и, как следствие, «расплывчатость» представлений об их взаимодействии;
- изменчивость, динамический характер причинно-следственных отношений.

Все эти НЕ-факторы (по А.С. Нариньяни [18]) – неопределённость, неточность, неполнота, нечёткость располагаемой информации – буквально «пронизывают» стадию ИП.

В 2004 г. вышел в свет журнал РАИИ «Новости искусственного интеллекта», содержащий рубрику «Моделирование НЕ-факторов – ключевое направление ИИ в начале XXI-го века». После работы А.С. Нариньяни «НЕ-факторы: краткое введение» [19] в нём была опубликована замечательная статья Ю.Р. Валькмана [20], где были рассмотрены истоки и методы моделирования НЕ-факторов, показана их роль в образном мышлении, предложена классификация НЕ-факторов в системе координат «НЕ-факторы – методы – объекты» (рисунок 3). В заключение сформулирован «основной вопрос» искусственного интеллекта (ИИ) начала XXI-го века: «*Может ли искусственная система называться интеллектуальной, если она не моделирует какие-либо НЕ-факторы?*»

В работе [21] были прослежены основные направления исследования НЕ-факторов, выделены информационные НЕ-факторы и НЕ-факторы развития сложных систем (рисунок 4), предложена методика логико-алгебраического представления НЕ-факторов на базе расширенных логических матриц, идей логического пространства Л. Витгенштейна и семиотического пространства Ю.М. Лотмана.

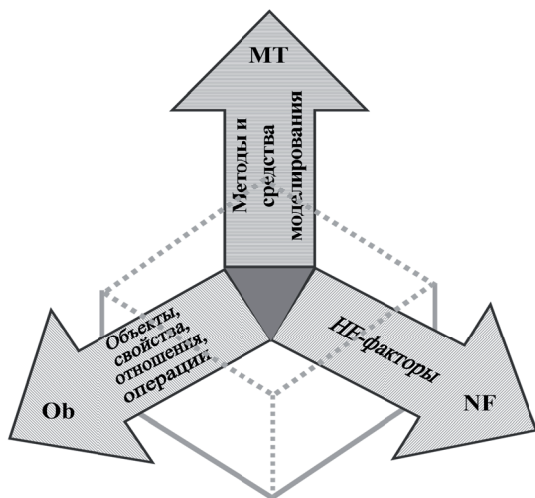


Рисунок 3 – Трёхмерное пространство моделирования НЕ-факторов (по Ю.Р. Валькману)



Рисунок 4 – Два основных класса НЕ-факторов (по В.Б. Тарасову)

## 2 Представление проектных знаний: модельно-параметрическое пространство, когнитивная графика, формальная теория проектирования и семиотические модели

Развитие интеллектуальных технологий в сфере проектирования предполагает представление и обработку проектных знаний. Речь идёт об использовании баз знаний для синтеза проектных решений, применении различных методов рассуждений для пополнения проектных знаний, интеллектуальном анализе проектных данных и обнаружении в них знаний, построении дружественного интерфейса пользователя систем автоматизированного проектирования.

вания и пр. (см., например, [22]). Предварительно приведём классификации видов мышления и соответствующих им знаний.

В психологии разные виды мышления часто характеризуют диадами: образное и вербально-логическое, интуитивное и аналитическое, практическое и теоретическое, репродуктивное и продуктивное мышление и т.д. Следует отметить, что в ИП центральное место занимают процессы образно-интуитивного и продуктивного мышления. Аналогичные диады встречаются и при классификации проектных знаний: эвристические и алгоритмические, неявные и явные, интуитивно-образные (трудно передаваемые) и вербализуемые (легко передаваемые), «поверхностные» (субъективные эвристические правила) и «глубинные» (теории, формальные модели), «горячие» (сформированные и накопленные в ходе собственной деятельности) и «холодные» (полученные извне) знания. Легко понять, что в ИП преобладают неявные, эвристические, «горячие», трудно передаваемые знания проектантов. Поэтому целесообразно разделить методы представления знаний на: 1) традиционные методы представления знаний в ИИ; 2) специализированные (оригинальные) методы представления проектных знаний и их отображения в вычислительных средах.

В проектировании СТС знания специалистов используются для построения различных моделей. *Модель* – это объект (реальный, знаковый или воображаемый), отличный от исходного объекта-оригинала, но способный заменять его в рамках решаемых задач. По А.А. Ляпунову, модель: а) находится в некотором соответствии с объектом-оригиналом; б) способна замещать его в определенных отношениях; в) обеспечивает при её исследовании информацию о самом моделируемом объекте. Модель есть приближённое, неполное отражение проектируемого объекта, абстрагированное от ряда признаков [23]. Структура и законы функционирования модели должны быть проще (или, в каком-то смысле, доступнее для изучения), чем структура и законы функционирования объекта.

Ещё в 1970-е годы в ИИ был выделен ряд классов моделей знаний: сетевые, алгебраические, логические, лингвистические модели. Здесь имеются в виду семантические сети, сценарии, фреймы, реляционные и алгебраические системы, продукционные системы, логика предикатов 1-го порядка, неклассические логики, лингвистические и нечёткие переменные. Затем стали строиться гибридные классы моделей знаний: логико-алгебраические, логико-лингвистические, нейро-логические, фреймово-продукционные, и т.д. Наиболее известными специализированными методами работы с проектными знаниями являются И/ИЛИ-графы, пирамидальные и алгоритмические сети, многоуровневая логика.

Особый интерес для проектировщиков вызывает представление знаний в семантических и прагматических пространствах, а также выделение системы образующих модели мира в виде оппозиционных шкал.

Любой процесс проектирования предполагает построение и использование иерархии моделей. Пусть  $M = \{M_i\}$ ,  $i=1, \dots, n$  есть множество проектных моделей, а  $P = \{P_j\}$ ,  $j=1, \dots, q$  – множество проектных параметров. В работах [24-26] Ю.Р. Валькманом было введено и подробно исследовано модельно-параметрическое пространство ( $\langle M, P \rangle$ -пространство) как средство представления знаний исследователей и проектантов сложных систем. По аналогии с топологическими пространствами определены  $\langle M, P \rangle$ -окрестности его элементов. При построении такой окрестности в центре внимания оказывается либо какая-то модель, либо некоторый параметр. Всё остальное пространство рассматривается относительно данного элемента. Так например, окрестность  $\langle M, P \rangle$ -пространства относительно модели  $M_i$  есть подмножество параметров из  $P$ , непосредственно связанных с моделью  $M_i$ . В свою очередь, окрестностью  $k$ -го порядка называется подмножество всех элементов  $\langle M, P \rangle$ -пространства, связанных с  $M_i$  путями длиной не больше  $k$ . Пример фрагмента системы  $\langle M, P \rangle$ -окрестностей изображён на рисунке 5.

Таким образом,  $\langle M, P \rangle$ -пространства представимы в виде графов с двумя разными типами вершин, в частности, двудольных графов. В  $\langle M, P \rangle$ -пространстве можно определить расстояние, аналогичное семантическому расстоянию между понятиями в психолингвистике. Таким образом, возникает структура на знаниях типа семантического пространства.

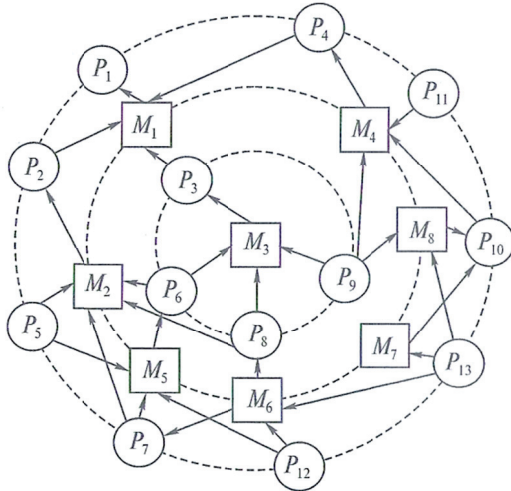


Рисунок 5 – Фрагмент модельно-параметрического пространства: пример  $\langle M, P \rangle$ -окрестностей 1-го, 2-го и 3-го порядков относительно модели  $M_3$ .

Интерпретации зависит тип шкалы. В частности, Д.А. Поспелов ввёл два типа оппозиционных шкал – «серые» и «чёрно-белые» [27]. Для «серых» шкал середина есть точка перехода, понимаемая как *противоречие* «и +, и – одновременно», а для «чёрно-белых шкал» середина – точка разрыва или перескока на другие шкалы, где имеем «ни +, и ни –» (*полная неопределённость*). Обычно биполярные шкалы имеют еще несколько промежуточных точек; в общем случае вместо точек могут рассматриваться нечёткие или лингвистические значения [28].

Различные формы биполярности выделяются по двум критериям: а) сила связи (оппозиции) между ними; б) статус нейтральной точки. Так, в случае сильной оппозиции отрицательная часть шкалы является зеркальным отражением положительной части, и эти две области считаются взаимно исключающими. Ослабление оппозиции означает возможность одновременного сосуществования положительных и отрицательных оценок, а в дальнейшем возникают новые связи между ними (примерами могут служить связи между сильными положительными и слабыми отрицательными модальностями).

В [29] Ю.Р. Валькман предложил систему оппозиционных шкал для описания проектирования: «объективное-субъективное», «традиционное-оригинальное», «коллективное-индивидуальное», «конкретное-абстрактное», «информативное-когнитивное», «логическое-метафорическое». Нетрудно понять, что ИП связано с понятиями, расположенными ближе к правым краям вышеуказанных шкал. Так, в контексте ИП важную роль играет наглядное отображение субъективных представлений проектанта. В отличие от ГОСТ 2.118-2013 на техническое предложение, закрепляющего традиции рассмотрения начального этапа проектирования, модельно-параметрическое пространство является оригинальным, как абстрактным, так и достаточно наглядным способом представления индивидуальных знаний.

В контексте систематизации языков проектирования подробнее остановимся на шкале «логическое-метафорическое». В общем случае язык, имеющий знаковую природу, обеспечивает выражение мыслей и выступает как главный инструмент человеческого общения. Здесь мы будем трактовать термин «язык» как совокупность средств, необходимых для полу-

чения, переработки и передачи информации. Языки проектирования предназначены для представления и преобразования описаний в процессе проектирования.

Следуя В.В. Налимову [30], рассмотрим семантическую шкалу языков моделирования. На одном конце этой шкалы (слева) находятся искусственные языки, в особенности жёсткие математические языки, где каждому знаку приписано чётко определённое значение (логические исчисления, алгебраические системы, алгоритмы), а на другом конце (справа) – мягкие языки, которым свойственна многозначность (или даже неопределённость выражений – по В.В. Мартынову [31]). На этой шкале естественный язык займёт положение ближе к правому краю.

Из данного примера видно, что на основе системы биполярных шкал можно построить многомерное семантико-прагматическое пространство методов и языков проектирования.

Рассмотрение семантической шкалы языков позволяет сформулировать два важных принципа – принцип соответствия и принцип взаимодополняемости. Согласно *принципу соответствия* модели, методы, процедуры, технологии проектирования должны соответствовать сложности проектируемого объекта и характеру информации, располагаемой проектировщиком (проектантом, конструктором, расчётчиком) на данной стадии проектирования. Чем больше уровень неопределённости проектных данных, тем меньше потребная точность их описания, но тем больше эффективность применения научных методов проектирования. Язык описания для каждого этапа проектирования должен отвечать уровню неопределённости на этом этапе. В частности, попытки применения дифференциальных или интегральных уравнений в ИП нередко оказываются малоэффективными как раз из-за того, что эти методы не соответствуют высокому уровню неопределённости. Здесь требуются более *гибкие средства*, связанные с поддержкой образного мышления.

В ИП именно отсутствие жёсткой связи между знаками (исходной текстовой и графической информацией) и значениями требует создания новых графических средств, ориентированных на извлечение неявных, интуитивно-образных знаний и представлений.

Проблему интенсификации научного исследования путём использования возможностей интерактивной компьютерной графики сформулировали и активно изучали Д.А. Поспелов и А.А. Зенкин [32], который применял визуализацию в столь абстрактной математической области, как теория чисел. Важнейшей характеристикой графики является её способность сжимать информацию и непосредственно воздействовать на интуитивно-образное мышление человека. Когнитивная графика связана с визуализацией внутреннего содержания, смысла научных абстракций. Примерами когнитивных компьютерно-графических образов математических абстракций служат: распределение простых чисел по спирали Улама, инвариантные множества обобщённой проблемы Варинга, графические образы чисел  $\pi$ ,  $e$ .

А.А. Зенкин выделял иллюстративную и когнитивную функцию графических изображений. Иллюстративная функция графики визуализирует уже известное знание и обеспечивает узнаваемость изображаемого объекта, а когнитивная функция графики, направленная на визуализацию внутреннего содержания, смысла научных абстракций, способствует порождению нового знания. К иллюстративной графике относятся такие средства проектирования как чертежи, рисунки, гистограммы, фотографии, видеофильмы о создании и испытаниях СТС и пр. Между иллюстративной и когнитивной графикой нет чёткой границы: концептуальное сжатие и наглядное представление известного знания может подсказать новую идею или гипотезу, для подтверждения которой могут также применяться картинки или диаграммы. В системах автоматизации научных исследований и ИП когнитивная компьютерная графика выполняет такие функции как: повышение мотивации и преодоление психологических барьеров на пути к генерации нового знания; стимуляция продуктивного (образного) мышления; поддержка мерелогических переходов в диаде «конкретизация-обобщение» (т.е. поддержка



вариантов грануляции информации проектантом). Так, в русле поддержки образного мышления проектанта возникают следующие переходы: «образ – задача» (визуализация образа для переформулирования задачи, представления её на другом языке с целью лучшего понимания); «образ – решение» (графический образ наталкивает на решение проектной задачи); «решение – образ» (графическое отображение решения проектной задачи – основная ситуация проектирования).

По мнению Ю.Р. Валькмана [33], предметом когнитивной графики являются графические метафоры, в особенности, методы и средства их создания и интерпретации. В привлечении графических метафор и заключается отличие когнитивной графики от других парадигм визуализации.

Принцип Н. Бора гласит: «Противоположности – не противоречия; они - дополнения». По *принципу взаимодополняемости*, решение проектных задач предполагает применение иерархии дополняющих друг друга моделей, а также языков описания различной жёсткости. Соответственно, мягкие модели когнитивной графики следует дополнить строгими логическими и алгебраическими формализмами. ИП предполагает сочетание естественного и искусственных, геометрического и логико-алгебраического языков.

Для ИП, как и для других этапов проектирования, несомненный интерес представляют *исчисления*, т.е. системы, состоящие из некоторого количества исходных объектов и набора правил построения новых объектов из исходных или уже построенных объектов. Теория исчислений является одним из основных разделов математической логики. Именно логические исчисления стали первым примером полностью формализованных дедуктивных систем.

Канонические исчисления опираются на понятие формальной системы, полностью абстрагирующейся от смысла слов используемого языка. *Формальная система* задаётся четвёркой  $FS = \langle T, R, A, I \rangle$ , где  $T$  – множество базовых термов (алфавит системы),  $R$  – множество синтаксических правил,  $A$  – множество аксиом,  $I$  – множество правил вывода.

Под исчислением понимается формальная теория, полученная из аксиом с помощью правил вывода. В случае *полуформальной аксиоматической теории*, например геометрии Эвклида, задаются только аксиомы, а правила вывода считаются общеизвестными. Другой частный случай логического исчисления – это теория естественного вывода, когда множество аксиом пусто, а задаются только правила вывода.

Для систем автоматизированного проектирования чрезвычайно важной является автоматизация проектных рассуждений, связанная с поиском вывода в исчислениях. Классическая система рассуждений состоит из следующих главных компонентов: 1) формальный язык представления знаний; 2) семантика, определяющая его смысл и значения истинности; 3) множество правил вывода для порождения новых знаний; 4) память, обеспечивающая хранение знаний; 5) механизм управления, выбирающий на каждом шаге вывода необходимые посылки и правила. Здесь первые три компонента относятся собственно к логике (логической части системы рассуждений), а последние два представляют собой реализацию логики (управляющую часть системы рассуждений). До сих пор наиболее популярной теорией, определяющей логику рассуждений, остаётся исчисление предикатов, а для управляющей части – теория вычислимости и вычислительной сложности.

Главным достоинством дедуктивного рассуждения (вывода от общего к частному) как перехода от посылок к заключению, опирающегося на логический закон, является гарантированное сохранение истинности. Заключение, полученное дедуктивным путём, не нуждается в проверке и подтверждении. Однако в ИП центральное место занимают недедуктивные рассуждения – индукция (от частного к общему), абдукция, выводы по аналогии или выводы на основе прецедентов. Все эти виды рассуждений характеризуются неопределённостью и не гарантируют истинности заключения при истинности посылок.

В [34] типология рассуждений была распространена на методы проектирования, которые также разделялись на дедуктивные и индуктивные. Там же были сформулированы следующие аксиомы проектирования: а) аксиома неразрешимости; б) аксиома неопределённости; в) аксиома противоречивости; г) аксиома многокритериальности и невозможности «сквозного» правила предпочтения. Эти аксиомы относятся к так называемой *общей задаче проектирования*, понимаемой как прямое преобразование требований ТЗ в проектные решения и соответствующую документацию. Из неразрешимости общей задачи проектирования вытекает необходимость её декомпозиции на совокупность локальных задач в русле многоуровневой параллельно-последовательной схемы проектирования. Ввиду неполной определённости исходных данных и ограничений в общей задаче проектирования возникает потребность их прогнозирования, обмена информацией и проектными решениями между подразделениями предприятия. В силу логической противоречивости общей задачи проектирования требуется организация итерационных циклов коррекции проектных решений и перепроектирования. Из последней аксиомы вытекает эвристический характер обобщённого критерия оценки проектных решений.

В 1995 г. Ю.Р. Валькманом была предложена идея «исчисления» моделей [35]. Цель создания «исчисления» моделей состоит в разработке формального аппарата, обеспечивающего оценку возможности перехода от одной формальной теории к другой при трансформациях и интеграции моделей. Соответственно, «исчисление» моделей можно рассматривать как «оболочку» или среду, в которую заключаются классические и нетрадиционные формализмы, применяемые в ИП.

Возможный набор математических аппаратов для ИП опирается на аппараты: интегрального и дифференциального исчисления, исчисления предикатов, математической статистики, алгебры фреймов, а также теории множеств, теории графов, алгебры матриц, планирования экспериментов и др. В центре находится аппарат исчисления предикатов, что соответствует подходам общей теории проектирования (General Design Theory) – основного тренда науки о проектировании конца XX-века. Среди наиболее интересных работ в этой области надо отметить результаты двух японских учёных – С. Осуги [36, 37] и Х. Йошикавы [38].

Главная цель С. Осуги заключалась в построении интеллектуальных САПР для уровня концептуального проектирования. На стадии предварительного проектирования под моделью СТС (в узком смысле) понимается множество описаний её функциональных характеристик, связанных со структурой СТС. Работа с моделью производится с помощью правил преобразования модели, которые выражают проектные знания и хранятся в базе знаний. Для описания СТС в модели необходимо организовать циркуляцию больших объёмов данных, что требует интеграции проектных данных и проектных знаний.

Работа с моделью разрабатываемого объекта подразумевает наличие процессов двух типов: а) процессов анализа и оценки; б) процессов изменения и уточнения. Обычно модель СТС разбивают на две части: представление структуры разрабатываемого объекта и представление его функциональных характеристик, в том числе структурных ограничений.

В качестве языка представления проектных знаний С. Осуга [36] предложил расширение классической и многосортной логики с помощью предикатов высокого порядка, что позволяет выразить иерархическую структуру разрабатываемой СТС, а также с использованием наследования и других конструкций объектно-ориентированного подхода. Построенная им многоуровневая логика есть расширенная логика предикатов, которая включает структуру СТС (в виде структуры данных в базе данных) и её функциональные характеристики в виде предикатов. Структура данных включена в логику предикатов за счёт расширения её синтаксиса. В результате была построена интегрированная интерактивная интеллектуальная САПР KAUS (Design Knowledge Acquisition and Utilization System) [37].

По Х. Йошикаве, общая теория проектирования (ОТП) [38] – это формальная теория, которая стремится описать проектирование в терминах теории множеств и представляет этап предварительного проектирования с помощью частичных отображений из пространства технических требований (функций) в пространство проектных решений (атрибутов). Отличительной особенностью предложенного им подхода является широкое использование топологических понятий в ОТП. При этом весь процесс проектирования понимается как процесс развития, который переводит модель проектируемого объекта из одного состояния в другое вплоть до достижения требуемого решения. Основными характеристиками этого процесса являются «грубость» (robustness) и «сходимость». Грубость процесса проектирования означает, что при малых изменениях проектных требований будут наблюдаться малые изменения функциональных характеристик проектируемого объекта. Близкий смысл имеет термин «сходимость»: малые корректировки промежуточных проектных решений не влияют на схождение к конечному решению и приводят к созданию изделия с требуемыми свойствами.

Центральными процедурами проектирования здесь оказываются формирование метаэпистем (знаний о проектных знаниях), построение сценария, определяющего последовательность проектных процедур и создание метамодели, а также отображение пространства функций в пространство атрибутов. В ОТП используются различные типы рассуждений. Вначале главную роль играют абдуктивные рассуждения: на основе выделенных характерных свойств процесса проектирования путём абдукции формируются гипотезы проектных понятий и строятся аксиомы (аксиомы представления, соответствия, топологической структуры множества идеальных проектируемых объектов). Затем, исходя из ряда определений, путём дедукции выводятся теоремы ОТП, а в итоге создаётся модель процесса проектирования.

Следует отметить, что многие постулаты формальных систем, классической логики и теории рассуждений не соответствуют реалиям проектирования. В классической математической логике любое высказывание или предикат являются либо истинными, либо ложными, однако проектировщики в реальной жизни оперируют не точными знаниями, а предположениями, имеющими промежуточные значения истинности, которые требуют подтверждения или опровержения. Для классической логики одним из основных является закон непротиворечивости, и она не приемлет противоречий, тогда как в проектировании, например, в теории решения изобретательских задач (ТРИЗ) Г.С. Альтшуллера, обнаружение противоречия является основой алгоритма решения задачи. Значение (смысл) любого термина в математической логике определяется в точном соответствии с интерпретацией и остаётся неизменным в период работы системы рассуждений. Напротив, смысл термина в человеческих рассуждениях часто меняется в зависимости от опыта и контекста. Например, смысл термина «язык» в лингвистике и в проектировании различен. В традиционной математической логике значения истинности не изменяются со временем, тогда как мнения проектировщиков могут меняться после получения новой информации; как следствие, проектные решения могут пересматриваться и корректироваться.

В теории рассуждений процессы вывода определяются алгоритмами, т.е. являются точно предсказуемыми. Между тем, реальные рассуждения проектировщиков и проектные решения порой являются непредсказуемыми, а в процессе их подготовки происходят логические разрывы и скачки в другом направлении. Полученный в формальных исчислениях вывод может быть точно объяснён и повторен, например, с помощью процедур возврата назад, а в случае интуитивных рассуждений проектировщиков часто порождаются заключения, истоки и способы получения которых остаются неизвестными или непонятными. Этот список несоответствий можно было бы продолжить. Поэтому проблемы построения формализованных теорий, соответствующих реальным когнитивным возможностям проектировщиков, всё ещё остаются открытыми.

В отличие от вышеупомянутых вариантов логической и логико-топологической теории проектирования, Ю.Р. Валькман обращает особое внимание на необходимость привлечения в науку проектирования прагматической категории «контекст» модели, что означает переход от формальных систем к семиотическим моделям. В самом деле, когнитивный контекст проектирования неразрывно связан с семантикой, а ситуативный контекст – с прагматикой.

Уже в работе [35] Ю.Р. Валькман раскрывает ряд ключевых ролей контекстов моделей. Во-первых, посредством контекстов проектанты описывают траектории получения моделей, особенности их трактовки, специфику интерпретации соответствующих параметров и т.п. Тем самым проектные знания погружаются в вычислительную среду. Во-вторых, контексты представляют собой средства обеспечения адекватности моделей. В-третьих, они служат в качестве механизмов отчуждения моделей от их создателей. В-четвёртых, они являются основой для корректного проведения преобразований модели. В-пятых, они обеспечивают возможность интеграции моделей в законченные структуры, обусловленные проектируемой СТС. В-шестых, контексты предоставляют возможность определения единого терминологического базиса для общения разработчиков смежных профессий (проектантов, конструкторов, технологов).

Аргументируя в [39] необходимость использования семиотических моделей в ИИ, Д.А. Поспелов отмечал, что современные технические устройства работают на *досемиотическом уровне*, в силу чего они способны моделировать лишь простейшие формы поведения при решении творческих задач. В отличие от технических систем высшие животные и человек решают сложные задачи на *семиотическом уровне*, что позволяет им находить такие способы решения, которые невозможно реализовать без привлечения знаковых систем.

С целью преодоления ряда существенных ограничений формальных систем, таких как гипотеза о замкнутости мира, статический характер моделей знаний, жёсткость интерпретации, монотонность вывода (см. таблицу), Д.А. Поспелов [39] предложил расширение формальной системы, названное им *семиотической моделью*:  $SM = \langle T, R, A, I, \alpha_T, \beta_R, \gamma_A, \delta_I \rangle$ , где первые четыре компонента характеризуют формальную систему,  $\alpha_T$  – правила изменения алфавита,  $\beta_R$  – правила изменения синтаксических правил,  $\gamma_A$  – правила изменения аксиом,  $\delta_I$  – правила изменения правил вывода.

Здесь формальная система может пониматься как состояние сложной системы, а смена её состояния определяется изменением различных параметров формальной системы. В отличие от формальной системы, имеющей жёсткий синтаксис и жёсткую семантику, здесь синтаксис, семантика и прагматика являются гибкими, как и в случае многих знаков.

Основные различия между формальными системами и семиотическими моделями показаны в таблице.

Таблица – Основные различия между формальными системами и семиотическими моделями

ФОРМАЛЬНЫЕ СИСТЕМЫ	СЕМИОТИЧЕСКИЕ МОДЕЛИ
Замкнутые миры	Открытые миры
Статические модели знаний	Динамические модели знаний
Постоянная интерпретация	Переменная интерпретация
Единственная классическая логика	Сосуществование разных логик
Монотонные рассуждения	Возможность немонотонного вывода
Нет мультirezолюции	Мультirezолюция возможна

В монографии [1] Ю.Р. Валькман предложил интерпретацию модельно-параметрического подхода в виде семиотической метамодели, где под  $T$  понимается модельно-параметрический базис для проектных задач синтеза и анализа,  $R$  – синтаксически кор-

ректные процедуры синтеза моделей на основе базы данных,  $A$  – модели-аксиомы, введённые пользователем,  $I$  – процедуры определения и обоснования семантически корректных моделей,  $\alpha_T$  – процедуры построения и модификации  $\langle M, P \rangle$ -базиса,  $\beta_R$  – правила изменения синтаксически корректных процедур,  $\gamma_A$  – база данных и знаний по эвристическим приёмам ИП,  $\delta_I$  – правила определения семантической корректности.

### 3 От семиотического моделирования к прикладной и когнитивной семиотике

#### 3.1 О семиотике

*Семиотика* (термин Дж. Локка и Ч.С. Пирса [40]) или *Семиология* (термин Ф. де Соссюра [41]) есть наука о знаковых системах. Она возникла как самостоятельная дисциплина в XX-м веке, хотя ещё Дж. Локк ввёл этот термин и писал, что важная задача учения о знаках состоит в рассмотрении природы знаков, которыми ум пользуется для понимания вещей или для передачи своего знания другим. Ранние семиотические идеи восходят к концепции «возможных миров» Г. Лейбница. Один из видных представителей русской школы семиотики Г.Г. Шпет понимал её как «онтологическое учение о знаках вообще» (см. [42]). По определению другого крупного отечественного учёного в этой области Ю.М. Лотмана [43] семиотика есть наука о коммуникативных системах и знаках, используемых в процессе общения.

Знак – это объект, который представляет или замещает какой-либо другой объект или процесс. Для каждого знака обязательным является наличие двух сторон: означающей (материальной) стороны, воспринимаемой органами чувств, и означаемой стороны – значения, изучаемого в семантике. Так, для слов естественного языка означающей стороной является написание слова, а означаемой стороной – само его значение. При этом знаки одной семиотической системы (к примеру, слова русского языка) могут быть означающей стороной для знаков другой семиотической системы (например, математического языка), надстроенной над ними. В целом, установление связи между означающим и означаемым является необходимой предпосылкой любого понимания. Процесс производства и функционирования знаков называется *семиозисом*.

Классификацию знаков в зависимости от вида отношений между означаемым и означающим построили Ч.С. Пирс [40] и Ч.У. Моррис [44]. Ими выделены различные типы знаков: изобразительные знаки, указатели и символы. Так, знаки, основанные на реальном сходстве между означающим и означаемым, называются *изобразительными (иконическими)* знаками. Действие иконического знака основано на фактическом подобии означающего и означаемого, например, географической карты и местности, ярлыка с изображением принтера и принтера и т.п. Внутри этой группы Ч.С. Пирсом различаются образы и диаграммы. У последних отношения между составными частями означающей стороны сходны с отношениями между составными частями означаемой стороны.

Знаки, в которых между означающей и означаемой сторонами устанавливаются отношения смежности, называют *указателями* или *индексами* (например, стрелка – указатель пути к определённому месту). Действие указателей опирается на ассоциацию по смежности, например, дым есть указатель огня, замедление работы компьютера есть симптом вируса и т.п. Иными словами, в каждый момент времени указатель означает то, на что указывается.

В свою очередь, знаки, в которых отношение между означаемой и означающей сторонами условно, называются *символами*. Здесь связь между означаемым и означающим носит договорный характер: она определяется некоторым (часто неформальным) соглашением. Превращение иконических знаков в символы соответствует переходу от поверхностного к более глубокому пониманию. Знаки также делятся на собственные и переносные. *Собственные*

знаки используют для обозначения вещей, ради которых они изобретены, а *переносные* – тогда, когда сами вещи применяют для обозначения чего-либо другого.

Семиотический подход предполагает наличие, по крайней мере, трёх сторон изучения знаковых систем: синтактики, семантики и прагматики или, иначе, плана выражения, плана содержания и плана значения [44, 45]. *Синтактика* есть раздел семиотики, который исследует отношения между знаками внутри данной знаковой системы. *Семантика* охватывает сферу отношений между знаками и тем, что они обозначают, а *прагматика* – сферу отношений между знаками и теми, кто ими пользуется. Таким образом, прагматика учитывает требования конкретного адресата сообщения, в то время как у семантики его нет.

Семантика, как общая теория значения, подразделяется на *теорию референции* и *теорию смысла*. В теории референции рассматривают внешнее отношение знаковой формы к объектам реального мира (денотатам), а в основе теории смысла лежит внутренняя связь знаковой формы с соответствующими языковыми значениями и представлениями (смысл или интенционал знака). В первом случае семантическое отношение называют референцией (или десигнацией) знака, а во втором – означиванием (или сигнификацией).

Онтологическое моделирование в семиотике может опираться на анализ отношений в треугольнике Фреге (рисунок 6) или в родственных ему структурах: треугольнике Огдена-Ричардса и треугольнике Ульмана.

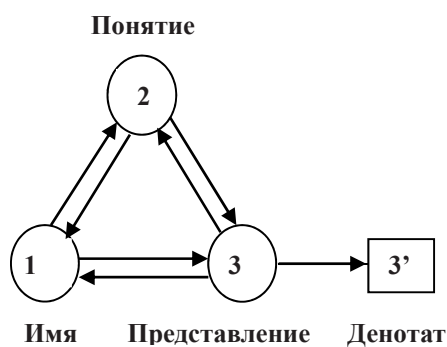


Рисунок 6 – Вариант треугольника Фреге

В контексте проектирования отношение  $1 \rightarrow 2$  соответствует проектным описаниям, а обратное отношение  $2 \rightarrow 1$  – системе обозначений. Отношение  $1 \rightarrow 3$  задаёт проектные представления, например, в виде чертежей, а обратное отношение именованья  $3 \rightarrow 1$  позволяет соотнести проектную или конструкторскую документацию со стадией проектирования. Отношение  $2 \rightarrow 3$ , соответствующее конкретизации, характеризует переход с более раннего на более поздний этап проектирования или с более высокого на более низкий уровень проектной деятельности. Обратное отношение  $3 \rightarrow 2$ , соответствующее обобщению, характеризует обратные переходы. Наконец, отношение  $3 \rightarrow 3'$  отображает переход от чертёжно-графической документации к реальной СТС.

### 3.2 От прикладной семиотики к когнитивной семиотике

В последние два десятилетия активно развиваются новые научные направления на стыке когнитивной науки и теории языка: когнитивная лингвистика, когнитивная семантика, онтологическая семантика, когнитивная семиотика. Независимо от изучаемого или проектируемого объекта все рассуждения о нём неизбежно строятся на языке. Язык – это система знаков, служащая средством человеческого общения, мышления и выражения. Язык возникает тогда, когда система сигналов превращается в систему знаков.

В основе *когнитивной лингвистики* лежит постулат о неразрывной связи между знанием и языком. Здесь язык понимается и как единый механизм познания, и как проявление общей способности человека познавать мир, других людей и самого себя, а традиционное разделение лингвистики на фонологию, синтаксис, морфологию и пр. уходит на второй план. По сути, в когнитивной лингвистике язык рассматривается как средство быстрого доступа к познавательным процессам. При этом изучаются связи отдельных когнитивных процессов и способностей человека с языком и формы их взаимодействия [31].

Таким образом, одной из главных категорий когнитивной лингвистики становится категория «знания»: при этом исследуются как виды знаний, так и способы их языкового представления. Специалисты по когнитивной лингвистике стремятся понять, как протекают про-

цессы восприятия, узнавания, осмысления, абстрагирования и как происходит формирование и накопление знаний. Одним из главных инструментов познания выступает языковая коммуникация, связанная с обменом знаниями и обеспечением взаимопонимания.

*Когнитивная семантика* изучает проблемы концептуализации (конструирования смыслов) и представления знаний как концептуальных структур [46, 47]. В ней исследуются соотношения между опытом, познанием и языком. В частности, значение изучается как когнитивный феномен. Ключевое положение теории значения Дж. Лакоффа состоит в том, что значение всегда включает интенциональность и человеческое понимание. Базовым для когнитивной семантики является интерпретирующий подход, что сближает её с герменевтикой.

В свою очередь, *онтологическая семантика* – это теория значения в языке, которая использует онтологию как основной ресурс для извлечения и представления смысла текста, построения логического вывода на знаниях, полученных из текстов, а также для генерации и восстановления текстов, исходя из представления их смысла [48].

*Когнитивная семиотика* (КС) представляет собой сравнительно новую научную парадигму, сочетающую знаковые и когнитивные структуры. Её зарождение и развитие демонстрирует, что познание отнюдь не является антитезой теории знаков. Одним из основателей этой дисциплины может по праву считаться Ю.Р. Валькман [49-51]. Он указывал, что появление КС предполагает движение когнитивных наук и семиотики встречным курсом. С одной стороны, КС есть наука о получении знаний из знаковых структур, а с другой стороны, речь идёт о синтезе знаковых конструкций, представляющих соответствующие структуры знаний. В основе методов КС лежит идея взаимодействия трёх миров (согласно эволюционной эпистемологии К. Поппера) – реального мира, мира знаков и ментального мира. Типичным способом получения знаний на основе модели знака является анализ связей в треугольнике Фреге (см. рисунок 6).

Существуют две разные позиции относительно истоков и периода возникновения КС. Ряд авторов относят появление КС к 1990-м годам и связывают его с расцветом когнитивной науки. Другие учёные считают, что у основоположников семиотики Ч.С. Пирса и Ч.У. Морриса изначально содержались зачатки КС. Широко известно высказывание Ч.С. Пирса «мы думаем только в знаках» [40]. При этом он допускал возможность взаимодействия между познанием и семиозисом. Более того, сам семиозис включает у него две одинаково важные части: производство знаков и их интерпретацию.

В свою очередь, Ч.У. Моррис ввёл прагматическое измерение семиозиса и предложил классификацию знаков, непосредственно связанную с типологией знаний [44]. Так, он различал:

- знаки-идентификаторы (т.е. знаки, которые выражают «где-знания?»);
- знаки-десигнаторы (знаки, отвечающие на вопрос «что такое?»);
- оценочные знаки, связанные с предпочтением, которые характеризуют «почему-знания?»;
- прескриптивные знаки, соответствующие «как-знаниям?»;
- формирующие знаки, или знаки систематизации (направляющие поведение интерпретатора в отношении других знаков).

Для обоснования той или иной позиции следует уточнить методологические установки классической и КС в плане описания человеческого поведения. В первую половину XX-го века, когда разрабатывались основы семиотики как общей теории знаков, двумя противоположными идеологическими течениями были ментализм и бихевиоризм.

*Ментализм* – это научная доктрина, согласно которой адекватная характеристика человеческого поведения и самой жизни невозможна без привлечения сознания или психики в целом как «средств объяснения». По сути, ментализм как философия сознания близок к *психо-*

логизму, так как его адепты утверждают, что любые попытки объяснить психические свойства, состояния и процессы путём учёта только физических и физиологических аспектов функционирования являются признаком редукционизма, т.е. недопустимым упрощения сущности дела. В соответствии с познавательными установками ментализма, для понимания человеческих действий и поступков требуется обращение к ментальным (т.е. описывающим ненаблюдаемый внутренний мир) категориям.

Во втором десятилетии XX-го века в исследовании поведения на смену ментализму пришёл *бихевиоризм*, исключивший психику из сферы научного знания. Классический бихевиоризм изучал поведение животных и человека на основе объективных наблюдений и экспериментов по схеме «стимул-реакция». Потом появился *необихевиоризм*, основанный на выражении научных понятий (и не только поведенческих) с помощью измерений. До 1960-х годов идеи бихевиоризма и необихевиоризма занимали господствующее положение. Затем на первый план вышло новое течение ментализма – когнитивизм, моделирующее познающего субъекта и процессы познания.

КС в нашей стране имеет глубокие корни; её зарождение связано как с семиотической школой Ю.М. Лотмана, так и с концепцией *семиотического моделирования* и *прикладной семиотики* Д.А. Пospelова. Согласно Ю.М. Лотману, основными вопросами для любой семиотической системы являются, во-первых, отношение к миру, лежащему за её пределами, а во-вторых, отношение статики к динамике [43]. Эти идеи открытости и динамики семиотической системы нашли отражение в формализованной семиотической модели Д.А. Пospelова. Ряд других соображений Ю.М. Лотмана, в частности, мысль о неоднородности (как минимум, двуязычности) семиотической системы как инструмента понимания, две модели коммуникации, в том числе автокоммуникация, когда в результате сдвига контекста и введения добавочного кода в «диалоге человека с самим собой» порождается новый смысл, а также концепция культуры как коллективного семиотического интеллекта, имеют прямое отношение к КС и ИИ.

Прикладная семиотика направлена на использование знаковых систем при разработке новых компьютерных, в том числе интеллектуальных, технологий. В русле интеграции семиотики и инженерии знаний было предложено опираться на знания как объекты знаковой природы. Стали появляться семиотические базы знаний, семиотические решатели, семиотические системы управления сложными объектами [45]. Всё это можно рассматривать как интеграцию семиотики, когнитологии и формальных систем.

Ещё в 1970 г. Д.А. Пospelов ввёл понятие «метазнак» [52]. Он писал, что кроме знаков, значениями которых выступают предметы или явления реального мира, можно рассматривать знаки знаков (метазнаки), значениями которых служат знаки семиотической системы первого уровня. Таковую систему следует отнести к знаковой системе второго уровня. Путём индукции нетрудно ввести системы знаков любого  $k$ -го уровня. Введение метауровня в знаковые представления явным образом предполагает в них внутреннюю интерпретируемость действий.

Метазнак в отличие от обычного знака несёт в себе «заряд активности». В треугольнике Фреге появляется дополнительная вершина, с которой связан фрагмент некоторой структуры на множестве знаков, которая играет роль денотата метазнака. Соответственно, возникает новая базовая семиотическая модель, которую Г.С. Осипов предложил называть «квадратом Пospelова» [53].

Истоки когнитивных течений восходят к пониманию роли знаний как ведущего фактора, определяющего эффективность действий человека. Когнитивистский подход в любой области подразумевает исследование ЖЦ знаний.

В КС центральное место занимает определение соответствий между знаковыми структу-



рами и структурами знаний. Здесь замечательным примером является описанное ещё в [45] соотношение знаков и фреймов и введение структуры знака-фрейма, т.е. треугольника «имя – протофрейм – экзофрейм» (рисунок 7).

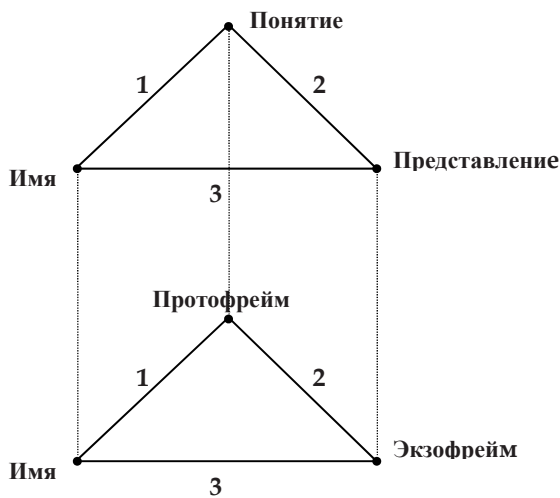


Рисунок 7 – К определению знака-фрейма

Зарубежные работы по КС появились существенно позже публикаций Д.А. Пospelова по семиотическому моделированию. Одним из пионеров КС считается Т. Даддезио, написавший книгу «О разуме и символах: релевантность когнитивных наук семиотике» [54]; среди часто цитируемых авторов выделим Й. Златева [55]. В России термин КС впервые использовал В.В. Рыков [56]. Первоначально он рассматривал КС как инструмент для интеграции и совершенствования систем извлечения знаний из текстов. Впоследствии было разработано описание бизнес-процессов в виде знаковых структур и проведена семантическая разработка таких описаний с целью получения новых знаний об этих бизнес-процессах.

Определение КС [49] легко расширить и не ограничиваться только уровнем знаний. С одной стороны, её можно интерпретировать как науку о развитии моделей и методов познания и конструирования смысла на основе знаковых структур. В рамках КС можно рассматривать весь спектр проблем синтеза и анализа смыслов, определения значений разных знаковых структур, понимания ситуативного поведения людей и их коллективов, познания мира в целом. С другой стороны, она призвана заниматься проблемами семиозиса, формирования семиотических отношений для когнитивных процессов и систем любого уровня (восприятие, представление, мышление, понимание).

Проиллюстрируем эту мысль на примере *когнитивного семиозиса* – процесса порождения и трансформации знаний, который определяет их синтаксис (представление в структурном мире), семантику (представление в ментальном мире) и прагматику (представление в реальном мире). Семиозис разворачивается внутри контура циркуляции знаний (рисунок 8<sup>2</sup>), который типичен для целенаправленной системы, где синтаксис характеризует структуру знаний, семантика – их смысл, содержание, а прагматика – их реальное воплощение в действиях или продуктах.

Рецепторы (сенсоры) обеспечивают входную информацию из внешнего мира, которая поступает в систему восприятия, где начинается процесс её обработки и представления. Синтаксическое наполнение продуцирует начальные структуры знаний, которые затем сохраняются, и начинается процесс развёртывания смысла. Этот смысл вместе с целью семиозиса определяют стадию семантизации знаний. Смысл подключает процесс формирования пове-

С помощью знака-фрейма можно пояснить шесть базовых процедур работы со знаниями и информацией. Первая пара процедур 1 связывает имена и понятия. Ей отвечают поиск информации по адресу и ассоциативный поиск информации по содержанию. Вторая пара процедур 2 соотносит понятия и представления. Речь идёт о синтезе конкретных представлений на базе понятия, т.е. порождении экзофреймов на базе протофрейма. Обратная процедура, которая связана с поиском понятия (протофрейм) по представлению (экзофрейм), соответствует приобретению знаний в ИИ. Наконец, третья пара процедур соединяет имена и представления.

<sup>2</sup> Схема семиозиса была в общих чертах обрисована Д.А. Пospelовым на его лекциях по прикладной семиотике в МЭИ.

дения, в результате чего появляются программы действий. Действия приводят к изменениям внешней среды, которые становятся частью прагматики знаний. Эти изменения регистрируются сенсорами, и цикл повторяется.



1са

Каждый модуль контура семиозиса на рисунке 8 можно трактовать как конечный автомат, который осуществляет переход от одной системы символов (знаков) к другой:

- из словаря Среды в словарь Восприятия;
- из словаря Сенсоров в словарь Базы Знаний;
- из словаря Восприятия в словарь Порождения Действий;
- из словаря Знаний в словарь средств Осуществления Действий;
- из словаря Планов в словарь Среды;
- из словаря Действий в словарь Сенсоров.

В каждом модуле должна обеспечиваться эффективность семиозиса с помощью специальных процедур обработки (сжатия) символической информации. Сжатие информации обеспечивается тремя операциями: группирование, фокусировка и комбинаторный поиск. В результате такого сжатия группы символов, которые характеризуются свойством общности, заменяются единичными символами.

В своих последних работах по КС [57,58] Ю.Р. Валькман рассмотрел различные виды и схемы семиозиса, обратился к гештальтам как основным ресурсам когнитивных семиотических операций, показал взаимосвязи между гештальтами и метафорами. Наряду со ставшим традиционным противопоставлением гештальта структуре им были прослежены переходы между ними. Выдвинута идея о том, что именно переходы от знаков к гештальтам обеспечивают понимание как текстов, так и ситуаций. Но все эти темы заслуживают отдельного обсуждения.

## ЗАКЛЮЧЕНИЕ

В статье сделана попытка подвести некий итог творческой деятельности доктора технических наук, профессора Юрия Роландовича Валькмана, ключевыми характеристиками которого были междисциплинарность и исключительная широта интересов. Он ставил и разрабатывал научные проблемы, относящиеся к различным и, на первый взгляд, слабо связанным

между собой областям: теории проектирования и семиотике, ИИ и проблемам образного мышления, компьютерной лингвистике и когнитивным наукам. После окончания университета и прихода Ю.Р. Валькмана в Институт кибернетики АН Украины, его первыми научными темами стали *обработка результатов испытаний* сложных объектов и *ИИП*. Затем возникли задачи *интеллектуализации САПР*, вопросы *инженерии проектных знаний*, в частности, построения специализированных моделей представления знаний на ранних стадиях проектирования – *модельно-параметрического пространства, семантико-прагматического пространства проектировщика* на базе системы оппозиционных шкал. Осознав ограниченность классических исчислений в плане отражения прагматических характеристик проектной деятельности, Ю.Р. Валькман предложил пути перехода от формальных к семиотическим моделям проектирования.

Ещё автор «Божественной комедии» сказал: «...сомнение доставляет мне не меньшее наслаждение, чем знание». Для Ю.Р. Валькмана как учёного важнейшим двигателем познания стали сомнение и со-мнение (свободное высказывание и отстаивание мнений, согласование мнений, формирование общих мнений на научных конференциях). Обладая даром сомнения и умением задавать вопросы, он поистине наслаждался общением с коллегами на конференциях КИИ (он активно участвовал почти во всех конференциях КИИ: от КИИ-1992 до КИИ-2016), «Интеллектуальные САПР», «Знания – диалог – решение». В 2000-е годы он неоднократно выступал с пленарными докладами на международных конференциях «Интегрированные модели и мягкие вычисления в искусственном интеллекте», «Интеллектуальный анализ информации», и др.

Потребности моделирования семантики и углублённого, «доформального» исследования *НЕ-факторов знаний* привели Ю.Р. Валькмана на международную междисциплинарную конференцию по компьютерной лингвистике и интеллектуальным технологиям «Диалог». Особую роль в его научной судьбе сыграл научный семинар «Отражение образного мышления и интуиции специалиста в системах ИИ» (Переславль-Залесский, 1998) и доклад Д.А. Поспелова «Метафора, образ и символ в познании мира», который определил крутой поворот к проблемам образного мышления, гештальтам, метафорам, взаимосвязям между знаками, образами и знаниями в КС.

Важнейшая миссия в жизни Юрия Роландовича была – «хранитель знаний». Он стал главным историографом русскоязычного сообщества специалистов в области ИИ, на протяжении многих лет фиксируя на видео основные конференции и семинары, доклады и дискуссии, неформальные мероприятия и события. Благодаря ему мы можем найти сегодня на YouTube и других ресурсах выступления и лекции ведущих учёных по всему спектру проблем, относящихся к ИИ и смежным наукам, разработке интеллектуальных систем и когнитивным технологиям.

Нить времён связывает между собой поколения учёных; их новые научные горизонты, нерешённые, но поставленные для других задачи – всё это остаётся людям. И память о Ю.Р. Валькмани – в его книгах и статьях. И в наших сердцах.

## Благодарности

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, проекты №17-07-01374 и №18-57-00014. Автор выражает искреннюю благодарность С.В. Смирнову и Н.М. Боргесту – инициаторам написания данной статьи.

## СПИСОК ИСТОЧНИКОВ

- [1] **Валькман, Ю.Р.** Интеллектуальные технологии исследовательского проектирования: формальные системы и семиотические модели. – Киев: Port-Royal, 1998. – 250 с.
- [2] **Валькман, Ю.Р.** От моделей жизненных циклов к С-технологиям/ Ю.Р. Валькман // Сборник научных трудов VI-й Национальной конференции по искусственному интеллекту (КИИ'98, Пущино, 5-11 октября 1998г.). Пущино: РАИИ, 1998. Т.2. – С.671-676.
- [3] **Kusiak, A.** (Ed.). Concurrent Engineering: Automation, Tools and Techniques. – New York: John Wiley and Sons, 1993. – 534 p.
- [4] **Pahl, G.** Engineering Design – A Systematic Approach / G. Pahl, W. Beitz, J. Feldhusen, K.-H. Grote. – Heidelberg: Springer-Verlag, 2007. – 617p.
- [5] **Тарасов, В.Б.** Искусственный интеллект в комплексной разработке машиностроительной продукции/ В.Б. Тарасов // Вестник МГТУ им. Н.Э. Баумана. Сер. «Машиностроение». 1995. №3. – С.10-15.
- [6] **Жук, Д.К.** Системные методы в программировании жизненных циклов новой техники / К.Д.Жук // Автоматизация проектирования сложных систем. – Мн: ИТК АН БССР, 1976. – С.15-26.
- [7] **Дубровина, Е.В.** Программные комплексы моделирования процессов эксплуатации сложных технических систем / Е.В. Дубровина, С.В. Игнатов, В.В. Литвинов и др. – Киев: Наукова думка, 1994. – 244 с.
- [8] **Кашуба, Л.А.** Параллельное проектирование средствами CAD/CAM/CAE в жизненном цикле изделий машиностроения / Л.А. Кашуба // Программные продукты и системы. 1998. №3. – С.24-30.
- [9] **Tarassov, V.B.** Concurrent Engineering and AI Methodologies / V.B. Tarassov, L.A. Kashuba, N.A. Cherepanov // Proceedings of the IFIP International Conference on Feature Modeling and Recognition in Advanced CAD/CAM Systems. Valenciennes, 1994. Vol.2. – P.869-888.
- [10] **Kimura, F.** Product Life Cycle Modeling for Inverse Manufacturing / F. Kimura, H. Suzuki // Proceedings of IFIPWG 5.3 International Conference on Life Cycle Modeling for Innovative Products and Processes (PROLAMAT'95, November 29-December 1, 1995)/ Ed. by F.L. Krause, H. Hansen. – Berlin: Springer-Verlag, 1996. – P.81-89.
- [11] **Тарасов, В.Б.** Концепция МетаКИП: от компьютерно интегрированного производства к Internet/ Intranet-сетям предприятий / В.Б. Тарасов // Программные продукты и системы. 1998. №3. – С.19-22.
- [12] **Норенков, И.П.** Основы автоматизированного проектирования. 3-е изд. – М.: МГТУ им. Н.Э. Баумана, 2006. – 448 с.
- [13] **Диксон, Дж.** Проектирование систем: изобретательство, анализ и принятие решений: Пер. с англ. – М.: Мир, 1969. – 440 с.
- [14] **Половинкин, А.И.** Автоматизация поискового конструирования (искусственный интеллект в машинном проектировании) / А.И. Половинкин, Н.К. Бобков, Г.Я. Буш и др. – М.: Радио и связь, 1981. – 344 с.
- [15] **Борисов, А.Н.** Интеллектуальные системы принятия проектных решений / А.Н. Борисов, А.В. Алексеев, Э.Р. Вилломс, Н.Н. Слядзь, С.А. Фомин. – Рига: Зинатне, 1997. – 317 с.
- [16] **Боргест, Н.М.** Ключевые термины онтологии проектирования: обзор, анализ, обобщения / Н.М. Боргест // Онтология проектирования. 2013. №3. – С.9-31.
- [17] **Валькман, Ю.Р.** Мульти-modalность, мультимедийность, мультиконтекстность, мультимодельность представления понятий в базах знаний / Ю.Р. Валькман // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник трудов IV-й Международной научно-практической конференции (ИММВ'2007, Коломна, 28-30 мая 2007г.). – М.: Физматлит, 2007. Т.1. – С.76-83.
- [18] **Нариньяни, А.С.** НЕ-факторы и инженерия знаний: от наивной формализации к естественной прагматике/ А.С. Нариньяни // Сборник трудов IV-й Национальной конференции по искусственному интеллекту с международным участием (КИИ-94, Рыбинск, 15-21 сентября 1994 г.). – Тверь: ЦПС-АИИ, 1994. Т.1. – С.9-18.
- [19] **Нариньяни, А.С.** НЕ-факторы: краткое введение/ А.С. Нариньяни // Новости искусственного интеллекта. 2004. №2. – С.52-63.
- [20] **Валькман, Ю.Р.** Моделирование НЕ-факторов: основа интеллектуализации компьютерных технологий/ Ю.Р. Валькман // Новости искусственного интеллекта. 2004. №2. – С.64-81.
- [21] **Тарасов, В.Б.** НЕ-факторы: от семиотического анализа к методам формализации / В.Б. Тарасов // Новости искусственного интеллекта. 2004. №2. – С.95-114.
- [22] **Тарасов, В.Б.** Интеллектуальные системы в проектировании / В.Б. Тарасов // Новости искусственного интеллекта. 1994. №3. – С.24-67.
- [23] **Шрейдер, Ю.А.** Системы и модели / Ю.А. Шрейдер, А.А. Шаров. – М.: Радио и связь, 1982. – 152 с.
- [24] **Валькман, Ю.Р.** Модельно-параметрическое пространство в исследовательском проектировании: цели построения, определения, структура и свойства / Ю.Р. Валькман // Вопросы когнитивно-информационной поддержки постановки и решения новых научных проблем. Киев: ИК НАН Украины, 1995. – С.103-115.

- [25] **Валькман, Ю.Р.** Модельно-параметрическое пространство: теория и применение / Ю.Р. Валькман, В.И. Гриценко, А.Ю. Рыхальский. – Киев: Наукова думка, 2012. – 192 с.
- [26] **Валькман, Ю.Р.** Модельно-параметрическое пространство – средство представления знаний исследователей сложных систем / Ю.Р. Валькман, А.Ю. Рыхальский // Управляющие системы и машины. 2009. №1. – С.20-30.
- [27] **Поспелов, Д.А.** Знания и шкалы в модели мира / Д.А. Поспелов // Модели мира. – М.: РАИИ, 1997. – С.69-84.
- [28] **Тарасов, В.Б.** Анализ и моделирование НЕ-факторов на полярных шкалах / В.Б. Тарасов // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник трудов международного научно-практического семинара (ИММВ-2001, Коломна, 17-18мая 2001г.). – М.: Физматлит, 2001. – С.65-71.
- [29] **Валькман, Ю.Р.** Видеообразы в операциях исследовательского проектирования: отношения между абстрактным и конкретным, логичным и метафоричным, объективным и субъективным, информационным и когнитивным / Ю.Р. Валькман // Сборник научных трудов V-й Национальной конференции по ИИ (КИИ-96, Казань, 5-11 октября 1996г.). Т.III. –Казань: ЗАО ПО «Спецтехника» – АИИ, 1996. –С.389-393.
- [30] **Налимов, В.В.** Вероятностная модель языка. О соотношении естественных и искусственных языков. 2-е изд. – М.: Наука, 1979. –304 с.
- [31] **Мартынов, В.В.** В центре сознания человека. – Мн.: БГУ, 2009. – 272 с.
- [32] **Зенкин, А.А.** Когнитивная компьютерная графика.– М: Наука, 1991. –192 с.
- [33] **Валькман, Ю.Р.** Графическая метафора – основа когнитивной графики / Ю.Р. Валькман // Сборник трудов IV-й Национальной конференции по ИИ (КИИ-94, Рыбинск, 15-21 сентября 1994 г.). – Тверь: ЦПС–АИИ, 1994. Т.1.– С.94-100.
- [34] **Жук, К.Д.** Построение современных систем автоматизированного проектирования / К.Д. Жук, А.А. Тимченко, А.А. Родионов и др. – Киев: Наукова думка, 1983. – 248 с.
- [35] **Валькман, Ю.Р.** Исчисление моделей – основа интеллектуализации процессов исследовательского проектирования / Ю.Р. Валькман // Программные продукты и системы. 1995. №4. – С.18-23.
- [36] **Осуга, С.** Обработка знаний: Пер. с япон. – М.: Мир, 1989.– 293 с.
- [37] **Ohsga, S.** Toward Intelligent CAD Systems / S.Ohsuga // Computer-Aided Design.1989.Vol.21, №5. P.315-337.
- [38] **Yoshikawa, H.** General Design Theory as a Formal Theory of Design / H. Yoshikawa // Intelligent CAD / H. Yoshikawa, D. Gorsard (Eds.). – Amsterdam: Elsevier, 1989. – P.51-61.
- [39] **Поспелов, Д.А.** Ситуационное управление: теория и практика. – М.: Наука, 1986.–288 с.
- [40] **Пирс, Ч.С.** Логика как семиотика: теория знаков: Пер. с англ. / Ч.С. Пирс // Метафизические исследования. Вып. 11. Язык. – СПб, 1999. – С.199-217.
- [41] **Соссюр де, Ф.** Курс общей лингвистики: Пер с франц. – М.: Едиториал УРСС, 2004. –278 с.
- [42] **Почепцов, Г.Г.** История русской семиотики. – М.: Лабиринт, 1998. – 336 с.
- [43] **Лотман, Ю.М.** Семиосфера. – СПб.: Искусство-СПБ, 2000. –704 с.
- [44] **Моррис, Ч.У.** Основания теории знаков: Пер. с англ. / Ч.У. Моррис // Семиотика. Антология. 2-е изд. – М.: Академический проект, 2001. – С.45-97.
- [45] **Поспелов, Д.А.** Прикладная семиотика / Д.А. Поспелов, Г.С. Осипов // Новости искусственного интеллекта. 1999. №1. – С.9-35.
- [46] **Лакофф, Дж.** Женщины, огонь и опасные вещи: Что категории языка говорят нам о мышлении: Пер. с англ. – М.: Языки славянской культуры, 2004. – 792 с.
- [47] **Кузнецов, О.П.** Когнитивная семантика и искусственный интеллект / О.П. Кузнецов // Искусственный интеллект и принятие решений. 2012. №4. – С.32-42.
- [48] **Nirenburg S., Raskin V.** Ontological Semantics. Cambridge MA: MIT Press, 2004. – 440 p.
- [49] **Валькман, Ю.Р.** О когнитивной семиотике / Ю.Р. Валькман // Интеллектуальный анализ информации. Сборник трудов XII-й международной научной конференции им. Т.А. Таран (ИАИ-2012, Киев, 16-18 мая 2012г.). – Киев: Просвіта, 2012. – С.19-30.
- [50] **Валькман, Ю.Р.** Когнитивность семиотики / Ю.Р. Валькман // Интеллектуальный анализ информации. Сборник трудов XIII-й международной научной конференции им. Т.А. Таран (ИАИ-2013, Киев, 15-17 мая 2013г.). – Киев: Просвіта, 2013. – С.69-81.
- [51] **Валькман, Ю.Р.** Когнитивная семиотика: семиозисы и гештальты / Ю.Р. Валькман // Труды XIV-й Национальной конференции по искусственному интеллекту с международным участием (КИИ-2014, Казань, 24-27 октября 2014). – Казань: Изд-во РИЦ «Школа», 2014. Т.1.– С.180-188.
- [52] **Поспелов, Д.А.** Системный подход к моделированию мыслительной деятельности / Д.А. Поспелов // Проблемы методологии системного исследования. – М.: Мысль, 1970. – С.333-358.
- [53] **Осипов, Г.С.** От ситуационного управления к прикладной семиотике / Г.С. Осипов // Новости искусственного интеллекта. 2002. №6. – С.3-7.

- [54] **Daddesio, Th.C.** On Minds and Symbols: the Relevance of Cognitive Science to Semiotics. – Berlin: Mouton de Gruyter, 1994. – 263 p.
- [55] **Zlatev, J.** Cognitive Semiotics: an Emerging Field for Transdisciplinary Study of Meaning / J. Zlatev // The Public Journal of Semiotics. 2012. Vol. IV. №1. – P.2-24.
- [56] **Рыков, В.В.** Извлечение знаний – реализация когнитивной семиотики. – <http://rykov-cs.narod.ru/dlg9.html>.
- [57] **Валькман, Ю.Р.** Когнитивная семиотика: гештальты и метафоры / Ю.Р. Валькман // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник трудов VIII-й международной научно-практической конференции (ИММВ'2015, Коломна, 18-20 мая 2015г.). – М.: Физматлит, 2015. Т.1.–С.54-59.
- [58] **Валькман, Ю.Р.** Когнитивная семиотика: гештальты и знаки, целостность и структура / Ю.Р. Валькман // Труды XV-й Национальной конференции по искусственному интеллекту с международным участием (КИИ-2016, Смоленск, 3-7 октября 2016 г.). – Смоленск: Универсум, 2016. Т.2. – С.250-258.
- 

## FROM ONTOLOGIES OF DESIGNING TO COGNITIVE SEMIOTICS

**Yury R. Valkman<sup>1</sup>**, **Valery B. Tarassov<sup>2</sup>**

<sup>1</sup>*International Research and Training Center of Information Technologies and Systems NAS and MES of Ukraine, Kiev, Ukraine*

<sup>2</sup>*BaumanMoscowStateTechnicalUniversity, Moscow, Russia*  
*Vbulbov@yahoo.com*

### Abstract

Some fundamentals of lifecycle concept for Complex Technical Systems are considered. The Complex Technical Systems lifecycle is a meta-system for computer-aided design, computer-integrated manufacturing and maintenance, enabling the mapping of Complex Technical Systems state changes during the period of its existence. Main lifecycle trends are both extension and detailing that leads to non-traditional design strategies and emergence of new design phases at the crossroad of research and development, design and tests. Within the framework of ontological approach design phases, levels, aspects and tasks are viewed. Here the phase of preliminary design (also called «research design») seen as a sort of «analytical synthesis» is of primary concern. An important role of various Non-Factors (uncertainty, imprecision, contradiction, fuzziness) in preliminary design is shown. A three-dimensional space for Non-Factors modeling is introduced and their classifications are suggested. On early stages of design a model-based approach is crucial: it deals with both models and parameters. Some original techniques and tools of representing design knowledge in a model-parametric space are developed. Besides, a multi-dimensional semantic-pragmatic space of design methods and languages based on bipolar scales is outlined. Two keynote principles for design languages – compatibility principle and complementarity principle – are formulated on the basis of semantic scale for languages. On the one hand, the support of designer's creative thinking requires the use of both illustrative and cognitive graphics. On the other hand, two variants of General Design Theory are analyzed. By General Design Theory a formal design theory is meant that tends to describe the design concepts in terms of sets, topological spaces and calculi, in particular, higher-order predicate calculus. The analysis of formal systems shortcomings from the viewpoint of making real design decisions is performed. As a result, the need in bringing semiotic models into design science is justified. In the final part of the paper, basic definitions of semiotics together with sign classifications are given. The peculiarities of the transition from semiotic modeling to Applied Semiotics and Cognitive Semiotics are discussed. Cognitive Semiotics faces the problems of both knowledge acquisition from sign systems and the synthesis of sign-based structures in the context of knowledge engineering. Professor Yury R. Valkman was one of the pioneers in Cognitive Semiotics.

**Keywords:** *design ontology, complex technical system, lifecycle, preliminary design, knowledge non-factors, model-parametric space, general design theory, semiotics, applied semiotics, cognitive semiotics.*

**Citation:** *Valkman YuR, Tarassov VB. From ontologies of designing to cognitive semiotics [In Russian]. Ontology of designing. 2018; 8(1): 8-34. - DOI: 10.18287/2223-9537-2018-8-1-8-34.*

### Acknowledgment

The work was supported by the Russian Foundation for Basic Research, projects No.17-07-01374 and No.18-57-00014. The author expresses sincere thanks to S.V. Smirnov and N.M. Borgest – initiators of the writing of this paper.

## References

- [1] **Valkman YuR.** Intellectual'nye technologiyi issledovatel'skogo proectirovaniya [Intelligent Technologies for Preliminary Design]. [In Russian]. – Kiev: Port-Royal, 1998. – 250 p.
- [2] **Valkman YuR.** Ot modeley zhiznennykh tsykhov k C-technologuyam [From Lifecycle Models to C-Technologies]/ Sbornik nauchnykh trudov VI Natsional'noy konferencii po iskusstvennomu intellektu CAI'98 [Proceedings of the VI<sup>th</sup> National Conference on Artificial Intelligence]. [In Russian]. – Puschino: RAAI, 1998. Vol.2. – P.671-676.
- [3] **Kusiak A.** (Ed.). Concurrent Engineering: Automation, Tools and Techniques. New York: John Wiley and Sons, 1993. – 534 p.
- [4] **Pahl G.** Engineering Design – A Systematic Approach/ G.Pahl, W.Beitz. 2<sup>nd</sup> edition. Heidelberg: Springer-Verlag, 1996. – 544 p.
- [5] **Tarassov VB.** Iskusstvenniy intellekt v kompleksnoy razrabotke machinostroyitel'noy produktsii [Artificial Intelligence in a Complex Development of Machine-Building Production] // Vestnik MGТУ imeni Baumana. Ser. «Machinostroyeniye» [BMSTU Herald. Machine-Building Ser.]. [In Russian]. 1995. No3. - P.10-15.
- [6] **Zhuk KD.** Systemniye metody v programmirovanii zhiznennykh tsykhov novoy tekhniki [Systemic Methods in Programming Lifecycles of New Technique] // Avtomatizatsiya proektirovaniya slozhnykh system [Computer-Aided Design of Complex Systems]. [In Russian]. – Minsk: ITC of Belorussian Academy of Sciences, 1976. – P.15-26.
- [7] **Dubrovina EV, Ignatov SV, Litvinov VV, et al.** Programniye komplekxy modelirovaniya processov ekspluatatsii slozhnykh tekhnicheskikh system [Software Complexes for Modeling the Processes of Using Complex Technical Systems]. [In Russian]. – Kiev: Naukova Dumka, 1994. – 244 p.
- [8] **Kashuba LA.** Parallel'noye proektirovaniye sredstvami CAD/CAM/CAE v zhisennom tsykle izdeliy machinostroyeniya [Parallel Design by Using CAD/CAM/CAE Tools in the Lifecycle of Machine-Building Products]/ L.A. Kashuba // Programmye produkty i systemy [Software and Systems]. [In Russian]. 1998. No3. - P.24-30.
- [9] **Tarassov VB, Kashuba LA, Cherepanov NA.** Concurrent Engineering and AI Methodologies // Proceedings of the IFIP International Conference on Feature Modeling and Recognition in Advanced CAD/CAM Systems. – Valenciennes, 1994. Vol.2. – P.869-888.
- [10] **Kimura F, Suzuki H.** Product Life Cycle Modeling for Inverse Manufacturing // Proceedings of IFIP WG 5.3 International Conference on Life Cycle Modeling for Innovative Products and Processes (PROLAMAT'95, November 29-December 1, 1995)/ Ed. by F.L. Krause, H. Hansen. – Berlin: Springer-Verlag, 1996. – P.81-89.
- [11] **Tarassov VB.** Conceptsya MetaKIP: ot computerno integrirovannogo proizvodstva k Internet/ Intranet–setyam predpriyatiy [MetaCIM Concept: from Computer-Integrated Manufacturing to Internet/Intranet-Networks of Enterprises] // Programmye produkty i systemy [Software and Systems]. [In Russian]. 1998. No3. – P.19-22.
- [12] **Norenkov IP.** Osnovy avtomatizirovannogo proectirovaniya [Bases of Computer-Aided Design]. 3<sup>rd</sup> ed. [In Russian]. – Moscow: Bauman Moscow State Technical University, 2006. – 448 p.
- [13] **Dixon JR.** Design Engineering: Inventiveness, Analysis and Decision-Making. – New York; McGraw Hill, 1966. – 354 p.
- [14] **Polovinkin AI, Bobkov NK, Bush GYa, et al.** Avtomatizatsiya poiskovogo konstruirovaniya (iskusstvenniy intellekt v machinnom proectirovanii) [Computer-Aided Search Design (Artificial Intelligence in Machine-Based Design)] [In Russian]. – Moscow: Radio i svyaz, 1981. – 344 p.
- [15] **Borisov AN, Alekseev AB, Vilums ER, Slyadz NN, Fomin SA.** Intellectual'nye systemy prinyatiya proectnykh resheniy [Intelligent Systems of Design Decision-Making]. [In Russian]. – Riga: Zinatne, 1997. – 317 p.
- [16] **Borgest NM.** Keywords of Ontology of Designing: Review, Analysis, Generalization [In Russian]. Ontology of Designing. 2013. No3. – P.9-31.
- [17] **Valkman YuR.** Mul'timodalnost, mul'timediynost, mul'ticontextnost, mul'timodelnost predstavleniya ponyatiy v bazakh znaniy [Multi-Modal, Multi-Media, Multi-Context, Multi-Model Concept Representation in Knowledge Base] // Integrirovanniye modeli i myagkie vychisleniya v iskusstvennom intellekte. Sbornik trudov IV mezhdunarodnoy nauchno-practicheskoy konferencii [Integrated Models and Soft Computing in Artificial Intelligence. Proceedings of the 4<sup>th</sup> International Scientific-Practical Conference]. [In Russian]. – Moscow: PhysMathLit, 2007. Vol.1. – P.76-83.
- [18] **Narin'yan AS.** Ne-factory i ingeneriya znaniy: ot naivnoy formalizatsii k yestestvennoy pragmatike [Non-Factors and Knowledge Engineering: from Naïve Formalization to Natural Pragmatics] // Sbornik trudov IV natsional'noy konferencii po iskusstvennomu intellektu CAI'94 [Proceedings of the IV<sup>th</sup> National Conference on Artificial Intelligence]. [In Russian]. – Tver: AAI, 1994. Vol.1. – P.9-18.
- [19] **Narin'yan AS.** Ne-factory: kratkoye vvedenie [Non-Factors: a Short Introduction] // Novosti iskusstvennogo intelekta [Artificial Intelligence News]. [In Russian]. 2004. No2. – P.52-63.
- [20] **Valkman YuR.** Modelirovaniye ne-factorov: osnova intellektuslizatsii computernykh technologiy [Non-Factors Modeling: the Basis for Intelligent Software Technologies] // Novosti iskusstvennogo intelekta [Artificial Intelligence News]. [In Russian]. 2004. No2. – P.64-81.

- [21] **Tarasov VB.** Ne-factory: ot semioticheskogo analiza k metodam formalizatsii [Non-Factors: from Semiotic Analysis to Formalization Methods] // *Novosti iskusstvennogo intellecta [Artificial Intelligence News]*. [In Russian]. 2004. No2. – P.95-114.
- [22] **Tarasov VB.** Intellectualnye systemy v proektirovanii[Intelligent Systems in Design] / V.B. Tarasov // *Novosti iskusstvennogo intellecta [Artificial Intelligence News]*. [In Russian]. 1994. No3. – P.24-67.
- [23] **Shreider YuA, Sharov AA.** Systemy I modeli [Systems and Models]. [In Russian].– Moscow: Radio i svyaz, 1982. – 152 p.
- [24] **Valkman YuR.** Modelno-parametricheskoye prostranstvo v issledovatel'skom proektirovanii:tseli postroyeniya, opredeleniya, structura i svoystva [Model-Parametric Space in Preliminary Design: Construction Goals, Definitions, Structure and Properties] // *Voprosy kognitivno-informatsionnoy podderzhki postanovki i resheniya novykh nauchnykh problem [Issues of Cognitive-Information Support in Putting and Solving New Scientific Problems]*. [In Russian]. – Kiev: Glushkov Institute of Cybernetics of NASU, 1995. – P.103-115.
- [25] **Valkman YuR, Gritsenko VI, RychaskyAYu.** Modelno-parametricheskoye prostranstvo: teoriya i primeneniye [Model-Parametric Space: Theory and Applications]. [In Russian].–Kiev: Naukova Dumka, 2012. – 192 p.
- [26] **Valkman YuR, RychaskyAYu.** Modelno-parametricheskoye prostranstvo – sredstvo predstavlenniya znaniy issledovatelye slozhnykh system [Model-Parametric Space – a Knowledge Representation Tool for Researchers of Complex Systems] /*Upravlyauschiye systemy i mashiny [Control Systems and Machines]*. [In Russian].2009. No1. – P.20-30.
- [27] **Pospelov DA.** Znaniya i shcaly v modelimira [Knowledge and Scales in the World Model] // *Modeli mira [World Models]*. [In Russian]. – Moscow: RAAI, 1997. – P.69-84.
- [28] **Tarasov VB.** Analiz i modelirovaniye ne-factorov na polyarnykh shcalakh [Analysis and Modeling of Non-Factors on Bipolar Scales] // *Integrirovanniye modeli i myagkie vychisleniya v iskusstvennom intellecte. Sbornik trudov mezhdunarodnogo nauchno-practicheskogo seminar [Integrated Models and Soft Computing in Artificial Intelligence. Proceedings of the International Scientific-Practical Workshop]*. [In Russian]. – Moscow: PhysMath-Lit, 2001. – P.65-71.
- [29] **Valkman YuR.** Video-obrazy v operatsiyakh issledovatel'skogo proektirovaniya: otnosheniya mezhdru abstractnym i concretnym, loguichnym i metaphorichnym, obyektivnym i subyektivnym, informatsionnym I kognitivnym [Video-images in Operations of Preliminary Design: Relationships between Abstract and Concrete, Logical and Metaphorical, Objective and Subjective, Informational and Cognitive] // *Sbornik nauchnykh htrudov V natsional'noy konferencii po iskusstvennomu intellectu CAI'96 [Proceedings of the V<sup>th</sup> National Conference on Artificial Intelligence]*. [In Russian]. – Kazan': AAI, 1996. Vol.3. – P.389-393.
- [30] **Nalimov VV.** Veroyatnostnaya model yazyka. O sootnoshenii mezhdru yestestbennymi I iskusstvennymi yazykami [Probabilistic Model of Language. On the Relationship between Natural and Artificial Languages.2<sup>nd</sup> ed. [In Russian]. – Moscow: Nauka, 1979. – 304 p.
- [31] **Martynov VV.** V tsentre soznaniya cheloveka [In the Center of Human Conscience]. [In Russian]. – Minsk: BSU, 2009. – 272 p.
- [32] **Zenkin AA.** Kognitivnaya computernaya graphika [Cognitive Computer Graphics]. [In Russian]. – Moscow: Nauka, 1991. –192 p.
- [33] **Valkman YuR.** Graphicheskaya metaphora – osnova kognitivnoy graphiki [Graphic Metaphor – a Basis of Cognitive Graphics] // *Sbornik trudov IV natsional'noy konferencii po iskusstvennomu intellectu CAI'94 [Proceedings of the IV<sup>th</sup> National Conference on Artificial Intelligence]*. [In Russian]. – Tver: AAI, 1994. Vol.1. – P.94-100.
- [34] **ZhukKD, Timchenko AA, Rodionov AA, et al.** Postroyeniye sovremennykh system automatizirovannogo proektirovaniya [Development of Modern Computer-Aided Design Systems]. [In Russian]. – Kiev: Naukova Dumka, 1983. – 248 p.
- [35] **Valkman YuR.** Ischisleniye modeley – osnova intellectualizacii processov issledovatel'skogo proektirovaniya [Model Calculus – a Basis for Making Intelligent Processes of Preliminary Design] // *Programmnye producty i systemy [Software and Systems]*. [In Russian]. 1995. No4. – P.18-23.
- [36] **Ohsuga S.** Obrabotka znaniy [Knowledge Processing]. [In Russian]. – Moscow: Mir, 1989. – 293 p.
- [37] **Ohsuga S.** Toward Intelligent CAD Systems// *Computer-Aided Design*. 1989. Vol.21, No5. – P.315-337.
- [38] **Yoshikawa H.** General Design Theory as a Formal Theory of Design // *Intelligent CAD / H. Yoshikawa, D. Gorsard (Eds.)*. – Amsterdam: Elsevier, 1989. – P.51-61.
- [39] **Pospelov DA.** Situatsionnoye upravleniye: teoriya i praktika [Situational Control: Theory and Practice]. [In Russian]. – Moscow: Nauka, 1986. – 288 p.
- [40] **Peirce ChS.** Issues of Pragmatism. Vol.II. Logical Foundations of the Theory of Signs. – Mineola NY: Dover Publications, 1958. – 388 p.
- [41] **Saussure de F.** Cours de linguistique générale. – Genève: L'Arbre d'Or, 2005. – 254 p.
- [42] **Pocheptsov GG.** Istoriya russkoy semiotiki [History of Russian Semiotics]. [In Russian]. – Moscow: Labyrinth, 1998. – 336 p.



- [43] **Lotman YuM.** Semiosphera [The Semiosphere]. [In Russian]. – St. Petersburg: Iskusstvo-SPb, 2000. – 704 p.
- [44] **Morris ChW.** Signification and Significance: A Study of the Relations of Signs and Values. Chapter 1. Signs and the Act. – Cambridge MA: MIT Press, 1964. – 99 p.
- [45] **Pospevov DA, Osipov GS.** Prikladnaya Semiotika [Applied Semiotics] // Novosti iskusstvennogo intellecta [Artificial Intelligence News]. [In Russian]. 1999. No1. – P.9-35.
- [46] **Lakoff G.** Women, Fire and Dangerous Things: What Categories Reveal about the Mind. – Chicago: University of Chicago Press, 1987. – 614 p.
- [47] **Kuznetsov OP.** Kognitivnaya semantika i iskusstvenniy intellect [Cognitive Semantics and Artificial Intelligence] // Iskusstvenniy intellect i prinyatiye resheniy [Artificial Intelligence and Decision Making]. [In Russian]. 2012. No4. – P.32-42.
- [48] **Nirenburg S, Raskin V.** Ontological Semantics. Cambridge MA: MIT Press, 2004. – 440 p.
- [49] **Valkman YuR.** O kognitivnoy semiotike [On Cognitive Semiotics] // Intellectual'niy analiz informatsii. Sbornik trudov XII mezhdunarodnoy konferencii imeni T.A.Taran [Information Mining. Proceedings of the XII<sup>th</sup> T.A. Taran International Conference]. [In Russian]. – Kiev: Prosvita, 2012. – P.19-30.
- [50] **Valkman YuR.** Kognitivnost semiotiki (Semiotics Cognitivity) // Intellectual'niy analiz informatsii. Sbornik trudov XIII mezhdunarodnoy konferencii imeni T.A. Taran [Information Mining. Proceedings of the XIII<sup>th</sup> T.A. Taran International Conference]. [In Russian]. – Kiev: Prosvita, 2013. – P.69-81.
- [51] **Valkman YuR.** Kognitivnaya semiotika: semiozisy i geshtalty [Cognitive Semiotics: Semiosises and Gestalts] // Trudy XIV natsional'noy konferencii po iskusstvennomu intellectu CAI 2014 [Proceedings of the XIV<sup>th</sup> National Conference on Artificial Intelligence]. [In Russian]. – Kazan': RPC Shkola, 2014. Vol.1. – P.180-188.
- [52] **Pospelov DA.** Systemniy podkhod k modelirovaniyu myslitel'noydeyatelnosti [A Systemic Approach to the Modeling of Intellectual Activity] // Problemy metodologuii systemnogo issledovaniya [Methodological Problems of Systemic Research]. [In Russian]. – Moscow: Mysl', 1970. – P.333-358.
- [53] **Osipov GS.** Ot situatsionnogo upravleniya k prikladnoy semiotike [From Situational Control to Applied Semiotics] // Novosti iskusstvennogo intellecta [Artificial Intelligence News]. [In Russian]. 2002. No6. – P.3-7.
- [54] **Daddesio ThC.** On Minds and Symbols: the Relevance of Cognitive Science to Semiotics. – Berlin: Mouton de Gruyter, 1994. – 263 p.
- [55] **Zlatev J.** Cognitive Semiotics: an Emerging Field for Transdisciplinary Study of Meaning // The Public Journal of Semiotics. 2012. Vol. IV. No1. – P.2-24.
- [56] **Rykov VV.** Knowledge Acquisition – Cognitive Semiotics Implementation [In Russian]. – <http://rykovcs.narod.ru/dlg9.html>.
- [57] **Valkman YuR.** Kognitivnaya semiotika: geshtalty i metaphory [Cognitive Semiotics: Gestalts and Metaphors] // Integrirovanniye modeli i myagkie vychisleniya v iskusstvennom intellecte. Sbornik trudov VIII mezhdunarodnoy-nauchno-practicheskoy konferencii [Integrated Models and Soft Computing in Artificial Intelligence. Proceedings of the VIII<sup>th</sup> International Scientific-Practical Conference]. [In Russian]. – Moscow: PhysMathLit, 2015. Vol.1. – P.54-59.
- [58] **Valkman YuR.** Kognitivnaya semiotika: geshtalty i znaki, tselostnost i structura [Cognitive Semiotics: Gestalts and Metaphors, Integrity and Structure] // Sbornik trudov XV natsional'noy konferencii po iskusstvennomu intellectu CAI 2016 [Proceedings of the XV<sup>th</sup> National Conference on Artificial Intelligence]. [In Russian]. – Smolensk: Universum, 2016. Vol.2. – P.250-258.

## Сведения об авторах



**Валькман Юрий Роландович** (1948-2017). Родился 26 марта 1948 г. в городе Пайде в Эстонии. Окончил в 1971 г. Киевский государственный университет им. Т.Г. Шевченко. Работал в Институте кибернетики им. В.М. Глушкова в отделе, который занимался разработкой систем автоматизированного проектирования сложных объектов и автоматизированной обработкой экспериментальных данных. В 1986 г. Ю.Р. Валькман защитил кандидатскую диссертацию по специальности 05.13.11 – математическое и программное обеспечение вычислительных машин и систем, а в 1997 г. в Твери – докторскую диссертацию на тему «Исследование сложных объектов в модельно-параметрическом пространстве» (по специальности 05.13.16).

В 1970-1990-е годы принимал участие в разработке крупных программно-информационных комплексов систем автоматизации исследований и проектирования сложных объектов новой техники: ТЕМП – система обработки резуль-

татов лётных испытаний (Жуковский), ЧЕРТЁЖ – система автоматизации исследовательского проектирования сложных изделий кораблестроения (Ленинград), ИЗИН – система управления базами данных, ориентированная на обработку результатов исследований (Калинин), ГЕЛИОГРАФ – автоматизированная система научных исследований гидрофизических объектов (Севастополь).

В Институте кибернетики Ю.Р. Валькман проработал около 25 лет вплоть до его структурной реорганизации и переориентации прикладной тематики. Вместе с академиком В.И. Скурихиным он перешёл в образованный в 1997 г. МНУЦ информационных технологий и систем НАН и МОН Украины на должность заведующего отделом распределённых интеллектуальных систем.

В 2000-е годы вёл активную преподавательскую работу в качестве профессора кафедры математических методов системного анализа Национального технического университета Украины «Киевский политехнический институт им. Игоря Сикорского». Также работал по совместительству профессором Киевского национального университета им. Тараса Шевченко, Киево-Могилянской академии.

Автор более 300 научных публикаций, в том числе 5 монографий, 9 учебно-методических пособий. Среди его лучших книг: «Информационные технологии в испытаниях сложных объектов: методы и средства», «Интеллектуальные технологии исследовательского проектирования: формальные системы и семиотические модели», «Модельно-параметрическое пространство: теория и применение».

Профессор Ю.Р.Валькман был одним из ведущих специалистов Украины в области интеллектуальных систем и технологий, вице-президентом Международной ассоциации создателей и пользователей интеллектуальных систем, членом Советской, а затем Российской ассоциации искусственного интеллекта и Международной ассоциации когнитивных исследований. Круг научных интересов Ю.Р.Валькмана был весьма широк: теория проектирования и искусственный интеллект, экспертные системы, базы знаний, интеллектуальные информационные системы в исследовании и моделировании сложных технических объектов, прикладная семиотика, инженерия образов и когнитивные науки.

Профессор Ю.Р. Валькман был одним из главных организаторов международных научных конференций по искусственному интеллекту, проводившихся на Украине: «Знание – Диалог – Решение», «Интеллектуальный анализ информации», членом программных комитетов международных научных конференций: «Международный конгресс по интеллектуальным системам и информационным технологиям», «Интеллектуальные САПР», «Интегрированные модели и мягкие вычисления в искусственном интеллекте», «Открытые семантические технологии проектирования интеллектуальных систем», «Гибридные и синергетические интеллектуальные системы» и др. Уже, будучи тяжело больным, он организовал Круглый стол по когнитивной семиотике на конференции КИИ-2016. С 2011 г. был активным членом редколлегии журнала «Онтология проектирования», также работал в составе редколлегий международных научных журналов «Программные продукты и системы», «Интеллектуальные системы и технологии», «Радиоэлектроника, информатика, управление».

Ю.Р. Валькман внёс большой вклад в подготовку кадров по системному анализу и ИИ на Украине. Награждён Почётной грамотой Министерства образования и науки Украины (2006), Грамотой Председателя Верховной Рады Украины за весомый личный вклад в развитие отечественного образования, подготовку высококвалифицированных специалистов, плодотворную научно-педагогическую деятельность, многолетний добросовестный труд и высокий профессионализм (2009).

**Yury Rolandovich Valkman** (1948-2017) was graduated from Shevchenko Kiev State University in 1971. He received his candidate's degree (in 1986) in computer science and doctoral degree (in 1997) in CS applications in research and design. Almost 25 years he worked at Glushkov Institute of Cybernetics of Ukrainian Academy of Sciences. Later on he moved to the International Research-Education Center of Information Technologies as the Head of Department of Distributed Intelligent Systems. He was also Professor at National Technical University of Ukraine «Sikorsky Kiev Polytechnic Institute». He was the author and co-author of more than 300 publications, including 5 monographs.



**Тарасов Валерий Борисович**, 1955 г. рождения. Окончил МВТУ им. Н.Э. Баумана в 1978 г., к.т.н. (1982), доцент (1999). Заместитель заведующего кафедрой «Компьютерные системы автоматизации производства» МГТУ им. Н.Э. Баумана. Член научного совета Российской ассоциации искусственного интеллекта, вице-президент Российской ассоциации нечётких систем и мягких вычислений. В списке научных трудов более 250 работ в области искусственного интеллекта, теории проектирования и САПР, многозначных и нечётких логик, когнитивных наук.

**Valery Borisovich Tarasov** (b. 1955) graduated from the Bauman Moscow State High Technical School (actually Bauman Moscow State Technical University), PhD (1982), Deputy Head of the CIM Department at BMSTU. He is the member of the Scientific Council of Russian Association for Artificial Intelligence, Vice-President of Russian Association for Fuzzy Systems and Soft Computing. He is the author and co-author of more than 250 publications in the fields of AI, design theory and CAD, many-valued and fuzzy logics, cognitive sciences.

УДК 597.97

## ФОРМАЛИЗАЦИЯ ПОЗНАВАТЕЛЬНОГО ПРОЦЕССА НА ОСНОВЕ БАЗИСА МОДЕЛЕЙ

**С.В. Микони**

Санкт-Петербургский институт информатики и автоматизации Российской академии наук,  
Санкт-Петербург, Россия  
smikoni@mail.ru

### Аннотация

Вся умственная деятельность человека представляет собой моделирование его внешнего и внутреннего мира и основывается на применении моделей. Познавательную и конструктивную деятельность человека невозможно представить через огромное многообразие используемых им частных моделей. В работе предлагаются структурная, функциональная и операционная составляющие, присущие в разной степени любой разрешимой модели. Универсальность составляющих объясняется их принадлежностью модели языка предикатов первого порядка, положенной в основу любых математических моделей. Эти базовые свойства модели используются для описания модели знания, приобретаемого человеком в процессе познания окружающей и внутренней среды. Сам познавательный процесс осуществляется также с применением моделей, названных в работе моделями инструментального знания. В качестве базовой модели познавательного процесса принята «пирамида познания» Акоффа. Определены используемые в ней понятия. Каждый этап формирования модели нового знания и используемого для его приобретения инструментального знания описывается с применением базовых моделей, что позволяет представить познавательный процесс в терминах моделирования.

**Ключевые слова:** модель, свойства модели, моделирование, познание, онтологическая модель, знание, умение, данные, информация.

**Цитирование:** Микони, С.В. Формализация познавательного процесса на основе базиса моделей / С.В. Микони // Онтология проектирования. – 2018. – Т. 8, №1(27). – С. 35-48. – DOI: 10.18287/2223-9537-2018-8-1-35-48.

### Введение

Мировоззрение складывается из совокупности частных моделей, описывающих фрагменты окружающего и внутреннего мира человека. Работа [1] посвящена обоснованию того, что любая активность человека основана на моделировании, как неотъемлемого, внутренне присущего компонента любых внешних действий человека (познания реальности или преобразования её). Автор приводит многочисленные убедительные примеры применения моделей в познавательной и практической деятельности человека. Моделирование познавательной деятельности человека автор иллюстрирует на примере «пирамиды познания» Р. Акоффа [2], включающей следующие уровни познания: добычу данных (*data mining*), извлечение из них информации, пополнение (формирование) знания, его понимание и мудрость. Как это и бывает во многих случаях, русский и английский смысл словосочетания ”*data mining*” не совпадают. Буквально, по-русски, оно и означает добычу данных, но носители языка понимают под этим извлечение из полученных данных информации, под которой понимаются полезные (информативные) для познающего субъекта данные.

В работе [1] утверждается, что любую модель можно представить комбинацией из трёх типов моделей. К ним автор относит: модель *состава* (перечень существенных частей систе-

мы), модель *структуры* системы (перечень существенных связей между частями системы) и модель *чёрного ящика* (перечень существенных связей системы с окружающей средой).

Утверждение автора, что «все наши знания, явные и неявные, существуют только в виде моделей», не подлежит сомнению. Сложнее обстоит дело с базисом моделей. Модель чёрного ящика трактуется автором не в математическом, а в системном смысле – в качестве целостного образа моделируемого объекта. В математической интерпретации термин «ящик», как модель «вещи в себе или для нас» имеет три разновидности: чёрный, серый и белый. В понятие *чёрного ящика* вкладывается неизвестность формальной модели функции  $\psi$ , осуществляющей отображение  $\psi: X \rightarrow Y$ . Под *серым ящиком* понимается тот факт, что получена приближённая функция  $\varphi$ , например, в виде многочлена или уравнения регрессии, воспроизводящая отображение  $\varphi: X \rightarrow Y$ . Под *белым ящиком* понимается известность реальной функции  $f$ , осуществляющей отображение  $f: X \rightarrow Y$ .

Недостатком предложенного базиса моделей является взаимозависимость первых двух «типов моделей». Любая структура представима графом – отношением на множестве вершин (сущностей). Нуль-граф – граф без связей и представляет собой модель состава, т.е. совокупность не связанных между собой сущностей. Следовательно, модель состава является частным, хоть и вырожденным, вариантом структуры.

Целью настоящей работы является формализация познавательного процесса с применением уточнённого базиса моделей.

## 1 Формализация общих свойств модели

Наибольшей общностью для формирования типовых свойств модели обладает модель языка предикатов первого порядка, представленная многосортной алгебраической системой  $\langle A, C, F, P \rangle$  [3]. В ней:  $A$  – множество предметных переменных (носитель),  $C$  – множество констант (домен значений переменной),  $F$  – множество функций,  $P$  – множество предикатов.

В силу тождественности (изоморфизма) предиката, как логической функции  $p: C_1 \times \dots \times C_n \rightarrow B, B = \{\text{True}, \text{False}\}$ , и  $n$ -местного отношения  $R \subset C_1 \times \dots \times C_n$ , справедлива другая форма записи алгебраической системы:  $\langle A, C, F, R \rangle$ , в которой символ предикатов заменён на символ отношений  $R$ . Пару  $\langle F, R \rangle$  называют *сигнатурой* системы. Использование только одного из этих символов представляют собой частные случаи алгебраической системы. К ним относятся соответственно *реляционная система* (модель)  $\mathbf{B} = \langle A, C, R \rangle$  и *алгебра*  $\mathbf{A} = \langle A, C, F \rangle$  [4].

При интерпретации символов  $A$  и  $R$  реляционной системы множествами вершин  $V$  и связей  $E$  между ними (рёбер и/или дуг) переходим к модели графа  $G = (V, C, E)$ . Граф называется *помеченным*, если множество  $C$  содержит имена его вершин и связей, и *взвешенным* (нагруженным), если множество  $C$  содержит численные оценки вершин и связей. Графическое изображение реляционной системы является наглядной моделью любой структуры. Поэтому непредметно помеченный граф представляет *структурную составляющую* модели, которую назовём *C-моделью*.

Любая функция  $f \in F$  в алгебре  $\mathbf{A} = \langle A, C, F \rangle$  представляет собой отображение области её определения в область значений. Представим носитель  $A$  через множества определения  $X$  и значений  $Y$  функции  $f: X \subset A, Y \subset A$ . Пусть функция  $f$  отображает в область значений  $Y$   $n$  аргументов  $x_1, \dots, x_n: y = f(x_1, \dots, x_n)$ . Если нас не интересуют внутренние связи между аргументами отображения, выражение  $y = f(x_1, \dots, x_n)$  представляет собой функциональную модель (*Ф-модель*) или модель «чёрного ящика».

В алгебре  $\mathbf{A}$  функциям ставятся в соответствие одноимённые операции, например, сложение или умножение. Свойства операций определяются относительно элементов носителя

$A$ , являющихся их аргументами. По имени известной функции можно установить результат отображения  $f$ . Например, в таблице умножения двум сомножителям 5 и 5 ставится в соответствие результат 25. Однако его вычисление требует выполнения совокупности элементарных операций, описываемых соответствующим алгоритмом. Функция называется *вычислимой*, если существует вычисляющий её алгоритм [5].

Пошаговое выполнение алгоритма влечёт изменение внутреннего состояния объекта действия – от  $q(t)$  к  $q(t + 1)$ . Эти изменения привязываются к моментам времени  $t = 0, 1, 2, \dots$ , где  $t = 0$  фиксирует начальное состояние объекта. Закономерность смены состояний описывается *функцией переходов*  $f_p$ , дополняющей *функцию выхода* объекта  $f_v$ .

Полагая  $f_p, f_v \in F$  и  $q \in Q$ , получаем следующее расширение  $\Phi$ -модели, соответствующее модели конечного автомата:

$$(1) \quad M_{\text{КА}} = \langle X, Y, Q, f_p, f_v, t \rangle$$

В отличие от  $\Phi$ -модели модель конечного автомата  $M_{\text{КА}}$  характеризует не правило, а процесс реализации функции, т.е. является её операционной моделью (*О-моделью*). Последовательность выполнения операций, реализующая выраженную функцией переходов причинно-следственную связь между состояниями автомата, позволяет отнести О-модель к классу динамических моделей.

С развитием параллельных вычислений алгоритм получил расширительное толкование как набор инструкций, описывающих порядок действий исполнителя для достижения некоторого результата [6]. О-модель параллельных вычислений определяет, например, процесс перемещения меток в такой дискретно-событийной модели, как сеть Петри. В искусственном интеллекте О-моделью описывается процесс логического вывода следствия.

В предметных областях (ПрО) О-модель определяется как «упорядоченная совокупность (система) взаимосвязанных действий, направленных на достижение определённой цели» [7]. Примерами О-модели в ПрО являются инструкции и технологические карты. По существу, О-моделью обобщаются такие понятия, как метод, алгоритм, инструкция и т.п. Иными словами, О-модель, формализуя процедуры, относится к моделям *процедурного* типа.

Раскрытие чёрного ящика описывается структурно-функциональной моделью (*СФ-моделью*). СФ-модель представляет собой композицию  $\Phi$ -моделей, связи между которыми описываются С-моделью. На связях С-модели формируются функции *соотнесения* и *измерения*. Функции соотнесения представляются помеченным, а функции измерения – взвешенным графом. Метка связи двух вершин графа представляет собой имя отношения, в котором находятся эти вершины. По степени отвлечённости от предмета рассмотрения отношения делятся на *предметные* и *отвлечённые* (абстрактные). Примером предметного отношения является « $A$  атакует  $B$ ». Модели с предметными отношениями называют семантическими сетями. Примерами отвлечённых отношений являются: род-вид, целое-часть, класс-элемент, сущность-свойство, причина-следствие. По существу, семантическая сеть является предметной конкретизацией онтологической модели.

Функции измерения представляются *взвешенным* графом. Вес связи количественно характеризует либо меру близости вершин графа, либо пропускную способность связи. Близость может трактоваться, например, расстоянием между вершинами или мерой их сходства. Модели с весами связей в роли их пропускной способности относятся к *потокowym* моделям. Первичной характеристикой потока через дугу графа является величина непрерывного потока, либо интенсивность дискретного потока, например, потока заявок. Вторичными характеристиками потока являются, например, время достижения им смежной вершины или стоимость транспортировки.

Для преобразования С-модели в СФ-модель метки, входящие в носитель структуры, ассоциируются с функциями, реализуемыми  $\Phi$ -моделями, входящими в СФ-модель. Поскольку

элементами структуры являются вершины и ребра графа, для ассоциации меток с реализуемыми ими функциями используются отношения:

$$R_{fv} \subset A \times F, R_{fv} = \{(a_i, f_i)\}, a_i \in A, f_i \in F, i = \overline{1, n} \text{ и}$$

$$R_{fe} \subset E \times F, R_{fe} = \{(a_i, a_j), f_{ij}\}, a_i, a_j \in A, f_{ij} \in F, i \neq j, i, j = \overline{1, n}.$$

Функции СФ-модели рассматриваются с точки зрения их роли в модели.

С учётом объединения С-модели с Ф-моделью и разделения множества отношений на 3 вида  $R = R_c \cup R_{fv} \cup R_{fe}$  СФ-модель имеет вид:

$$(2) \quad M_{сф} = \langle A, R_c, R_{fv}, R_{fe}, C, F \rangle.$$

Для построения СФ-модели на основе С-модели необходимо дополнительно сформировать множество функций  $F$  и два множества  $R_{fv}$  и  $R_{fe}$ .

*Пример.* Представить схему решения выражения  $y = (6 \cdot x_1 + 3 \cdot x_2) \cdot 5$  СФ-моделью. Это выражение имеет 5 аргументов (5 входных вершин графа): 2 переменных –  $x_1$  и  $x_2$  и 3 константы – 6, 3 и 5. Оно содержит 3 операции умножения ( $\times$ ), одно сложение и операцию присваивания значения выражения (5 промежуточных вершин графа) выходной переменной  $y$  (одна выходная вершина графа). Таким образом, 11 вершин графа образуют носитель  $A = \{a_1, \dots, a_{11}\}$ , а отношение смежности на носителе  $R_c = \{(a_1, a_7), (a_2, a_7), (a_3, a_8), (a_4, a_8), (a_7, a_9), (a_8, a_9), (a_9, a_{10}), (a_5, a_{10}), (a_{10}, a_{11}), (a_{11}, a_6)\}$ .

Создание теоретико-множественной СФ-модели на основе известных множеств  $A, F, C, R_c$  заключается в формировании множеств  $R_{fv}, R_{fe}$  выражения (2):

- 1) множество функций  $F$  расширим переменными, считая их за разновидность функции одной переменной, и функциями связей:  $F = \{x_1, x_2, \times, +, :=, \text{пд}\}$ . Функция «пд» связи  $(a_i, a_j)$  означает передачу данных от  $i$ -й вершины графа к  $j$ -й вершине.
- 2) В примере множество констант  $C$  содержит целые числа 6, 3, 5:  $C = \{6, 3, 5\}$ .
- 3) На основе множеств  $A$  и  $F$  сформируем двумерное множество  $R_{fv} = \{(a_1, 6), (a_2, x_1), (a_3, 3), (a_4, x_2), (a_5, 5), (a_6, y), (a_7, \times), (a_8, \times), (a_9, +), (a_{10}, \times), (a_{11}, :=)\}$ .
- 4) На основе множеств  $A$  и  $R_c$  сформируем двумерное множество  $R_{fe} = \{(a_i, a_j), \text{пд}\}, i, j = \overline{1, 11}$ , отражающее присвоение всем дугам графа роли передачи данных.

Теоретико-множественная СФ-модель, построенная путём конкретизации выражения (2), пригодна для ввода в ЭВМ, являясь машинно-ориентированной (компьютерной) моделью. Этот пример показывает возможность порождения конкретных компьютерных моделей на основе частных формальных моделей алгебраической системы путём конкретизации их символов и их интерпретации свойствами предметного знания.

Графическая СФ-модель схемы решения задачи изображена на рисунке 1. Наряду с меткой  $a_i, i = \overline{1, 11}$ , каждой вершине графа присвоена переменная, либо константа, либо одна из операций: умножение, сложение, присвоение вычисленного значения.

Функциональные модели, входящие в состав СФ-модели, на рисунках обычно именуется блоками, благодаря чему СФ-модели в проектировании получили название блок-схем. Блок-схемы востребованы в силу своей наглядности и обзорности.

Ф- и СФ-модели содержат информацию о том, какие свойства объекта-оригинала моделируются, а О-модель – как решить задачу моделирования. Исходя из теории разрешимости, ни Ф-модель, ни СФ-модель не результативны в отсутствие О-модели. Именно она используется в моделировании с целью получения результата и в этом смысле является связующим звеном между моделируемыми свойствами объекта и реализующей их системой моделирования.

Таким образом, исходя из формального описания модели в языке предикатов первого порядка четвёркой  $\langle A, C, R, F \rangle$ , можно утверждать, что для выполнения моделирования Ф-

модель и СФ-модель должны содержать *операционную* составляющую. В совокупности с ней разрешимые модели обозначаются как ФО- и СФО-модели.

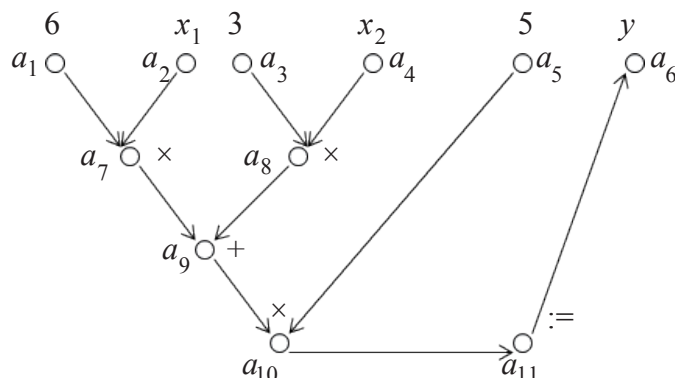


Рисунок 1 – Структурно-функциональная модель схемы решения задачи

ФО- и СФО-модели, не подготовленные к выполнению в среде моделирования, относятся к классу *описательных* или **D-моделей** (от *Descriptive*). Они существуют, но не погружены в среду моделирования. Примером D-модели в программировании является текст программы.

Описательная модель, погружённая в среду моделирования, т.е. представленная на её языке, относится к классу *выполнимых* или **E-моделей** (от *Executable*). Примером E-модели является компилированная программа, готовая к выполнению. Соответствующие файлы имеют расширения языка программирования и выполнения (\*.exe от *executive*). Выполняемая версия модели активна в породившей её системе моделирования.

Поскольку среда моделирования также описывается моделью, существует необходимость введения понятия *модель среды моделирования*. В широком смысле под ней будем понимать любую естественную или искусственную среду. Примером естественной среды моделирования является мозг человека, а искусственной среды – всё то, что им создано, например, программная система моделирования. Именно её и будем понимать в узком смысле под обобщённой моделью среды моделирования (под «машинной» **M-моделью**). Она, в свою очередь, может быть выражена через модели, входящие в базис моделей и их комбинации.

Машинной формой O-модели (**OM-моделью**) является, например, структурированный запрос, запрограммированный на языке SQL (*Structural Query Language*). Примером выполнимой ФО-модели (**ФОМ-модели**) служит модуль динамической библиотеки в программной системе. Примером выполнимой СФО-модели в объектно-ориентированном программировании (**СФОМ-модели**) является *класс*, объединяющий структуру данных и операции над ними.

Все разновидности моделей, описанных выше, сведены в таблицу 1. В верхней половине таблицы приведены 5 разновидностей описательных моделей, а в нижней половине – 5 разновидностей выполнимых моделей, встроенных в среду моделирования. Из них самостоятельными (разрешимыми) являются *описательные* ФО- и СФО-модели, и *выполнимые* ФОМ- и СФОМ-модели.

Предложенные элементарные и комбинированные типы моделей позволяют моделировать любой объект как изолированно, так и в составе среды моделирования. Используем их для описания процесса познавательной деятельности человека.

Таблица 1 - Виды моделей

<b>Вид модели</b>	<b>С-модель</b>	<b>Ф-модель</b>	<b>О-модель</b>	<b>М-модель</b>
С-модель	√			
Ф-модель		√		
О-модель			√	
ФО-модель		√	√	
СФО-модель	√	√	√	
СМ-модель	√			√
ФМ-модель		√		√
ОМ-модель			√	√
ФОМ-модель		√	√	√
СФОМ-модель	√	√	√	√

## 2 Уровни познания

Познание является многоэтапным процессом извлечения информации из первичных данных, полученных в результате прямого взаимодействия с познаваемым явлением, объектом, процессом. Р. Акофф в образной форме предложил различать пять уровней познания [8]: «Содержание людских умов и, следовательно, то, чему можно обучиться, разделяется на пять классов: *данные, информация, знание, понимание, мудрость*». Он представил их в форме ступеней «пирамиды познания» (рисунок 2).

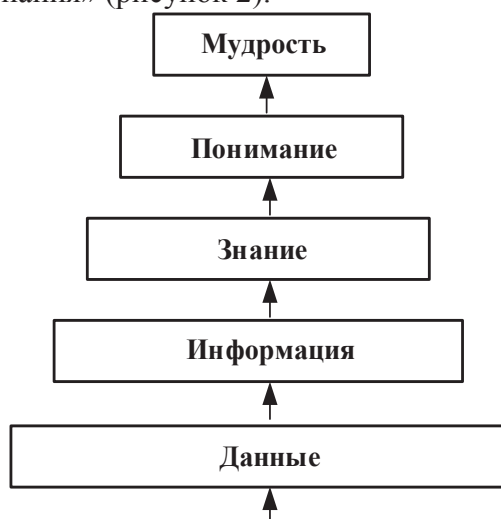


Рисунок 2 – Пирамида ступеней познания по Акоффу

На каждом уровне познания происходит обогащение знания об объекте познания – от примитивного знания фактов через создание индивидуальной модели знания до мудрости его применения – в соответствии с афоризмом Клода Андриана Гельвеция: «Знание некоторых принципов легко возмещает незнание некоторых фактов» [9]. В книге [1] правомерно утверждается, что «на каждом уровне познания происходит построение следующей новой модели знания, последовательное углубление переработки исходной эмпирической информации».

В связи с этим представляют интерес действия, которые предпринимаются для перехода к каждой последующей ступени познания. Они диктуются различием ступеней познания, которые заложены в определениях соответствующих понятий [10-13]. Не приводя и не обсуждая эти определения, проиллюстрируем различия понятий образными примерами.

В образной трактовке [1]: «данные подобны добытой руде, которую предстоит подвергнуть целой серии технологических преобразований, прежде чем получится нужный предмет



из металла, содержащегося в добытой руде». Продолжим такую трактовку, отталкиваясь от сырьевого значения английского слова *mining*, на практическом опыте автора, в студенческие годы принявшего участие в геологической экспедиции по поиску алмазов на северо-востоке Пермского края. Для геолога информацией об алмазе является наличие сопутствующей ему породы в образцах минералов. А роль данных играют камни, которые участник геологической партии собирает в труднопроходимой местности.

В пятидесятые годы прошлого века процесс обнаружения месторождения алмазов осуществлялся следующим образом. Участники геологической партии обходили обследуемую местность и к вечеру приносили в лагерь рюкзаки камней. Назовём их «сырыми» (необработанными) данными. Далее камни переправлялись на базу экспедиции, где измельчались в порошок, превращаясь в «обработанные данные». Затем осуществлялся отбор нужной породы промыванием порошка в реке. Положительным результатом такой камеральной обработки являлось нахождение алмазоносной породы – информации, интересующей геологов. По аналогии с изложенной технологией переход от сырых данных к информации осуществляется в три этапа: *добыча*, *обработка* и *отбор* данных. Согласно [14] отбор данных понимается как частный случай их *классифицирования* на два класса.

В терминах работы [1] полученная информация описывается моделью состава. В системе Пролог состав базы знаний описывается в секции *domains*. В терминах нашей работы информация представляет собой носитель *A*, на котором должны быть определены функциональные и иные отношения, причём каждый элемент носителя описывается Ф-моделью. В задачу следующего шага познания входит нахождение связей между элементами носителя. В системе Пролог связи между элементами носителя описываются в секции *clauses*. Итогом этого процесса является создание СФ-модели изучаемого явления. В терминах искусственного интеллекта эта модель характеризует *знание* «как устроено».

Секции *domains* и *clauses* представляют собой декларативную составляющую Пролог-программы. Процедурная часть программы задаёт способ получения результата, реализующий правила логического вывода и описываемый О-моделью. Эта модель характеризует *знание* «как решать» (*knowhow*).

Следует отметить, что для подтверждения на практике недостаточно знать найденную закономерность. Нужно ещё суметь ею воспользоваться. А это означает встраивание модели знания в систему моделирования и получение результата моделирования. Эта ступень практического познания отсутствует в «пирамиде познания» Р. Акоффа. Она характеризует *умение* как способность выполнять какое-либо действие по определённым правилам в среде моделирования.

**Понимание** определим как универсальную операцию мышления, связанную с усвоением нового содержания, включением его в систему устоявшихся идей и представлений [15]. В этом определении важна его последняя часть. Для понимания недостаточно только узнать что-то новое. Важно совместить его с уже имеющимся знанием, обосновать его истинность. В случае непротиворечивости нового знания с имеющимся знанием оно используется для пополнения имеющегося знания. В случае противоречивости нового знания с имеющимся знанием оно используется для уточнения (частичной перестройки) имеющегося знания. Пониманию близко и понятие осмысление, как процесс принятия решения о пополнении или перестройке имеющегося знания на основе полученного нового знания.

Понятие **мудрость** имеет различные трактовки. Одна из них означает *умение грамотно применять свои знания*, умение найти соответствие между моделью конкретного знания и моделью текущего состояния внешней среды.

Модель познавательного процесса в форме последовательности выполняемых этапов представлена на рисунке 3.

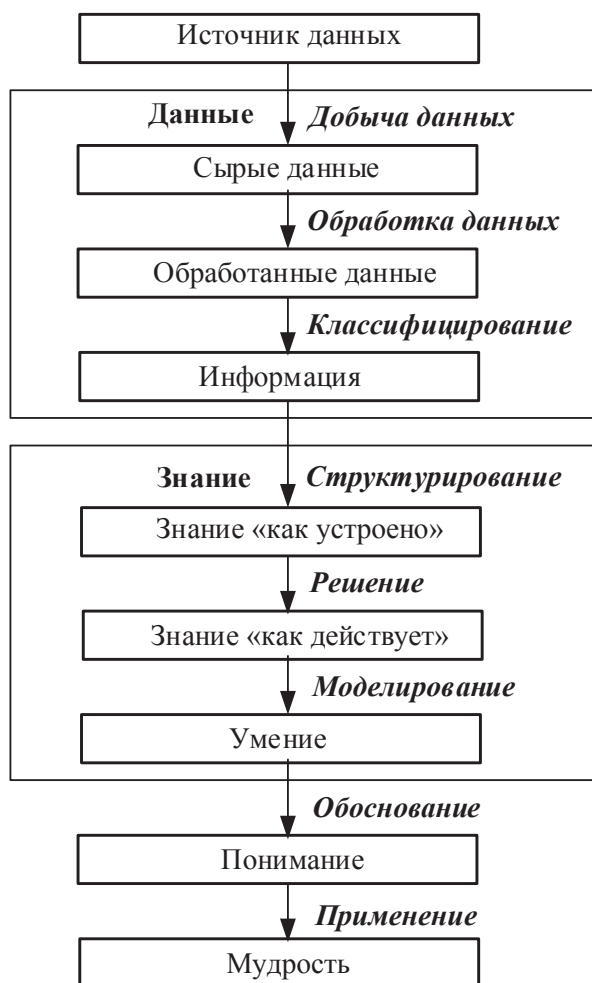


Рисунок 3 – Модель познавательного процесса

Операции над моделью данных объединены в блок *данные*, а операции над моделью знания – в блок *знание*. Этап *понимание* познавательного процесса требует применения моделей знания, аналогичных модели знания изучаемого явления, а на этапе *мудрость* дополнительно участвуют связанные с ней модели внешней среды.

Процесс моделирования познавательной деятельности проиллюстрируем на примере проектирования онтологической модели.

### 3 Проектирование онтологической модели

Проектирование онтологической модели можно разделить на три этапа: создание модели данных (определение состава модели знания в терминологии Ф. Тарасенко и информации – в терминах Р. Акоффа [1, 2]), создание модели знания и сопоставление полученной модели знания с имеющимся знанием. Процесс проектирования онтологической модели проиллюстрируем примером создания модели уровней управления сложным объектом [16]. Роль первичных данных при создании онтологической модели играют слова естественного языка. Ключевым словом выбранной Про является *управление*.

**Первый этап** (подготовительный) реализуется следующей последовательностью шагов (см. рисунок 4).

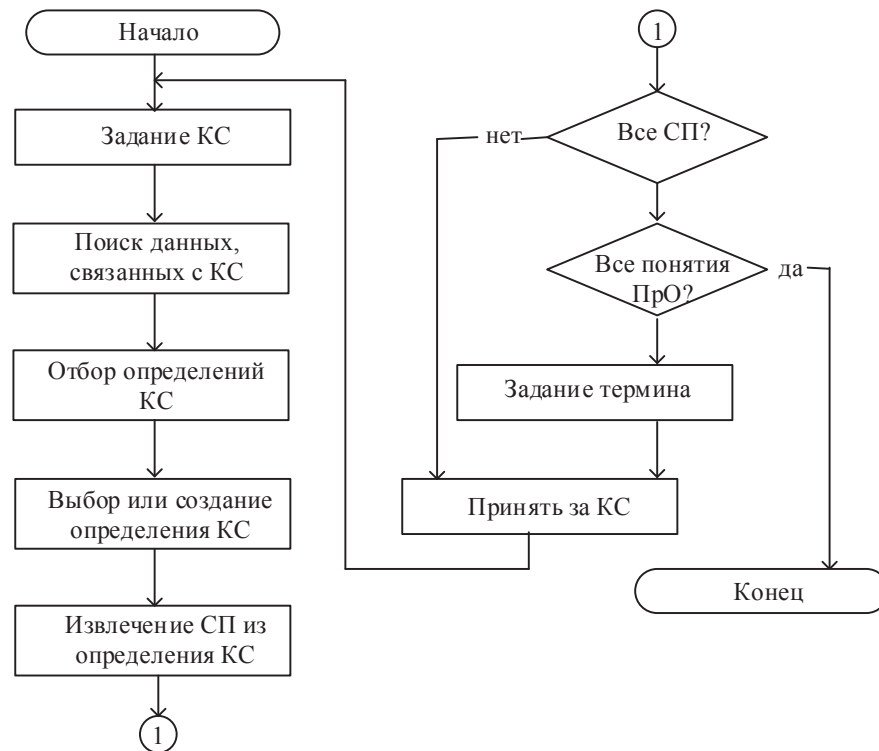


Рисунок 4 – Блок-схема алгоритма создания модели данных

- 1) Задание ключевого слова (КС) для поиска связанной с ним информации. В примере качестве начального ключевого слова принимается *управление*.
- 2) Осуществляется поиск информации по ключевому слову. Результат поиска – избыточная информация (сырые данные), связанная со словом *управление*.
- 3) Исключается вся информация, не относящаяся к определениям ключевого слова. Отсев ненужной информации можно совместить с поиском слова *управление*, конкретизируя связанную с ним информацию следующим образом: *определение понятия управление*. Результат поиска – обработанная информация в виде определений понятия *управление*.
- 4) Осуществляется выбор приемлемого определения понятия *управление* по критерию наибольшей общности. Если ни одно из определений не отвечает этому критерию, формулируется новое определение, содержащее существенные признаки всех частных определений. В [17] обобщённое определение понятия *управление* сформулировано как *выработка и осуществление воздействий одного объекта (субъекта управления) на другой объект (объект управления), предназначенных для поддержания его устойчивости и развития*.
- 5) Из принятого определения понятия извлекаются существенные признаки. В определении понятия *управление* ими являются *выработка воздействия, субъект управления, объект управления, устойчивость, развитие*.
- 6) Пока не сформулированы определения для всех существенных признаков из определения КС, очередной существенный признак (СП) принимается за КС для поиска его определений. Перейти к пункту 1.
- 7) Процесс расширения понятий ПрО продолжается до достижения ими границ ПрО, определяемых разработчиком модели. В качестве КС для поиска определений назначаются свойства, характеризующие понятие *управление*, такие например, как *управляемость* и *наблюдаемость*. Процедура логического вывода родственных понятий может дополняться понятиями, например, *система*, раскрываемое через её вещественные, энергетические

и информационные свойства, *качество*, которым оценивается система и т.д. Пока не определены все понятия ПрО, назначить в качестве КС исследуемый термин и перейти к пункту 1.

Сформированное множество понятий представляет собой состав модели знания.

**Второй этап** проектирования онтологической модели содержит выявление связей между понятиями, вошедшими в состав модели. Установленные логические связи между понятиями нельзя принять за окончательное отношение на множестве понятий. Определена лишь их принадлежность изучаемой ПрО, что характеризуется отношением *элемент-класс*.

Между понятиями онтологической модели устанавливаются следующие типы связи: логические (род-вид, целое-часть, объект-атрибут, причина-следствие), предметные, качественные и количественные связи. Количественные оценки характеризуют степень зависимости элементов модели в статических моделях и интенсивность обмена информацией (энергией, веществом) в динамических моделях. Выявление типа связи осуществляется парным сопоставлением понятий модели. Модель с установленными связями между её элементами представляет собой знание «как устроено».

Онтологическое моделирование заключается в обходе вершин семантической сети. Знание «как действует» в онтологической модели представляет собой совокупность правил вывода и организацию их применения. В ряде случаев оно оказывается очевидным на интуитивном уровне. Простейшим примером применения правил вывода в семантической сети является нахождение причинно-следственной цепочки факторов, последовательная активизация которых приводит к росту или уменьшению выходного фактора.

Знание «как действует» ещё не имеет практического воплощения. Оно проявляется в умении воспользоваться моделью полученного знания, что требует освоения среды моделирования. Если ею является мозг пользователя, то требуется ознакомление с моделью и правилами вывода. Если это программная система, необходимо ознакомиться с её языком и представить модель знания на языке среды моделирования, на котором и осуществляется процесс моделирования.

**Третий этап** познавательного процесса заключается в сопоставлении полученной модели знания с имеющимся знанием (понимание) и в определении области его применения (мудрость). При положительном исходе сопоставления новая модель знания включается в состав существующих моделей. При отрицательном исходе сопоставления принимается решение о поиске другой модели знания, согласующейся с текущей моделью внешней среды.

#### 4 Моделирование процесса познания

В процессе познания с применением предложенных типов моделей следует отделить поэтапно формируемую модель знания об объекте познания (ОП) от знания, используемого для создания этой модели. Назовём это знание *инструментальным* (сокращённо ИЗ), поскольку оно используется как инструмент для формирования нового знания. Каждая процедура перехода от одного к другому этапу познания требует применения своих типовых моделей. Рассмотрим взаимодействие моделей ИЗ и ОП на каждом этапе познавательного процесса. Если предположить, что процесс познания начинается с добычи данных об ОП и завершается мудростью, как умением правильно применить полученное знание в конкретной обстановке, то моделирование процесса познания должно охватывать как поэтапно преобразуемую модель знания об ОП, так и применяемые на каждом этапе модели ИЗ. Модели ИЗ и среды моделирования, участвующие в формировании модели знания на каждом этапе познания, сведены в таблицу 2, где наименования всех этапов познавательного процесса соответствуют рисунку 3.

Таблица 2 - Этапы моделирования познавательного процесса

№	Наименование этапа	Модели инструментального знания	Модель знания	Модели среды моделирования
1	Сырые данные	СФ- и О-модели поиска источника данных	$\langle A_{сд}, \emptyset \rangle$	СФ-модель хранения, О-модели загрузки и запроса данных в БД
2	Обработанные данные	О-модель подготовки данных	$\langle A_{од}, \emptyset \rangle$	СФ-модель хранения, О-модели загрузки и запроса данных в БД
3	Информация	Ф-модели классов, О-модель классифицирования	$\langle A_{сд}, F \rangle$ $A_k/R_{кл}$	СФ-модель хранения, О-модели загрузки и запроса данных в БД
4	Знание «как устроено»	С-модели отношений, О-модели структурирования	$\langle A, R_c, R_{f_1}, R_{f_2}, C, F \rangle$	СФ-модель базы знаний
5	Знание «как действует»	О-модель решения	$\langle X, Y, Q, f_{п}, f_{в}, t \rangle$	О-модели вывода знания
6	Умение	О-модель анализа среды моделирования	СФОМ-модель	М- и О-модели среды моделирования
7	Понимание	О-модель анализа непротиворечивости	Подтверждённая СФО-модель	Аналогичные СФ-модели базы знаний
8	Мудрость	О-модель анализа соответствия	$\langle \text{СФОМ}, M \rangle$	М- и О-модели внешней среды

- 1) **Сырые данные** являются результатом добычи данных из найденного источника данных [18, 19]. На этом этапе в качестве ИЗ привлекаются многоуровневые СФ- и О-модели поиска источника данных. Применительно к приведённому примеру эти модели представляют каталоги литературных источников и способы поиска нужной информации с применением соответствующих О-моделей. В общем случае результатом поиска является начальная модель знания  $\langle A_{сд}, \emptyset \rangle$ , где  $A_{сд}$  представляет собой множество сырых (необработанных) данных, например, слов изучаемой ПрО. При машинной реализации познавательного процесса данные фиксируются в базе данных (БД), для чего проектируются ER-модель (СФ-модель) и О-модели процедур загрузки и запроса данных.
- 2) **Обработанные данные**. Сырые данные подвергаются предварительной обработке. Например, она может состоять в выделении из найденных данных определений очередного термина ПрО. Получаемая модель предметного знания  $\langle A_{од}, \emptyset \rangle$  содержит множество обработанных данных  $A_{од}$ . Предварительная обработка данных может предшествовать их фиксации в БД с целью экономии её объёма.
- 3) **Информация**. Под информацией, требуемой для создания онтологической модели, подразумеваются определения всех включённых в модель понятий. На этом этапе формируется определение, содержащее существенные признаки всех частных определений каждого понятия. Эта операция соответствует формированию класса определений обобщённого понятия. Следовательно, ИЗ этого этапа должно содержать Ф-модели понятий-классов. Процедура классифицирования (формирования классов) выполняется с применением отношения классификации  $R_{кл}$  (С-модели на множестве  $A_{од}$ ) и О-модели классифицирования. Результатом этого процесса является фактор-множество  $A_{од}/R_{кл}$ . Элементы каждого класса по отношению  $R_{кл}$  представляют собой множество:  $A_k/R_{кл} = \{x/R_{кл}: x \in A_k\}$ ,  $k = \overline{1, m}$ . По терминологии [1]  $k$ -й класс  $A_k/R_{кл}$ ,  $k = \overline{1, m}$ , представляет собой *состав* (нуль-граф функций элементов)  $k$ -й модели знания.

- 4) **Знание «как устроено».** Для структурирования  $k$ -й модели знания используются  $S$ -модели типов отношений на множествах данных. На основе  $F$ -моделей классов и  $O$ -модели структурирования создаётся  $SF$ -модель знания. Для определения взаимосвязи понятий на основе численных данных используется корреляционная матрица. Высокое значение коэффициента корреляции указывает на необходимость анализа на предмет возможной содержательной взаимосвязи понятий. При её наличии формируется окончательная  $SF$ -модель  $PrO$ .
- 5) **Знание «как действует».** Разрешимость  $SF$ -модели  $PrO$  обеспечивается разработкой  $O$ -модели. Полученная  $SFO$ -модель представляет собой модель знания  $PrO$ .
- 6) **Умение** использование  $SFO$ -модели полученного знания требует освоения  $M$ -модели среды моделирования. С применением  $O$ -модели анализа среды моделирования  $SFO$ -модель нового знания представляется и решается на языке среды моделирования.
- 7) **Понимание.** Для подтверждения правильности полученной модели знания привлекаются  $SF$ - и  $O$ -модели аналогичного знания. При непротиворечивом исходе сопоставления  $SFO$ -модели нового знания с имеющимся знанием, выполненного с применением инструментальной  $O$ -модели,  $SFO$ -модель включается в состав имеющегося знания.
- 8) **Мудрость.** Для оценки применимости нового знания привлекаются  $M$ - и  $O$ -модели внешней среды. По положительному результату поиска  $M$ -модели внешней среды, соответствующей  $SFO$ -модели, принимается решение о возможности её применения.

## Заключение

На основе модели языка предикатов первого порядка формализованы свойства, присущие любой модели. Типы моделей, реализующие эти свойства по отдельности и в совокупности, образуют базис формальных моделей, необходимый и достаточный для формализации процесса моделирования. За основу моделирования познавательного процесса принята «пирамида познания» Р. Акоффа. Процесс получения нового знания разделён на три этапа: формирование модели данных (состава модели знания), создание модели знания и сопоставление полученной модели знания с имеющимся знанием об аналогичных моделях и внешней среде. Введено понятие инструментального знания, которое используется для выполнения всех этапов познавательного процесса.

## Благодарности

Автор выражает благодарность члену-корреспонденту РАН Р.М. Юсупову за предоставленные материалы по теме и полезные советы. Исследования, выполненные по данной тематике, проводились при финансовой поддержке гранта РФФИ № 17-01-00139 в рамках бюджетной темы № 0073–2018–0003.

## Список источников

- [1] **Тарасенко, Ф.П.** Моделирование и феномен человека. Часть I. Моделирование – инфраструктура взаимодействия человека с реальностью. – М.: Научные технологии, 2012. – 136 с.
- [2] **Акофф, Р.** Логико-лингвистические модели в системах управления / Р. Акофф, Д. Гринберг. – М. Энергия, 1981. – 231 с.
- [3] **Колмогоров, А.Н.** Введение в математическую логику / А.Н. Колмогоров, А.Г. Драгалин. – М.: Изд-во МГУ, 1962. – 120 с.
- [4] **Курош, А.Г.** Лекции по общей алгебре / А.Г. Курош. – СПб.: Лань, 2007. - 560 с.
- [5] **Хованский, А.Г.** Топологическая теория Галуа. Разрешимость и неразрешимость уравнений в конечном виде / А.Г. Хованский. – М.: Изд-во МЦНМО, 2008. – 296 с.

- [6] **Воеводин, В.В.** Параллельные вычисления / В.В. Воеводин, Вл.В. Воеводин. – СПб.: БХВ-Петербург, 2002. – 608 с.
- [7] **Тарасенко, Ф.П.** Прикладной системный анализ: Наука и искусство решения проблем / Ф.П. Тарасенко. – Томск: изд-во ТГУ, 2004. – 186 с.
- [8] **Акофф, Р.** Преобразование образования / Р. Акофф, Д. Гринберг. – Томск: Изд-во ТГУ, 2009. – 196 с.
- [9] **Гельвеций, К.А.** Сочинения / К.А. Гельвеций // В 2-х томах. Составление и общая редакция Х.Н. Момджяна. – М.: Мысль, 1973/1974. – 647/687 с.
- [10] **Новиков, А.М.** Методология: Словарь системы основных понятий / А.М. Новиков, Д.А. Новиков. – М.: Книжный дом ЛИБРОКОМ, 2013. – 208 с.
- [11] **Юсупов, Р.М.** Концептуальные и научно-методологические основы информатизации / Р.М. Юсупов, В.П. Заболотский. – СПб.: Наука, 2009. – 542 с.
- [12] **Исаев, Г.Н.** Моделирование информационных ресурсов: теория и решение задач. / Г.Н. Исаев. – М.: ИНФРА-М, 2010. – 224 с.
- [13] ГОСТ 33707-2016 (ISO/IEC 2382:2015) Информационные технологии (ИТ). Словарь.
- [14] **Микони, С.В.** Теория принятия управленческих решений / С.В. Микони. – СПб.: Лань, 2015. – 448 с.
- [15] **Ивин, А.А.** Словарь по логике / А.А. Ивин, А.Л. Никифоров. – М.: Гуманит. изд. центр ВЛАДОС, 1997. – 384 с.
- [16] **Микони, С.В.** О качестве онтологических моделей / С.В. Микони // Онтология проектирования. – 2017. – Т. 7, №3(25). – С. 347-360. – DOI: 10.18287/2223-9537-2017-7-3-347-360.
- [17] **Микони, С.В.** Формализация определений понятий как условие повышения качества содержательных моделей / С.В. Микони // Сборник докладов XX Международ. конф. по мягким вычислениям и измерениям (SCM-2017). 24-26.05.2017. – СПб.: СПбГЭТУ (ЛЭТИ). 2017. – С. 19-22.
- [18] **Chen, M.S.** Data mining: an overview from a database perspective. / M.S. Chen, J. Han, P.S. Yu // Knowledge and Data Engineering. IEEE Transactions – 1996. – 8 (6). – P. 866-883.
- [19] **Пятецкий-Шапиро, Г.** Data Mining и перегрузка информацией / Г. Пятецкий-Шапиро // Вступительная статья к книге: Анализ данных и процессов / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. // 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.

## FORMALIZATION OF THE COGNITIVE PROCESS USING THE BASIS OF MODELS

S.V. Mikoni

*St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russia  
smikoni@mail.ru*

### Abstract

All human mental activity is a modeling of its external and internal world and is based on the application of models. The cognitive and constructive activity of man can't be imagined through the immense variety of particular models he uses. The paper proposes the structural, functional and operational components inherent in different degrees of any solvable model. The universality of the components is explained by their belonging to the model of the predicate language of the first order, which is the basis of any mathematical models. These basic properties of the model are used to describe the model of knowledge acquired by a person in the process of understanding the surrounding and internal environment. The cognitive process itself is also carried out with the use of models called models of instrumental knowledge. As a basic model of the cognitive process, the "pyramid of cognition" of Ackoff is adopted. The concepts used in it are defined. Each stage of the formation of a model of new knowledge and the instrumental knowledge used to acquire it is described using basic models, which allows us to represent the cognitive process in terms of modeling.

**Key words:** *model, model properties, modeling, cognition, ontological model, knowledge, skill, data, information.*

**Citation:** *Mikoni SV. Formalization of the cognitive process using the basis of models [In Russian]. Ontology of designing. 2018; 8(1): 35-48. - DOI: 10.18287/2223-9537-2018-8-1-35-48.*

## Acknowledgment

The author is grateful to Corresponding Member of the Russian Academy of Sciences R.M. Yusupov for providing materials on the topic and useful advice. The studies carried out on this subject were carried out with the financial support of the RFBR grant 17-01-00139 within the framework of the budget theme 0073-2018-0003.

## References

- [1] *Tarasenko FP*. Modeling and the phenomenon of man. Part I. Modeling - the infrastructure of human interaction with reality [In Russian]. – Moscow: Nauchnye tekhnologii, 2012.
- [2] *Akoff R, Grinberg D*. Logic-linguistic models in control systems [In Russian]. – Moscow: Ehnergiya, 1981.
- [3] *Kolmogorov AN, Dragalin AG*. Introduction to mathematical logic [In Russian]. – Moscow: Publishing house MSU, 1962.
- [4] *Kurosh AG*. Lectures on general algebra [In Russian]. – SPb.: Lan', 2007.
- [5] *Hovanskij AG*. Topological Galois theory. Solvability and insolvability of equations in a finite form [In Russian]. – Moscow: Izd-vo MCNMO, 2008.
- [6] *Voevodin VV, VoevodinVIV*. Parallel Computing [In Russian]. – SPb.: BHV-Peterburg, 2002.
- [7] *Tarasenko FP*. Applied Systems Analysis: Science and the Art of Problem Solving: A Textbook [In Russian]. – Tomsk: Publishing house TSU, 2004.
- [8] *Akoff R, Grinberg D*. Transformation of education [In Russian]. – Tomsk Publishing house TSU, 2009.
- [9] *Gel'vecij KA*. Compositions [In Russian]. In two volumes. Compilation and general revision X.N. Momdzhyan. – Moscow: Mysl', 1973/1974.
- [10] *Novikov AM, Novikov DA*. Methodology: Dictionary of the system of basic concepts [In Russian]. – Moscow: Publishing house LIBROKOM, 2013.
- [11] *Yusupov RM, Zabolotskij VP*. Conceptual and scientific-methodological bases of informatization [In Russian]. – SPb.: Nauka. 2009.
- [12] *Isaev GN*. Modeling Information Resources: Theory and Problem Solving [In Russian]. – Moscow: INFRA-M, 2010.
- [13] GOST 33707-2016 (ISO/IEC 2382:2015) Information technology (IT). Dictionary [In Russian].
- [14] *Mikoni SV*. Theory of administrative decision making [In Russian]. – SPb.: Lan', 2015.
- [15] *Ivin AA, Nikiforov AL*. Dictionary of logic [In Russian]. – Moscow: Gumanit. izd. centr VLADOS, 1997.
- [16] *Mikoni SV*. On the quality of ontological models [In Russian]. *Ontology of Designing*. 2017; 7(3): 347-360. DOI: 10.18287/2223-9537-2017-7-3-347-360.
- [17] *Mikoni SV*. Formalization of definitions of concepts as a condition for improving the quality of content models [In Russian]. *Proc. XX Int. Conf. on Soft Computing and Measurements SCM-2017, 24-26.05.2017*. – SPb.: SPbGETU (LETI). 2017: 19-22.
- [18] *Chen MS, Han J, Yu PS*. Data mining: an overview from a database perspective. *Knowledge and data Engineering/ IEEE Transactions*. 1996; 8 (6): 866-883.
- [19] *Pyateckij-Shapiro G*. Data Mining and Information Overload [In Russian] / Introductory article to the book: Barsegyan AA, Kupriyanov MS, Holod II, Tess MD, Elizarov SI. *Data and Process Analysis [In Russian]*. – SPb.: BHV-Peterburg, 2009.

---

## Сведения об авторе



*Микони Станислав Витальевич*, 1936 г. рождения. Окончил Ленинградский институт инженеров железнодорожного транспорта им. Образцова в 1963 г., д.т.н. (1992), профессор (1994), ведущий научный сотрудник Санкт-Петербургского института информатики и автоматизации РАН. Член Российской ассоциации искусственного интеллекта (1998). В списке публикаций более 290 работ, из них 2 монографии и 7 учебных пособий в области технической диагностики, дискретной математики, системного анализа, теории принятия решений, искусственного интеллекта.

*Stanislav Vitalievich Mikoni* (b. 1936) graduated from the Obraztsov Institute of Engineers of Railway Transport (Leningrad) in 1963, D. Sc. Eng. (1992). Professor (1994). He is Russian Association of Artificial Intelligence member (1998). He is author and co-author of more than 290 publications in the field of technical diagnostic, discrete mathematic, system analyses, artificial intelligence, decision making theory.



УДК 821.161.1

## АВТОРСКИЙ ТЕРМИН: К ОПРЕДЕЛЕНИЮ ПОНЯТИЯ

Ю.В. Сложеникина<sup>1</sup>, А.В. Растягаев<sup>2</sup>, И.Ю. Кухно<sup>3</sup>

Московский городской педагогический университет (Самарский филиал), Самара, Россия

<sup>1</sup>goldword@mai.ru, <sup>2</sup>avr67@yandex.ru, <sup>3</sup>irina.kuhno@yandex.ru

### Аннотация

Авторы статьи констатируют, что в современном отечественном терминоведении отсутствует понятие и дефиниция авторского термина. Традиционной является точка зрения, что авторский термин - это термин, авторство которого достоверно известно либо это термин, названный в честь первооткрывателя научного явления. Авторы статьи считают данный подход ошибочным и предлагают собственную гипотезу: авторский термин существует только в пределах оригинальной научной гипотезы. Эта гипотеза описывается системой взаимосвязанных понятий и терминов, поэтому изолированного авторского термина не существует. Авторы статьи предлагают собственную дефиницию авторского термина. Авторский термин - созданный в рамках креативной научной концепции специальный знак, элемент системы понятий, взаимосвязанный с другими терминами и понятиями оригинальной гипотезы, соотносимый носителями языка для специальных целей с конкретной авторской идеей, необщепринятой научной картиной мира.

**Ключевые слова:** терминология, терминологическое поле, авторский термин, авторская интенция, научная теория.

**Цитирование:** Сложеникина, Ю.В. Авторский термин: к определению понятия / Ю.В. Сложеникина, А.В. Растягаев, И.Ю. Кухно // Онтология проектирования. – 2018. – Т.8, №1(27). – С.49-57. – DOI: 10.18287/2223-9537-2018-8-1-49-57.

### Введение

В 1961 г. в сборнике «Вопросы терминологии» была опубликована фундаментальная статья А.А. Реформатского «Что такое термин и терминология» [1]. Исходя из этимологического значения слова термин - «граница, предел», автор отграничил от свойств и признаков слов языка свойства и признаки слов науки, которые счёл возможным рассматривать в рамках терминологического поля. Основное когнитивное отличие терминов - это принципиальная связь с системой понятий какой-либо науки. Термины, по А.А. Реформатскому, имеют прямое отношение к потребностям теоретической мысли. Их количество в любой науке или научном направлении исчислимо, они служат для словесного отражения системы понятий данной науки.

Для термина важен не контекст употребления, а терминологическое поле, внутри которого термин однозначен. «Один и тот же» термин в разных терминологических полях - суть омоним, так как связи фиксируемого термином понятия в разных науках различны. Эти разные «координатные сетки» позволяют говорить об идиоматичности и парадигматичности терминологического поля каждой науки или научного направления. Если известен «терминологический ключ», то термин может жить самостоятельной жизнью и быть понятным вне контекста [1].

Взгляды А.А. Реформатского легли в основу современного понимания сущности термина, характеризующегося тремя основными признаками:

1) термин обозначает специальное понятие;

2) дефиниция термина служит эквивалентом его понятийного значения, термин и дефиниция взаимозаменяемы;

3) термин системен, за термином стоит не одно понятие, а вся совокупность взаимосвязанных понятий.

Дефиниция отражает его системные связи и место в терминосистеме [2].

## 1 История определения и представления авторского термина в терминоведении

Что такое авторский термин и терминология? А.П. Дьяченко в аннотации «Словаря авторских терминов, понятий и названий» в обоснование словника из 3700 терминов пишет: «В словаре приведены лишь только те термины, авторство которых достоверно или почти достоверно известно» [3, с.2]. Это значит, что авторским термином лексикограф считает специальную языковую единицу, вычлененную из текста с установленной атрибуцией. Например, «*абдоминализация сердца* – название в 1957 г. предложил Г.А. Рейнберг. Рейнберг обозначил им своеобразный способ лечения хронической коронарной недостаточности...» [3, с.5]. Такой подход вызывает несколько вопросов.

Во-первых, А.П. Дьяченко помещает в словарь так называемые мемориальные термины, например «*листерии* – название... в 1927 г. предложил Пири в честь английского хирурга, создателя антисептического направления в хирургии Джозефа Листера (1827-1912)» [3, с.164]. К какой стороне словарной статьи применимо понятие «авторский»? К левой, то есть собственно лексеме, называющей первооткрывателя явления? Или к правой, то есть дефиниции, сообщающей имя человека, предложившего обозначение? Такая неопределённость может порождать комические ситуации, как, например, с термином «ослиная сила»<sup>1</sup>.

Во-вторых, например, в словарной статье *литий* сообщается, что «фактически автором названия элемента является шведский химик и минералог Йенс Якоб Берцелиус, который предложил Арфведсону, тоже шведскому химику, открывшему новый элемент, назвать его литионом» [3, с.164]. Кому же должны быть отданы лавры учёного, введшего в химическую науку понятие литий? Его первооткрывателю или автору названия? При таком подходе возникает дилемма: кого мы считаем автором термина? К чему мы применяем понятие авторский? К термину или к явлению? Можем ли мы считать автором термина учёного, предложившего только формальную оболочку для именованного научного факта? Ведь само понятие, соотнесённое с этой формой, было разработано другим исследователем.

В-третьих, многие некогда специальные лексемы в настоящее время детерминологизировались и воспринимаются как общеупотребительные, например *лордоз* (Гиппократ), *лори* (Бюффон), *логарифм* (Нейпир), *невроз* (Куллен), *олигофрения* (Крепелин), *пищевая цепь* (И.И. Крупник), *психоанализ* (Фрейд), *радиус* (Рамус), *разность* (Видман) и др. Насколько для квалификации термина в качестве авторского важен аспект оригинальности? Можно ли считать авторским название, не соотносимое в общественном сознании с конкретной личностью?

Вопрос о соотношении термина и человека имеет продолжение. Если такого чёткого соотношения нет, то как авторские А.П. Дьяченко может квалифицировать (и делает это) термины, имеющие несколько авторов или даже предложенные коллективами и организациями. Например, «*лишай* (lichen) – авторами термина являются народные врачеватели – лечцы – в среде которых он возник» [3, с.165]. «*Ом* – этот термин в 1861 г. предложил Международный конгресс электриков в Париже» [3, с.212]. «*Пиквикский синдром* – название в 1956 г. пред-

---

<sup>1</sup> В 1884 г в английской журнале «Электрикал ревью» предлагалось ввести ещё одну единицу измерения мощности – ослиную силу, составляющую одну треть от лошадиной силы. *Прим. ред.*

ложила группа английских врачей» [3, с.230]. Каким образом развитию языка науки будет способствовать знание, что термин *собачий зуб* использовали «древние ассирийские и вавилонские врачи (VIII–VI в до н.э.)»? [3, с.283].

В-четвёртых, если произошло переименование понятия и новый знак закрепился в общественной коммуникации, будет ли в таком случае термин авторским? Например, Д. Папен предложил термин *развариватель*. В настоящее время этот предмет известен как *скороварка* [3, с.260]. К сожалению, в словаре отсутствует введение и предисловие, где бы автор мог разъяснить свою концепцию авторского термина.

В настоящее время в отечественном терминоведении и терминографии отсутствует однозначное определение авторского термина. Думается, что определить авторский термин можно по аналогии с понятием авторское кино. Известны режиссёры всех фильмов, но только малую часть из этих лент причисляют к категории авторское кино. Авторское кино удовлетворяет ряду критериев:

- 1) это фильм, который полностью делает сам режиссер;
- 2) главное в нём, не игра актёров, операторская работа, костюмы и проч., а авторская интенция, идея создателя;
- 3) режиссер ставит перед собой не коммерческие, популистские цели, а донесение до зрителя своих взглядов, убеждений;
- 4) он не заигрывает с аудиторией, не пытается ей понравиться, будучи уверенным, что сторонников его взглядов будет не много, но именно они обеспечат фильму непреходящий успех, а автору - пребывание в вечности;
- 5) обычно это кино интеллектуальное, в нём много символизма, оно не массовое, не для всякого зрителя, относится к элитарной культуре.

Попытку определения авторского термина с позиций переводчика предприняла В.Д. Табанакова в монографии «Авторский термин: знаю, интерпретирую, перевожу» [4]. Она стратифицирует специальные обозначения, которые считает возможным обозначить как авторские, и отмечает, что «...это термин одного конкретного человека – автора» [4, с.141]. Таковыми могут быть специальные единицы типа *зона Брока*, *зона Вернике*, названные в честь первооткрывателей. «Авторским мы считаем также термин, предложенный конкретным учёным... Авторским в буквальном смысле слова будут также варианты обозначений одного и того же понятия, предлагаемые разными исследователями или разными школами – так называемые «термины-дублиеты»» [4].

Принципиальная мысль В.Д. Табанаковой заключается в следующем: «Авторский термин - это такое специальное понятие (выраженное любой структурой предикативного характера), за которой стоит собственно авторское понимание и собственно авторская интерпретация» [4, с.168]. Поэтому максимально точный перевод авторского термина предусматривает: проведение лингвистического и логико-понятийного анализа, установление объёма и содержания понятия; проведение контекстуального и дискурсивного анализа, вскрытие авторской прагматики термина [4].

## **2 Дедуктивные размышления, предварительное формулирование дефиниции**

Если термин существует только как член системы, возможен ли вообще авторский термин или следует говорить всё-таки об авторской терминологии? Так, представления о совокупностях языковых единиц, объединённых общим признаком, каким-либо компонентом

значения в более или менее научном виде были оформлены уже в лингвистических трудах XIX в., например в работах М.М. Покровского<sup>2</sup>.

В 1931 г. Йост Трир в книге «Немецкая лексика понятийной области интеллектуальных свойств. История языкового поля с древнейших времен до начала 13 в.» опубликовал результаты докторской диссертации, защищённой в 1928 г. («Derdeutsche Wortschatz im Sinnbezirk des Verstandes. Die Geschichte eines sprachlichen Feldes. Von den Anfängen bis zum Beginn des 13. Jahrhunderts»). Именно в этой работе были представлены положения теории семантического поля, разработанной в языкознании к концу третьего десятилетия XX века. Й. Трир предложил семантически ёмкий, образный термин *семантическое поле*, закрепившийся в лексикологии. Идея семантического поля получила дальнейшее развитие в частности в трудах отечественных многих языковедов. Словарная статья «поле» представлена в «Лингвистическом энциклопедическом словаре» [5, с.380-381].

На наш взгляд, термин *семантическое поле* нельзя считать авторским. Заслуга Й. Трира состоит в находке удачного обозначения для идей, уже существовавших в науке. Его терминотворчество стало закономерным итогом развития общелингвистической теории. Отметим, что в языковедении много терминов, авторство которых устанавливается по первоисточникам, закреплено за учёными в энциклопедических лингвистических словарях, но эти метазнаки не воспринимаются филологами как авторские.

Например, по данным «Лингвистического энциклопедического словаря» [5]:

*аблаут* (Я. Гримм, с.9),  
*агглютинация* (Фр. Бопп, с.17),  
*адстрат* (М. Бартоли, с.19),  
*актант* (Л. Теньер, с.22),  
*аморфные языки* (А.В. Шлегель, с.512),  
*валентность* (С.Д. Кацнельсон, с.79),  
*гипонимия* (Дж. Лайонз, с.104),  
*графема* (Б. де Куртенэ, с.117),  
*диглоссия* (Ч.А. Фергюсон, с.136),  
*инкорпорирующие языки*,  
*изолирующие языки* (В. фон Гумбольдт, с.512),  
*конвергенция* (Е.Д. Поливанов, с.234),  
*крылатые слова* (В. Бюхман, с.246),  
*младограмматизм* (Ф. Царнке, с.302),  
*морф* (Ч.Ф. Хоккет, с.311),  
*морфема* (Б. де Куртенэ, с.312),  
*морфологическая аналогия* (Б. де Куртенэ, с.555),  
*морфонема* (Х. Улашин, с.315),  
*нейтрализация* (Н.С. Трубецкой, с.328),  
*ностратические языки* (Х. Педерсен, с.339),

*означаемое, означающее* (Ф. де Соссюр, с.343),  
*ономасиология* (А. Цаунер, с.346),  
*оппозиция* (Н.С. Трубецкой, с.348),  
*опрошение* (В.А. Богородицкий, с.349),  
*переразложение* (В.А. Богородицкий, с.370),  
*прагматика* (Ч.У. Моррис, с.389),  
*пресуппозиция* (Г. Фреге, с.396),  
*речевой акт* (Дж. Остин, с.412),  
*сигнификат* (Ч.У. Моррис, с.444),  
*синхрония, диахрония* (Ф. де Соссюр, с.451),  
*субстрат* (Я. Бредсдорф, с.497),  
*транспозиция* (Ш. Балли, с.519),  
*флективные языки, аффиксальные языки* (Ф. Шлегель, с.511),  
*фонология* (Б. де Куртенэ, с.556),  
*формы словоизменения, формы словообразования* (Ф.Ф. Фортунатов, с.116),  
*фузия* (Э. Сепир, с.563),  
*эсперанто* (Л.Л. Заменгоф, с.594),  
*языковой союз* (Н.С. Трубецкой, с.64) и др.

Совсем по-иному можно представить учение Н. Хомского о порождающей (генеративной, трансформационной) грамматике - психолингвистической теории, предложенной в конце 1950-х гг. [6]. Гипотеза отстаивала принципы, прямо противоположные доминировавшему в это время бихевиористскому подходу к языку. Как новаторские рассматриваются идеи Н. Хомского о структурных правилах для построения предложений и о врождённых лингвистических способностях. Н. Хомский предложил стройную, внутренне непротиворечивую теорию, терминологический аппарат которой вошёл в лингвистический обиход и использу-

<sup>2</sup> Покровский, Михаил Михайлович (1869-1942) - российский и советский филолог-классик, лингвист и литературовед. Профессор Московского университета, академик Академии наук СССР, создатель и первый заведующий кафедрой классической филологии в Московского института философии, литературы и истории имени Н. Г. Чернышевского. *Прим. ред.*

ется современными лингвистами. Предложенная система терминов позволила формализовать синтаксис как науку, сравнимую «по детальности с аппаратом описания морфологии» [7, с.98-99].

В работах «Аспекты теории синтаксиса» [8] и «Язык и мышление» [9] Н. Хомским был введён термин *языковая компетенция*, под которой понималось знание «говорящего-слушающего» о языке. В настоящее время термин оказался востребованным множеством наук: переводоведением, педагогикой, психологией, филологией и др. Он вышел за рамки концепции трансформационной грамматики, в разных науках наполнился отличным от других понятийным содержанием. С 2000-х гг. понятие языковой компетенции стало правовым термином, основой разработки стандартов по изучению иностранных и родного языков в странах Европы. В настоящее время использование термина языковая компетенция «не связано» к научной деятельности Н. Хомского, не ассоциируется с его синтаксической теорией.

Промежуточный вывод: термин, авторство которого известно, не есть авторский термин. Представления об авторском вкладе должны выходить на уровень гипотезы, теории. Учёный рассматривается прежде всего как основоположник оригинальной научной концепции. Эта концепция, в терминологии А.А. Реформатского, ограничивается системой координат, «координатной сеткой», чётко отделяющей креативную концепцию от общепринятой. Авторская теория устанавливает причинно-следственные, иерархические, линейные и прочие отношения, описываемые системой понятий и терминов. Эти термины находятся друг с другом в отношениях выводимости, сопоставленности / противопоставленности, включения и др. Единичный авторский термин может рассматриваться только как компонент авторской терминологии. Изолированного авторского термина не существует. Некий термин, введённый учёным в качестве компонента метаязыка своей научной теории, может покинуть рамки исходного понятийного поля, войти в узус. В таком случае он теряет родство с породившей его гипотезой и перестаёт восприниматься носителями языка для специальных целей как авторский.

### 3 Отграничение похожих явлений: дефинитивная вариантность термина

Как квалифицировать различные подходы к определению разными учёными одного и того же термина? Такая ситуация в лингвистике, например, активно обсуждалась применительно к термину «дискурс», констатировалась множественность его дефиниций. Действительно, осмысление и переосмысление учёными какого-либо научного понятия может актуализировать личностные подходы к толкованию научного факта. Преобладающими на уровне индивидуальных интенций учёного могут становиться вопросы методологии, цели и задачи исследования, различная научная и языковая компетентность, пространственно-временные координаты и проч. Можно предположить, что «...термин, как никакая другая языковая единица, подвержен семантическому варьированию, обусловленному различными исследовательскими установками» [10, с.237]. Традиционно наличие у термина разных дефиниций считают проявлением терминологической многозначности. Точнее говорить о факте семантической, а именно, дефинитивной вариантности: «связывать разные авторские дефиниции с выражением одного общего специального понятия позволяет их сигнификативное единство, общность терминируемого концепта» [10]. Концептуальная общность не позволяет нам квалифицировать индивидуальные толкования одной и той же специальной лексемы как авторские термины.

Противопоставление авторского термина и семантического (дефинитивного) варианта связано ещё с одной проблемой - оригинальностью термина. Варианты определения термина

находятся в пределах одного концепта, они имеют общую интегральную сему и разнятся дифференциальными. С авторским термином дело обстоит иначе - он соотносится с иным понятием. Сравним определения слова «бездна» в общелитературном языке и как термин в теории пассионарности Л.Н. Гумилева.

Общепотребительное значение	Термин Л.Н. Гумилева
Бездна – глубина, кажущаяся неизмеримой, не имеющей дна; пропасть [11, с. 36].	Бездна – пустота или вакуум, не являющийся частью материального мира [12].

В литературном языке видовое слово «бездна» входит в родовое понятие расстояния как гипоним. Как гумилевский термин бездна входит в родовое понятие пространства. Ещё большее семантическое различие имеют дифференциальные семы. В общепотребительном языке видовая сема указывает на отсутствие предела, дна; в термине понятие бездны соотносится с нематериальным миром.

Аналогично: в литературном языке и в терминологии биполярность – это двухполюсность, причём оба полюса равнозначны. У Л.Н. Гумилева биполярность – это возможность развития систем в двух направлениях – к усложнению и упрощению с лимитом в вакууме. Во-первых, у Гумилева биполярность не статическое явление, а процессуальное, во-вторых, полюса неравноценны.

Можно говорить о том, что общепринятый и авторский научные термины, на уровне формы явленные одинаковым словом, на уровне содержания соотносятся с разными понятиями или концептами, принадлежат к разным научным картинам мира.

#### 4 Центральное специальное понятие

Можно утверждать, что оригинальная авторская гипотеза строится вокруг одного или нескольких (немногих) базовых понятий. В нашем исследовании авторским будем считать специальное понятие, несущее основную идею, концепцию, смысловую нагрузку в тексте. При этом авторские термины должны представлять некую совокупность слов, а не единичные высказывания, или окказионализмы. В таком случае они могут быть оформлены лексикографически.

Центральным для пассионарной теории Л.Н. Гумилева является понятие специфически представленного этногенеза. В рамках изучения исторической науки теория этногенеза Л.Н. Гумилева демонстрирует спектр разнообразных проблем: закономерности изменений в этногенезе, присущие любому этносу, единообразие всех процессов этногенеза [13].

В работе [14] суть авторской теории этногенеза раскрывается следующим образом: «...история – сеть природных процессов этногенеза, её субъект – этнос – биосоциальный феномен, общность людей, обусловленная не столько генетически, сколько взаимным тяготением, комплиментарностью, общностью менталитета...». История этноса обусловлена ритмом этногенеза, особенностями ландшафта, культурными традициями, контактами с соседями. В своей теории этногенеза Л.Н. Гумилев показал, что человек не существует вне этноса, вне своего народа. Понятие этногенеза (по А.А. Реформатскому) является терминологическим ключом, с помощью которого можно открыть авторский замысел историка.

## Заключение

Поиск корреляций в трактовке теории процесса создания термина как слова, предназначенного для выражения специальных понятий, приводит к ключевому моменту: в мышлении человека осуществляются два процесса – эвристический и аналитический. Первый обеспечивает отбор информации, релевантной текущей ситуации, второй – формирует суждение на её основе. Остальная информация отторгается, не допускаясь к стадии анализа. Когнитивный подход к исследованию языковых явлений позволяет говорить о том, что человек через свой многообразный опыт познаёт окружающий его мир и структурирует индивидуальную картину мира как опору для разнообразной деятельности, в том числе научной [15-17].

Таким образом, в рамках данного исследования сформулируем дефиницию следующим образом: *авторский термин* - созданный в рамках креативной научной концепции специальный знак, элемент системы понятий, взаимосвязанный с другими терминами и понятиями оригинальной гипотезы, соотносимый носителями языка для специальных целей с конкретной авторской идеей, необщепринятой научной картиной мира.

## Список источников

- [1] *Реформатский, А.А.* Что такое термин и терминология / А.А. Реформатский // Вопросы терминологии: материалы Всесоюзного терминологического совещания. – М.: Изд-во АН СССР, 1961. – С. 46-54.
- [2] *Сложеникина, Ю.В.* Основы терминологии: Лингвистические аспекты теории термина / Ю.В. Сложеникина. – 2-е изд. – М.: Либроком, 2016. – 120 с.
- [3] *Дьяченко, А.П.* Словарь авторских терминов, понятий и названий / А.П. Дьяченко. – М.: Академкнига, 2003. – 375 с.
- [4] *Табанаква, В.Д.* Авторский термин: знаю, интерпретирую, перевожу / В.Д. Табанаква. – Тюмень: Изд-во ТюмГУ, 2013. – 208 с.
- [5] Лингвистический энциклопедический словарь / под ред. В.Н. Ярцевой. – М.: Советская энциклопедия, 1990. – 685 с.
- [6] *Chomsky N.* Syntactic Structures. Second edition (First edition published in 1957). — Berlin - New York: Mouton de Gruyter, 2002. — 118 p. — ISBN 3-11-017279-8.
- [7] *Кибрик, А.Е.* Генеративная лингвистика / А.Е. Кибрик // Лингвистический энциклопедический словарь / под ред. В.Н. Ярцевой. – М.: Советская энциклопедия, 1990. – С. 98-99.
- [8] *Хомский, Н.* Аспекты теории синтаксиса / Пер. с англ. под ред. и с предисловием В.А. Звегинцева. — М., Изд-во Моск. ун-та, 1972. — 129 с.
- [9] *Хомский, Н.* Язык и мышление / Пер. с англ. Б.Ю. Городецкого. – М.: Изд. МГУ, 1972. – 123 с.
- [10] *Сложеникина, Ю.В.* Терминологическая вариативность: семантика, форма, функция / Ю.В. Сложеникина. – 2-е изд., испр. – М.: ЛКИ, 2018. – 288 с.
- [11] *Ожегов, С.И.* Толковый словарь русского языка / С.И. Ожегов, Н.Ю. Шведова. – М.: ООО «ИТИ Технологии», 2003. – 944 с.
- [12] *Мичурин, В.А.* Словарь понятий и терминов теории этногенеза Л.Н. Гумилева // Гумилев Л.Н. Этносфера: история людей и история природы. – М.: Прогресс, 1993. – С. 531-540.
- [13] *Гумилев, Л.Н.* Древняя Русь и Великая степь / Л.Н. Гумилев. – М.: Айрис-пресс, 2011. – 735 с.
- [14] *Лобанкова, И.П.* Историко-археологический памятник Аркаим в контексте евразийства и пассионарной теории этногенеза Л.Н. Гумилева / Известия Российского государственного педагогического университета им. А.И. Герцена. -2014. - № 172. - С. 33-44.
- [15] *Залевская, А.А.* Когнитивизм, когнитивная психология, когнитивная наука и когнитивная лингвистика / А.А. Залевская // Когнитивная лингвистика: современное состояние и перспективы развития. – Тамбов, 1998. – С. 6-9.
- [16] *Сложеникина, Ю.В.* Философский диалог об именах: Платон, Сумароков, современное состояние / Ю.В. Сложеникина, А.В. Растягаев // Онтология проектирования. – 2015. – № 4 (18). С. 463-471. – DOI: 10.18287/2223-9537-2015-5-4-463-471.
- [17] *Микони, С.В.* О качестве онтологических моделей / С.В. Микони // Онтология проектирования. – 2017. – Т. 7, №3(25). – С. 347-360. – DOI: 10.18287/2223-9537-2017-7-3-347-360.

## AUTHOR'S TERM: TO DEFINITION OF THE NOTION

Yu.V. Slozhenikina<sup>1</sup>, A.V. Rastyagaev<sup>2</sup>, I.Yu. Kuhno<sup>3</sup>

Moscow City University, Samara brunch, Samara, Russia

<sup>1</sup>goldword@mail.ru, <sup>2</sup>avr67@yandex.ru, <sup>3</sup>irina.kuhno@yandex.ru

### Abstract

The authors of the article state that in modern domestic terminology there is no notion and definition of the author's term. The traditional point of view is that the author's term is a term whose authorship is reliably known or is a term named after the discoverer of a scientific phenomenon. The authors of the article consider this approach to be erroneous and offer their own hypothesis: the author's term exists only within the original scientific hypothesis. This hypothesis is described by a system of interrelated concepts and terms, so there is no isolated authorial term. The authors of the article offer their own definition of the author's term. The author's term is a special sign created in the framework of the creative scientific concept, an element of a system of concepts interconnected in other terms and concepts of the original hypothesis, correlated by native speakers for specific purposes with a specific author's idea, an uncommon scientific picture of the world.

**Key words:** terminology, terminological field, author's term, author's intention, scientific theory.

**Citation:** Slozhenikina YuV, Rastyagaev AV, Kuhno IYu. Author's term: to definition of the notion [In Russian]. *Ontology of designing*. 2018; 8(1): 49-57. - DOI: 10.18287/2223-9537-2018-8-1-49-57.

### References

- [1] **Reformatskiy AA.** What is the term and terminology [In Russian]. Questions of terminology: materials of the All-Union Terminological Meeting. Moscow, 1961: 46-54.
- [2] **Slozhenikina YuV.** Fundamentals of terminology: Linguistic aspects of the theory of term [In Russian]. Moscow, Librocom, 2016.
- [3] **Dyachenko AP.** Dictionary of author's terms, concepts and titles [In Russian]. Moscow, Academic book, 2003.
- [4] **Tabanakova VD.** Author's term: I know, I interpret, translate [In Russian]. Tyumen, Publishing house of Tyumen State University, 2013.
- [5] **Linguistic Encyclopaedic Dictionary/** Ed. V.N. Yartseva [In Russian]. Moscow, Soviet Encyclopedia, 1990.
- [6] **Chomsky N.** Syntactic Structures. Second edition (First edition published in 1957). — Berlin - New York: Mouton de Gruyter, 2002. — 118 p. — ISBN 3-11-017279-8.
- [7] **Kibrik, AE.** Generative Linguistics [In Russian]. Linguistic Encyclopaedic Dictionary / Ed. V.N. Yartseva. Moscow, Soviet Encyclopedia, 1990: 98-99.
- [8] **Chomsky N.** Aspects of the Theory of Syntax. Cambridge, Mass.: M.I.T. Press, 1965. — 251 p.
- [9] **Chomsky N.** Language and mind. Third edition. Cambridge University Press. 2006. — 209 p.
- [10] **Slozhenikina YuV.** Terminological Variability: Semantics, Form, Function. [In Russian]. Moscow, Publishing house "LCI", 2018.
- [11] **Ozhegov SI.** Dictionary of the Russian language [In Russian]. Moscow, ITI Technologies LLC, 2003.
- [12] **Michurin VA.** Dictionary of concepts and terms of the theory of ethnogenesis Gumilev [In Russian] Gumilev L.N. Ethnosphere: the history of people and the history of nature. Moscow, Progress, 1993: 531-540.
- [13] **Gumilev LN.** Ancient Rus and the Great Steppe [In Russian]. Moscow, Iris press, 2011.
- [14] **Lobankova IP.** Historical Archeological Monument of Arkaim in the Context of Eurasianism and the Passionary Theory of Ethnogenesis Gumilev [In Russian]. News of the Russian State Pedagogical University. A.I. Herzen. 2014; 172: 33-44.
- [15] **Zalevskaya AA.** Cognitivism, cognitive psychology, cognitive science and cognitive linguistics [In Russian]. Cognitive Linguistics: Current State and Prospects of Development. Tambov, 1998: 6-9.
- [16] **Slozhenikina YuV, Rastyagaev AV.** Philosophical dialogue about the name of Plato, Sumarokov and current state of problem [In Russian]. *Ontology of designing*. 2015. №4(18): 463-471. - DOI: 10.18287/2223-9537-2015-5-4-463-471.
- [17] **Mikoni SV.** On the quality of ontological models [In Russian]. *Ontology of Designing*. 2017; 7(3): 347-360. DOI: 10.18287/2223-9537-2017-7-3-347-360.



## Сведения об авторах



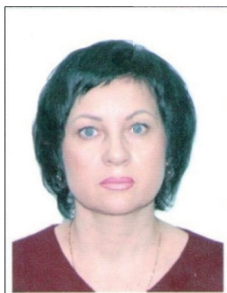
**Сложеникина Юлия Владимировна**, 1971 г. рождения. Окончила Самарский государственный педагогический университет в 1993 г., д.ф.н. (2006). Профессор кафедры филологии и массовых коммуникаций Московского городского педагогического университета (Самарский филиал). В списке научных трудов более 120 работ в области филологии, терминоведения.

**Slozhenikina Yulia Vladimirovna** (b.1971) graduated from the Samara State Teachers' Training University in 1993, Full D (2006). She is a Professor at Moscow City University (Samara brunch). She is co-author of more than 120 scientific articles and abstracts in the field of philology, science of terminology.



**Растягаев Андрей Викторович**, 1967 г. рождения. Окончил Самарский государственный университет в 1993 г., д.ф.н. (2009). Профессор кафедры филологии и массовых коммуникаций Московского городского педагогического университета (Самарский филиал). В списке научных трудов более 140 работ в области филологии, литературоведения.

**Rastyagaev Andrey Victorovich** (b. 1967) graduated from the Samara State University in 1993, FullD (2009). He is a Professor at Moscow City University (Samara brunch). He is co-author of more than 140 scientific articles and abstracts in the field of philology, literary criticisms.



**Кухно Ирина Юрьевна**, 1976 г. рождения. Окончила Самарский государственный педагогический университет в 2008 г., аспирант кафедры филологии и массовых коммуникаций Московского городского педагогического университета (Самарский филиал). Автор семи работ в области терминоведения.

**Kuhno Irina Yuryevna** (b.1976) graduated from the Samara State Teachers' Training University in 2008. She is a graduate student of the Department of Philology and Mass Communications of the Moscow City University (Samara branch). She is co-author of 7 scientific articles in terminology.

УДК 004.02

## ОНТОЛОГИЯ МЕДИЦИНСКОЙ ДИАГНОСТИКИ ДЛЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

В.В. Грибова<sup>1</sup>, М.В. Петряева<sup>2</sup>, Д.Б. Окунь<sup>3</sup>, Е.А. Шалфеева<sup>4</sup>

*Институт автоматизации и процессов управления ДВО РАН, Владивосток, Россия*

<sup>1</sup>gribova@iacp.dvo.ru, <sup>2</sup>margaret@iacp.dvo.ru, <sup>3</sup>okdm@dvo.ru, <sup>4</sup>shalff@iacp.dvo.ru

### Аннотация

Разработана онтология для представления знаний о диагностике заболеваний и синдромов, которая позволяет представлять знания о диагностике широкого круга заболеваний и синдромов. Знания о диагностике, в том числе с учётом его формы, этиологии, патогенеза, варианта течения представляются согласно клинической классификации заболевания в клинически значимой структуре: комплекс диагностических признаков, альтернативные симптомокомплексы с учётом преморбидных биологических, личностных и иных факторов, динамических проявлений признаков и клинических проявлений. При описании симптомов заболевания возможно их ранжирование по специфичности, а также описание условий, предрасполагающих к болезни или способствующих её развитию, что придаёт болезни специфические черты. Работа включает неформальное описание онтологии, а также её модель с описанием основных терминов, знаний, ситуаций и ограничений их целостности в виде онтологических соглашений. В работе приведён пример использования онтологии для формирования базы знаний о диагностике и дифференциальной диагностике заболеваний из группы «Болезни органов пищеварения». Онтология реализована на облачной платформе IASaaS и в настоящее время уже активно используется специалистами для создания баз знаний в различных областях медицины.

**Ключевые слова:** онтология, база знаний, диагностика заболеваний, синдром, симптомокомплекс, медицинские интеллектуальные системы, система поддержки решений.

**Цитирование:** Грибова, В.В. Онтология медицинской диагностики для интеллектуальных систем поддержки принятия решений / В.В. Грибова, М.В. Петряева, Д.Б. Окунь, Е.А. Шалфеева // *Онтология проектирования*. – 2018. – Т. 8, №1(27). – С.58-73. – DOI:10.18287/2223-9537-2018-8-1-58-73.

### Введение

Для принятия диагностических решений врачу необходимо учитывать множество факторов: симптомы и синдромы заболевания, его нозологические формы, этиологии, патогенез, клинические проявления с учётом индивидуальных особенностей пациентов. Всё это держать в памяти и принимать безошибочные и своевременные решения становится всё сложнее. Непрерывно появляются новые знания, а время на принятие врачом соответствующего решения не увеличивается. Как результат, растёт число врачебных ошибок, которые в некоторых странах доходят до 30% [1].

Целью государственной политики в области здравоохранения, определённой Правительством Российской Федерации на период до 2020 года, является повышение качества, безопасности и эффективности медицинских услуг с использованием передовых достижений науки [2, 3]. Качество медицинской помощи, оказываемой населению, зависит уже не только от уровня подготовки (компетентности) врачебного персонала, но и от систем, осуществля-

ющих поддержку решений врача, в том числе для диагностики заболеваний на различных этапах их развития [4].

К настоящему времени создан огромный спектр таких систем, но в повседневную практику врача они внедряются очень медленно. Основные причины этого:

- ориентированность моделей представления знаний на программистов или инженеров знаний и, как следствие, невозможность включить в процесс разработки и последующего сопровождения высококвалифицированных медицинских экспертов как равноправных участников разработки;
- несовместимость систем диагностики между собой, вследствие того, что системы с единым функциональным назначением имеют различные модели знаний, интерфейсы, программные и аппаратные средства реализации.

Решением данных проблем является создание специализированных оболочек, в которых модель представления знаний ориентирована на экспертов, соответствует широкому спектру заболеваний, а не конкретной группе (профилю). На основе такой оболочки могут быть созданы прикладные системы для целого ряда направлений медицины [5].

Целью работы является описание онтологии диагностики, включая дифференциальную диагностику заболеваний и синдромов с учётом динамики их развития, для разработки систем поддержки принятия решений (СППР) в медицине.

## 1 Представление знаний в медицинских интеллектуальных системах

В СППР основную роль играет база знаний, в которой сосредоточены проблемно- и предметно-ориентированные знания. Поскольку медицинское диагностирование является достаточно сложной проблемной областью, за несколько десятилетий накоплен опыт представления медицинских диагностических знаний разными способами: логические модели, продукционные модели, фреймы, семантические сети разных видов (от сетей произвольного вида до сетей, имеющих корневой узел, или близких к деревьям решений), объектно-ориентированные модели понятий и др. [6-16]. Однако СППР, построенные на таких моделях, обычно поддерживают не весь процесс диагностирования, а некоторые этапы, связанные с ним, такие как выдача ответов на запросы к информации (например, «the vital sign with the largest value»), оценка некоторых рисков и т.д. [14, 15]. Семантические сети произвольного вида используются также для визуализации содержимого, написанного на языке OWL, которое предназначено для хранения либо программной обработки данных [14, 16]. Также для визуализации часто используют блок-схемы диагностического процесса, тогда как сами знания этого процесса реализуются как процедурные знания в СППР [6, 17]. В качестве дополняющего способа описания связей между медицинскими сущностями используются таблицы [6, 17].

Сложность медицинского диагностирования связана с необходимостью рассмотрения развития внутренних процессов в организме и их не всегда интуитивно понятных связей (с внешними проявлениями или между собой). Поскольку ограничения некоторых способов не позволяют выразить естественным образом все знания современного диагностического процесса, иногда обращаются к комбинированию нескольких способов представления знаний [11, 12].

Онтология диагностики включает структуру описания информации, и правила её интерпретации и применения для диагностики. С появлением Protégé онтологии чаще стали представлять как иерархии классов понятий и свойств этих понятий [14-16]. Другим перспективным для представления знаний является онтолого-базированный семантический подход, который даёт возможность явно представить онтологию в виде семантической сети понятий с

корневым узлом, циклами и петлями [18, 19]. Онтологии, лежащие в основе большинства известных экспертных систем медицинской диагностики, являются существенно упрощёнными по сравнению с реальными концептуализациями этой области. Обычно в них не рассматриваются развитие патологических процессов во времени и взаимодействие различных типов причинно-следственных отношений.

Одной из первых онтологий медицинской диагностики, близкой к реальным представлениям медицины, была «Онтология медицинской диагностики острых заболеваний» [20]. Она описывает клиническую картину заболеваний в динамике патологического процесса (во времени), а также воздействие лечебных мероприятий и других событий на проявления заболеваний. На основе данной онтологии были разработаны базы знаний по диагностике заболеваний некоторых систем организма: органов дыхания (бронхиальная астма, пневмония), органов пищеварения (язвенная болезнь, острый аппендицит, острый и хронический панкреатит, острый и хронический колит), органов зрения (конъюнктивиты, кератиты, глаукома) и др. [21-23], реализовано несколько программных сервисов. Опыт более чем десятилетнего использования онтологии для формирования баз знаний по диагностике ряда заболеваний позволил накопить и выявить ряд ограничений: невозможность описания клинических проявлений заболевания для различных групп пациентов, игнорирование синдромального подхода к диагностике, бедность значений модальности признаков - обязательность и возможность, невозможность описать альтернативную диагностику и учесть в окончательном диагнозе формы, варианты течения и степень тяжести заболевания.

С учётом указанных факторов актуальным является усовершенствование онтологии диагностических знаний и системы медицинских понятий (т.е. разработка обновленной онтологии), позволяющих формализовать диагностику заболеваний для любого медицинского направления. Цель усовершенствования – приспособиться к современному диагностическому процессу, уровню развития медицинских знаний и к возможности дифференцированного учёта особенностей пациентов при диагностике.

## **2 Основные характеристики онтологии**

Онтология медицинской диагностики должна обладать следующими основными характеристиками.

1. *Формирование симптомокомплексов заболеваний с учётом категорий пользователей.* Использование референсных диапазонов вместо определённых «норм» позволяет описать клинические проявления, данные лабораторных и инструментальных показателей наиболее точно. При оценке результатов обследования различных групп людей становится очевидно, что «нормальные» значения показателя для одной группы не всегда оказываются нормальными для другой. Например, при беременности многие биохимические показатели организма женщины изменяются, поэтому для такой категории определены специальные референсные диапазоны значений этих показателей. У детей и подростков высокий уровень щелочной фосфатазы не только нормален, но и желателен, поскольку у ребенка должны расти здоровые кости. Однако такой же уровень у взрослого человека свидетельствует о заболеваниях: остеопорозе, метастазах опухолей костной ткани. Описание референсных значений (учёт пола, возраста, особенностей профессии, беременности, спорта и т.д.) для большинства клинических проявлений, лабораторных и инструментальных признаков повышают информационную значимость любого симптомокомплекса [24-27].

2. *Формирование клинической картины синдрома.* Широко вошёл во врачебную практику синдромальный подход к диагностике, дополняя классический подход, ориентированный на совокупность специфических и неспецифических симптомов (симптомокомплекс). Син-

дром - это сочетание (группа) симптомов, объединяемое общим патогенезом. В современных условиях синдромный уровень диагностики имеет определённые преимущества, особенно на догоспитальном этапе диагностического процесса. Он играет важную роль в определении нозологической сущности наиболее важных проявлений болезни или её осложнений. Диагноз может быть быстро установлен при наименьшем объёме диагностических исследований, и в то же время он достаточен для обоснования патогенетической терапии или направления больного в стационар для оперативного вмешательства (например, при синдроме острого живота).

3. *Расширенный ряд значений модальности.* Необходимость расширения понятия модальности связана с ранжированием симптомов заболевания по их специфичности. Для высокоспецифичного симптома модальность = «необходимость» (сжимающая боль за грудиной при стенокардии), для специфичного симптома = «характерность» (одышка при хроническом обструктивном бронхите или при бронхиальной астме), если симптом малоспецифичный, модальность = «возможность» (слабость, лихорадка, головная боль – для разных заболеваний).

4. *Единообразная формализация стадий хронических болезней и периодов развития острых заболеваний.* В современных системах должна быть возможность поставить диагноз и провести дифференциальную диагностику с другими заболеваниями в разные периоды развития болезни, анализируя развитие заболевания до обращения к врачу и учитывая, что пациент может обратиться к врачу в разное время от начала заболевания, как в первые часы, так и в момент, когда симптоматика угасает.

5. *Формирование альтернативных симптомокомплексов.* Важно иметь возможность формализовать разные подходы к выявлению достоверных признаков заболевания, чтобы в процессе диагностики выбрать наиболее щадящий, быстрый или недорогой.

6. *Уточнение диагнозов с учётом этиологии, патогенеза, варианта течения* и пр. для проведения дифференциальной диагностики заболеваний и выбора соответствующих методов лечения.

7. *Выделение признаков (и их комплексов) для групп заболеваний* [28]. Такое группирование делает более эффективным процесс поиска гипотез на основе базы знаний.

8. *Определение необходимых условий заболеваний* [29]. Различают условия, предрасполагающие к болезни или способствующие её развитию, препятствующие возникновению болезни и её развитию или модифицирующие, которые видоизменяют действие причинного агента и придают болезни специфические черты. Роль условий при возникновении патологических процессов и болезней различна: она может быть либо решающей, либо незначительной, что в свою очередь и определяет их диагностическую ценность.

9. *Учёт значений характеристик, изменённых воздействием событий* [29]. Наличие такого элемента причинно-следственных связей позволяет учитывать внешние воздействия, оказываемые на организм пациента на разных этапах заболевания. К числу наиболее значимых внешних условий можно отнести: экологические факторы (например, загрязнённый воздух, вода, воздействие на организм вредных промышленных, сельскохозяйственных, бытовых и других факторов); количественную и качественную неполноценность пищи; нарушение упорядоченности и оптимального соотношения труда и активного отдыха; социальные факторы (например, частые конфликтные ситуации) и т.д.

10. *Учёт разных вариантов динамики значений признаков* [29, 30]. Такие элементы знаний позволяют учитывать многообразие течения одних и тех же заболеваний у разных пациентов, представляющее врачебный опыт.

### 3 Описание онтологии медицинской диагностики

Диагностика в клинической медицине - раздел, изучающий содержание, методы и последовательные ступени процесса распознавания заболеваний по её симптомам (признакам заболевания). Обнаружение патогномичного симптома, т.е. встречающегося только при данном заболевании, достаточно для установления достоверного диагноза. Однако число патогномичных симптомов ограничено, поэтому в диагностике большинства заболеваний обычно ориентируются на симптомокомплексы. Новая онтология представляет используемые в медицинской диагностике причинно-следственные связи элементов симптомокомплексов с заболеваниями.

Каждое заболевание представлено альтернативными симптомокомплексами, необходимыми условиями для этого заболевания, и может содержать детализацию соответствующего диагноза. Симптомокомплекс заболевания состоит из комплекса жалоб и объективных исследований, комплекса лабораторных и инструментальных исследований и необходимых условий для симптомокомплекса. Количество симптомокомплексов определяется типом заболевания, необходимостью учёта преморбидных биологических, личностных и иных факторов. Роль применения симптомокомплексов велика в силу того, что позволяет объединить диагностически ценные признаки в «рамках» определённого условия.

Необходимое условие для заболевания обозначает то событие, без которого заболевание не случилось бы, например: укус клеща для клещевого энцефалита, проникающее ранение для травматического кератита.

В комплексах жалоб, объективных исследований, в лабораторных и инструментальных исследованиях представлены множества признаков, изменения значений которых являются симптомами заболевания.

Возможные причины заболевания представлены событиями или этиологическими факторами, которые привели или способствовали развитию заболевания. Они описываются модальностью и временными характеристиками, которые в свою очередь включают интервал до начала заболевания и продолжительность события.

Детализация диагноза - по форме, варианту, степени тяжести, стадии и т.п. - представляет набор признаков (или симптомокомплекс), позволяющий внести соответствующее уточнение к основному диагнозу. Так, например, при формализации заболевания «Острый холецистит» для более детальной (глубокой) диагностики и дифференциальной диагностики необходимо описание заболевания по форме (калькулёзный, некалькулёзный), по варианту течения (катаральный, флегмонозный, гангренозный), по степени тяжести (лёгкая, средней тяжести, тяжёлая).

В модели знаний будут предложения вида «вариант процесса изменения значений признака, характерный для некоторого заболевания», структура которых:

**(I)** <диагноз<sub>j</sub>, {симптомокомплекс<sub>kj</sub>}, [необходимое условие<sub>j</sub>]>;

**(II)** <симптомокомплекс<sub>k</sub>, {признак<sub>j</sub>, диапазон<sub>kj</sub> значений признака<sub>j</sub>}>;

**(IIIa)** <симптомокомплекс<sub>k</sub>, {признак<sub>j</sub>, {период<sub>i</sub>, продолжительность периода<sub>i</sub>, диапазон<sub>ij</sub> значений признака<sub>j</sub> в период<sub>i</sub>}>>.

Это позволяет описывать заболевания с учётом одной из главных сложностей диагностического процесса в медицине - необходимостью определить непрерывно развивающийся процесс (заболевание). Каждая болезнь развивается в течение большего или меньшего времени. С точки зрения быстроты развития болезней различают острейшие – до 4 дней, острые – около 5–14 дней, подострые – 15–40 дней и хронические, длящиеся месяцы и годы. В развитии болезни почти всегда можно различить следующие стадии: 1) начало болезни (иногда его называют латентным периодом); 2) стадия собственно болезни; 3) исход болезни.

Диагностика, как правило, проводится на стадии «собственно болезни». На этой стадии

выделяют следующие периоды развития: 1) период нарастания проявлений болезни; 2) период разгара (максимальной выраженности симптомов); 3) период угасания проявлений болезни (постепенное исчезновение клинической симптоматики).

Онтология медицинской диагностики задаёт структуру описания групп заболеваний, заболеваний, групп синдромов и синдромов.

(III) <группа диагнозов<sub>m</sub>, {(диагноз<sub>k</sub> | группа диагнозов<sub>k</sub>)}>.

Структура описания синдрома состоит из описания синдромов и групп синдромов. Каждая группа синдромов может содержать синдромы этой группы и группы синдромов более низкого уровня. Каждый синдром содержит клиническую картину, которая состоит из множества признаков. Графическое представление фрагмента онтологии представлено на рисунке 1.

В модели знаний могут быть указаны особые условия, необходимые для возникновения некоторого заболевания:

(IV) <диагноз<sub>k</sub>, событие<sub>u</sub> [, временной интервал<sub>ku</sub>]> или

(IVa) <диагноз<sub>k</sub>, событие<sub>u</sub>, временной интервал<sub>ku</sub> [, характеристика<sub>i</sub> события | диапазон значений события]>, если событие характеризуется не только моментом совершения, но и некоторой качественной или количественной характеристикой;

(V) <диагноз<sub>k</sub>, фактор<sub>j</sub> [, область значений фактора<sub>j</sub>]>.

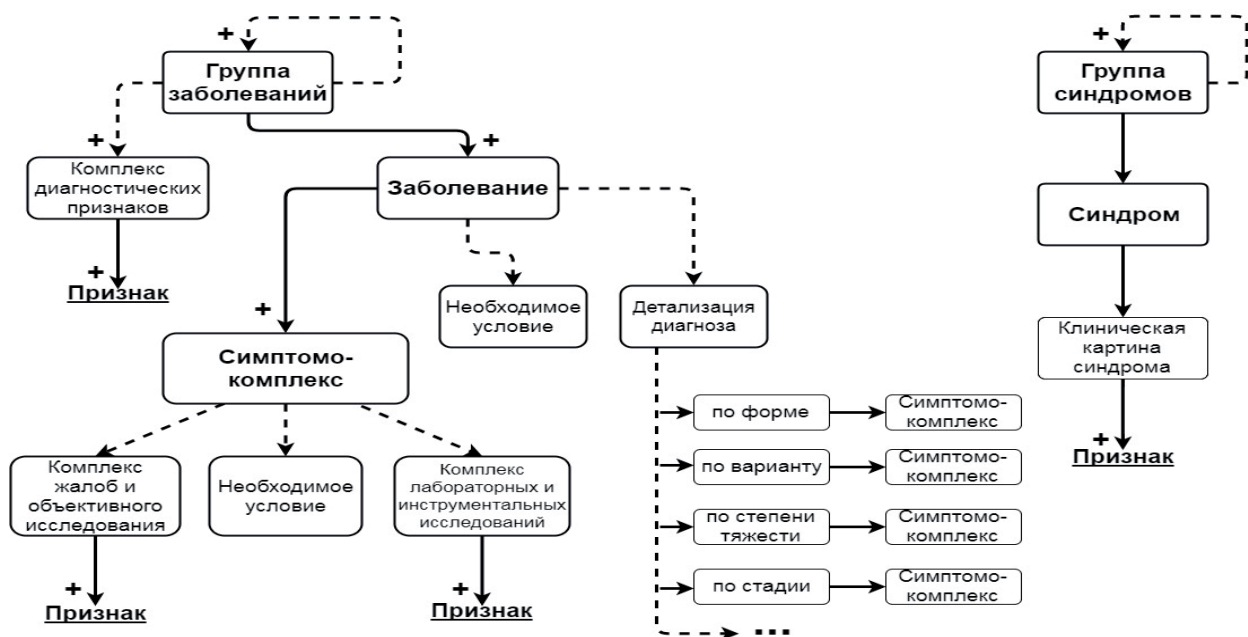


Рисунок 1 – Графическое представление фрагмента онтологии диагностики в форме семантической сети

Для описания изменения *развития* заболевания при воздействии на организм извне нужны предложения вида «вариант реакции процесса функционирования на воздействие события», имеющие следующую структуру:

(VI) <симптомокомплекс<sub>k</sub>, признак<sub>j</sub>, [диапазон значений<sub>i</sub>,] событие<sub>u</sub>, [временной интервал<sub>kju</sub>,] диапазон изменённых значений<sub>i</sub>>, если поэтапность изменений не важна, или

(VIa) <симптомокомплекс<sub>k</sub>, признак<sub>j</sub>, событие<sub>u</sub>, [временной интервал<sub>kju</sub>,] {период<sub>i</sub>, диапазон значений<sub>uk</sub> признака<sub>ji</sub>}>.

Событие может быть составным либо быть совокупностью событий, либо совокупностью события и факторов. Тогда нужны предложения вида «вариант реакции процесса на воздействие совокупности факторов»:

(VII) <симптомокомплекс<sub>k</sub>, признак<sub>j</sub>, [диапазон значений<sub>i</sub>,] событие<sub>u</sub>, количественная характеристика события<sub>u</sub>, [временной интервал<sub>kju</sub>,] {фактор<sub>j</sub>, [область значений фактора<sub>j</sub>,]} интересующий признак<sub>j</sub>, диапазон изменённых значений признака<sub>j</sub>>.

Онтология позволяет описывать заболевания с учётом этиологии, патогенеза, варианта течения, стадии и т.д. путём формирования дополнительных симптомокомплексов для более детальной (глубокой) диагностики или дифференциальной диагностики заболевания. По мере детализации диагноза для каждой единицы патологического процесса описываются свои дополнительные или специфичные симптомы.

Признак заболевания может быть простым или составным (альтернативное (А) представление показано на рисунке 2), его значения представляются по периодам динамики развития признака или заболевания в целом; могут задаваться необходимые условия для рассмотрения признака. Каждый период динамики характеризуется верхней и нижней границей длительности периода, единицей измерения границ.



Рисунок 2 – Графическое представление термина «Признак»

Простой признак имеет модальность (вхождение признака в клиническую картину); в каждом периоде признак может иметь более одного варианта значений. Каждый вариант значений задаёт множество возможных значений признака, необходимые условия для наличия признака, а также описание изменения значения этого признака под воздействием некоторых событий. Примером простого признака являются жалобы больного: жажда, изжога, тошнота, снижение аппетита с вариантами значений «имеется», «отсутствует».

Составной признак содержит описание наборов своих изменяющихся по периодам характеристик (рисунок 2) с модальностью вхождения характеристики признака в клиническую картину. Каждая характеристика может иметь множество собственных характеристик и множество вариантов значения этой характеристики. Примером составного признака является такая жалоба больного, как боль в животе. Этот признак имеет несколько характеристик: характер, локализация, интенсивность, выраженность, иррадиация, периодичность, причина усиления. Каждая характеристика может иметь одно или несколько различных значений. Например, характеристика характер боли (признака боль в животе) имеет значения: острая, тупая, колющая, режущая, пульсирующая, давящая, тянущая. Характеристика локализация имеет значения: правое подреберье, левое подреберье, эпигастрий, правая подвздошная об-



ласть, левая подвздошная область, мезогастрий. Характеристика интенсивность: слабая, умеренная, сильная, резкая, резчайшая.

Каждый вариант значений характеристики содержит множество возможных значений характеристики, необходимые условия для наличия характеристики, а также описание изменения значения этой характеристики под воздействием некоторых событий.

Термин «значение, изменённое воздействием события» позволяет описывать изменение признака в динамике, если после начала развития заболевания до обращения к врачу пациент сам предпринял какие-либо меры, или значения признаков (жалоб, объективного состояния) меняются под воздействием каких-либо событий или манипуляций, предпринятых врачом.

*Пример 1.* У пациента на момент обследования имеется жалоба: острая боль в животе. До обращения к врачу он принимал обезболивающий препарат (анальгин), острая боль в животе стала тупой.

Признак: «боль в животе»;

Характеристика: «характер»;

Значение характеристики: «острая»;

Воздействующее событие: «приём анальгина»;

Значение, изменённое воздействием события (значение характеристики): «тупая».

*Пример 2.* У больного дифтерийным конъюнктивитом имеются отёк, гиперемия, плёнки на слизистой век. После снятия врачом плёнок с конъюнктивы открылось кровотечение.

Признак: «конъюнктивит»;

Характеристика: «изменения на слизистой век»;

Значение характеристики: «отёк, гиперемия, дифтерийные плёнки»;

Воздействующее событие: «снятие плёнок»;

Значение, изменённое воздействием события (значение характеристики): «кровотечение».

Описанная онтология размещена на облачной платформе IASaaS (в формате семантической сети с корневой вершиной) и используется для формализации знаний. На рисунке 3 представлен фрагмент этой онтологии.

В основе любой интеллектуальной медицинской системы, базирующейся на знаниях, лежит база формализованных знаний, от качества и объёма которой напрямую зависит эффективность системы. Для единообразного представления и однозначного толкования баз знаний участниками разработки и пользователями (в том числе из разных учреждений) требуется общий свод всех используемых на практике терминов. Терминология должна быть общепринятой и понятной специалистам в области медицины, т.е. быть результатом онтологического соглашения в области медицины.

Общий свод используемых на практике терминов содержит названия наблюдений, все возможные их значения, а также их распространённые синонимы, употребляемые при заполнении историй болезней (поскольку медицинской терминологии свойственна «подвижность лексического состава» и его постоянное развитие: предусмотрена возможность сохранения национальных и интернациональных терминологических синонимов, дублетов (одышка – диспноэ, кожа – дерма) и частично совпадающих синонимов, введение в название метода или симптома имени его открывателя (симптом раздражения брюшины – симптом Щеткина-Блюмберга, симптом скольжения – симптом «рубашки» – симптом Воскресенского).

В сформированной коллективом лаборатории интеллектуальных систем ИАПУ ДВО РАН «Базе медицинской терминологии и наблюдений» выделены такие группы наблюдений как признаки, события и факторы (со своими подгруппами) [31]. В частности, группами признаков являются традиционные для специалистов *Жалобы*, *Данные объективного исследования*, *Данные лабораторных* и *Данные инструментальных исследований*. В этой базе сгруппированы термины разделов «патогенные факторы», «фармакология», «единицы измерения»

и другие. Такая «База наблюдений» является универсальным ресурсом, применяемым для формирования диагностических (и не только) баз знаний различных областей (профилей) медицины.

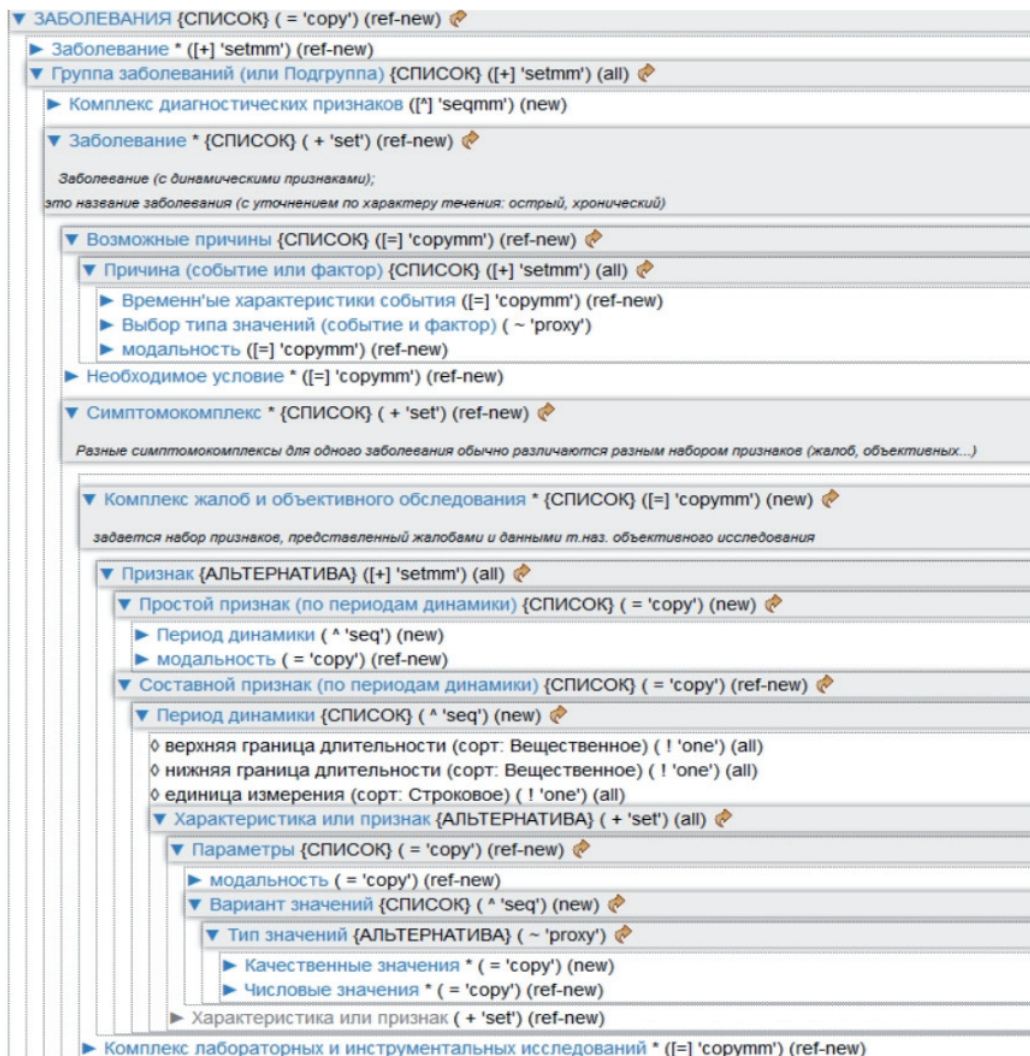


Рисунок 3 – Фрагмент онтологии диагностики, размещённой на облачной платформе IACPaas (скриншот)

#### 4 Использование онтологии для формирования баз медицинских знаний

На основе описанной онтологии, с использованием «Базы медицинской терминологии и наблюдений», был сформирован информационный ресурс на платформе IACPaas (<https://iacraas.dvo.ru/>) «База знаний о диагностике заболеваний и синдромов», включающий описание заболеваний из таких разделов медицины как «Гастроэнтерология», «Инфекционные болезни» и «Урология». В качестве примеров можно привести такие заболевания как: желчнокаменная болезнь, холецистит (острый, хронический) с различными его формами (катаральный, обструктивный, деструктивный...) и возможными вариантами осложнений (перитонит, холангит, околопузырный воспалительный инфильтрат...). На таком заболевании как геморрагическая лихорадка с почечным синдромом отчётливо демонстрируется актуальность периодов течения заболевания: инкубационный период, доолигурический период, олигурический период, полиурия и реконвалесценция со своей специфичностью клинической

картины. Описание знаний о диагностике такого заболевания как простатит с его формами позволит преодолеть современные трудности в ранней его диагностике при «омоложении» патологии и стёртости клинических проявлений.

Группы заболеваний объединяют концептуально связанные заболевания. Например, группа заболеваний «Болезни органов пищеварения» состоит из групп заболеваний – «Болезни пищевода», «Болезни желудка и двенадцатиперстной кишки», «Болезни кишечника», «Болезни печени», «Болезни желчного пузыря, желчевыводящих путей» и др. Каждая группа заболеваний группирует заболевания с общим комплексом диагностических признаков, который включает признаки, характерные для данной группы. Так для группы заболеваний «Болезни органов пищеварения» в комплекс диагностических признаков входит один признак «Боль в животе». Для группы «Болезни жёлчного пузыря и жёлчевыводящих путей» диагностический комплекс включает 5 признаков: «Боль в животе», «Тошнота», «Кожный зуд», «Повышение температуры тела», «Напряжение мышц передней брюшной стенки». Для группы заболеваний «Холециститы» диагностическими являются уже более 20 признаков: всё те же признаки плюс «Метеоризм», «Рвота», «Отрыжка», «Тошнота», «Горечь во рту», «Лейкоциты», «СОЭ», «Толщина стенки желчного пузыря на УЗИ» и т.д. Фрагмент информационного ресурса «База знаний о диагностике заболеваний» представлен на рисунке 4.

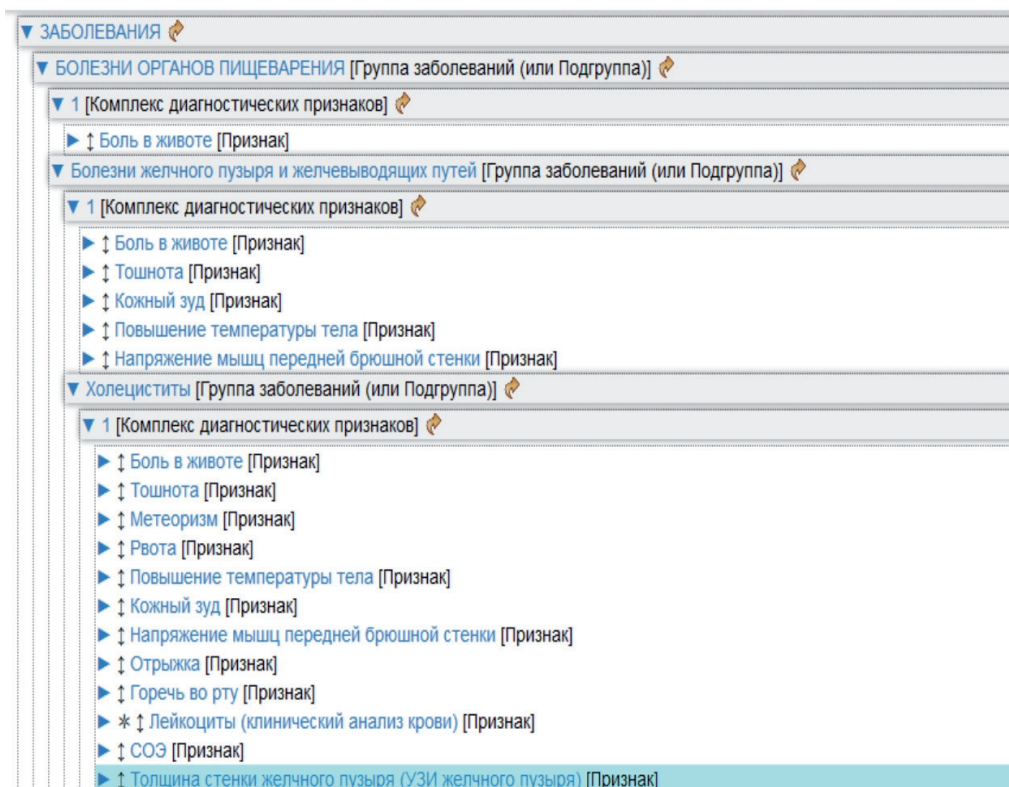


Рисунок 4 – Фрагмент информационного ресурса «Комплексы диагностических признаков групп заболеваний»

Согласно Международной классификации болезней в группу заболеваний «Холециститы» входит заболевание «Острый холецистит». Описание самого заболевания включает описание симптомокомплексов для нескольких возрастных групп пациентов (дети 0-1 год, дети 1-16 лет, взрослые 17-59 и т.д.) и необходимого условия или события, которое привело к возникновению этого заболевания.

Для более детальной (глубокой) диагностики и дифференциальной диагностики описаны формы заболевания (острый калькулёзный холецистит, острый некалькулёзный холецистит)

и варианты (катаральный, флегмонозный, гангренозный), также с описанием симптомокомплекса для каждой формы и варианта заболевания по возрастным группам. Фрагмент представлен на рисунке 5.

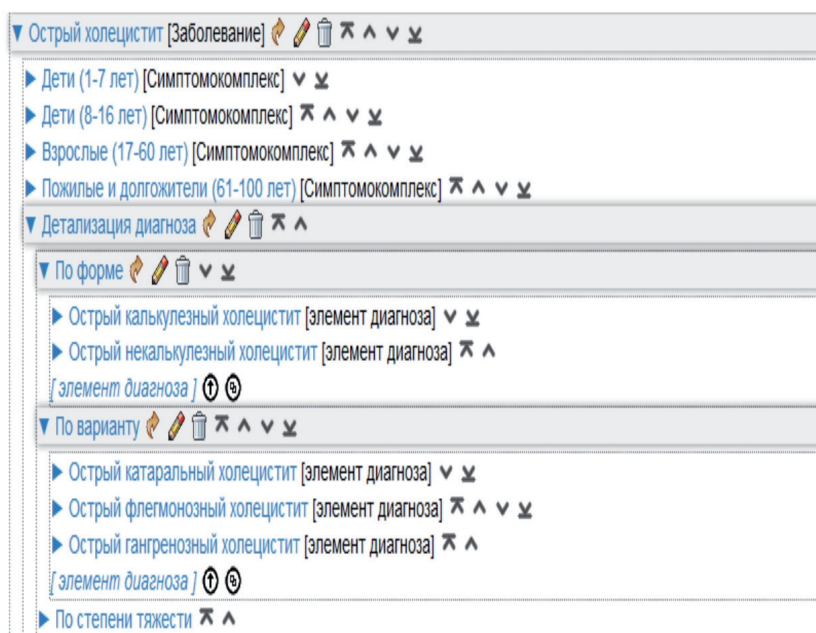


Рисунок 5 – Фрагмент информационного ресурса «Симптомокомплексы заболевания и детализация диагноза»

Каждый симптомокомплекс включает описание комплекса жалоб и объективного исследования и комплекса лабораторных и инструментальных исследований, которые в свою очередь включают описание патогномичных, специфических и неспецифических признаков заболевания с определением модальности (см. рисунок 4). В течение заболевания острый холецистит выделены периоды динамики: первый - 1-6 часов, второй - 7-24 часа, третий 1-7 суток. Онтология даёт возможность описать все варианты значений простых и сложных признаков с характеристиками во все периоды течения заболевания. Фрагмент информационного ресурса «Периоды развития заболевания» представлен на рисунке 6.

## Заключение

Предложенная онтология даёт возможность формировать современные знания о диагностике заболеваний в зависимости от их класса, формы, осложнений, выраженности клинических проявлений с возможностью дифференцированного учёта особенностей пациентов.

Онтология знаний по диагностике заболеваний и синдромов размещена на платформе IASaaS и в настоящее время уже активно используется специалистами для создания баз знаний в различных областях медицины. Сообщество экспертов в области медицины, разработчиков интеллектуальных медицинских систем, заинтересованное в создании баз знаний и их использовании для разработки интеллектуальных медицинских систем, имеют возможность присоединиться к этому процессу (для создания собственных баз, улучшения и развития созданных, а также для их использования в программных системах).

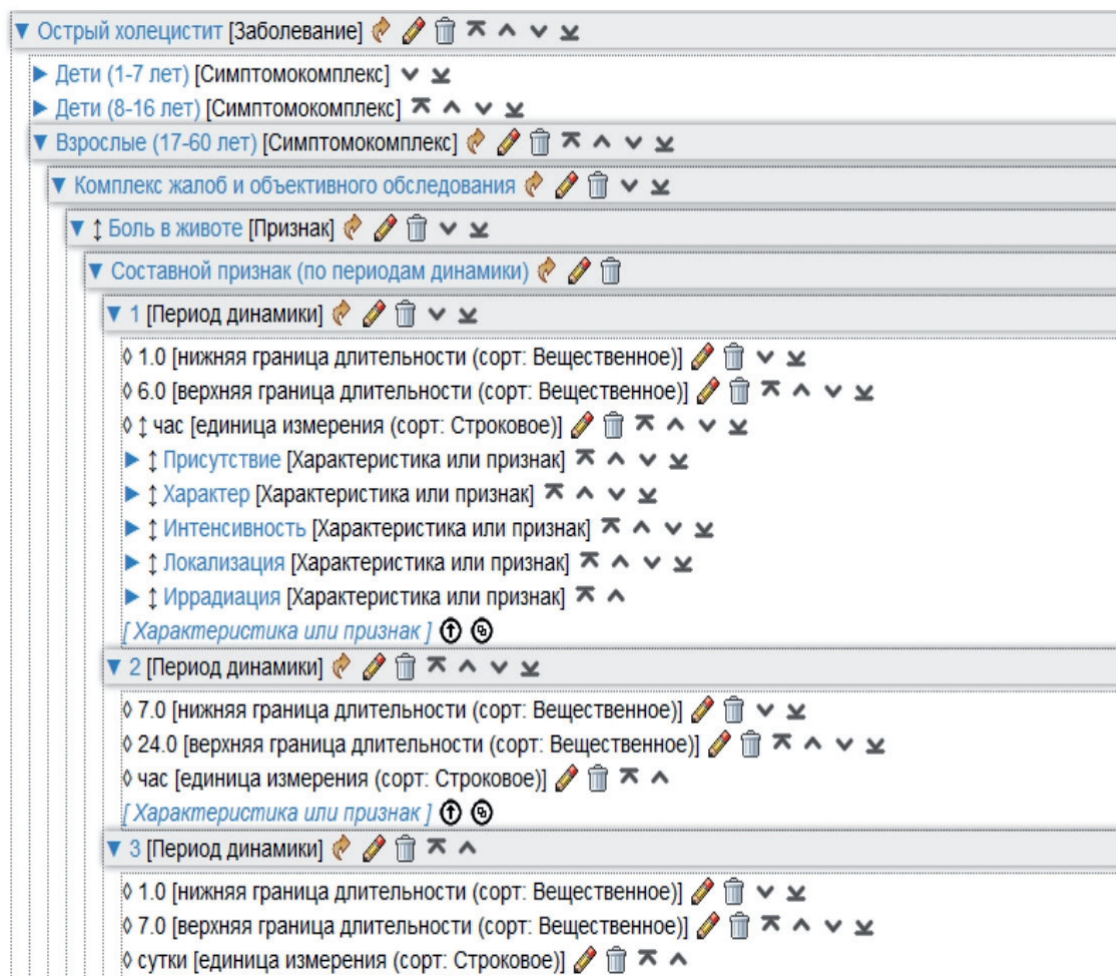


Рисунок 6 – Фрагмент информационного ресурса «Периоды развития заболевания»

## Благодарности

Работа выполнена при частичной финансовой поддержке РФФИ (проекты №№ 18-07-01079, 17-07-00956).

## Список источников

- [1] *Галанова, Г.И.* Врачебные ошибки – проблема не только врача // Менеджер здравоохранение. – 2014. - №8. – С.49-52.
- [2] Проблемные вопросы контроля качества и безопасности медицинской деятельности на современном этапе. - <http://rusmedunion.ru/?p=2222/>.
- [3] Распоряжение Правительства РФ №1662-р от 17.11.08.
- [4] Искусственный интеллект для Медицины – обзор. - [http://www.livemd.ru/tags/iskusstvennyj\\_intellekt/](http://www.livemd.ru/tags/iskusstvennyj_intellekt/).
- [5] *Грибова В.В.* Проект IASaaS. Комплекс для интеллектуальных систем на основе облачных вычислений / В.В. Грибова, А.С. Клещев, Д.А. Крылов и др. // Искусственный интеллект и принятие решений. – 2011. – №1. – С.27–35.
- [6] *Soltan, R.A.* Diagnosis of Some Diseases in Medicine via computerized Experts System / R.A. Soltan, M.Z. Rashad, B.El-Desouky // International Journal of Computer Science & Information Technology - 2013. - Vol. 5(5). - P.79-90.
- [7] *Загоруйко, Ю.А.* Семантическая технология разработки интеллектуальных систем, ориентированная на экспертов предметной области // Онтология проектирования. 2015. № 1 (15). Т.5. - С.30-46.

- [8] **Мельник, К.В.** Проблемы и основные подходы к решению задачи медицинской диагностики / К.В. Мельник, С.И. Ершова // Системы обработки информации, 2011, выпуск 2(92).
- [9] **Ле, Н.В.** Модель представления знаний при создании медицинской экспертной системы дифференциальной диагностики / Н.В. Ле, В.А. Камаев, Д.П. Панченко, О.А. Трушкина // Известия Волгоградского государственного технического университета. - 2014. - № 6. - С.42-50.
- [10] **Ле, Н.В.** Обзор подходов к проектированию медицинской системы дифференциальной диагностики / Н.В. Ле, В.А. Камаев, Д.П. Панченко, О.А. Трушкина // Известия Волгоградского государственного технического университета. 2014. Т. 20, вып. 6. С.50–58.
- [11] **Жаркова, О.С.** Построение систем поддержки принятия решений в медицине на основе деревьев решений / О.С. Жаркова, К.А. Шаропин, А.С. Сеидова, Е.В. Берестнева, И.А. Осадчая // Современные наукоемкие технологии. - 2016. - № 6-1. - С.33-37.
- [12] **Haug, et al** An ontology-driven, diagnostic modeling system // Journal of the American Medical Informatics Association. - 2013. - Т. 20. - №. e1. - С.e102-e110.
- [13] **Patil, R.S.** Modelling Knowledge of the Patient in Acid-base and Electrolyte Disorders / R.S. Patil, P. Szolovits, W.B. Schwartz // in Szolovits, P. (Ed.) Artificial Intelligence in Medicine, - Boulder, CO: Westview Press, 1982. - P.191-226.
- [14] **Uciteli A. et al.** Ontology-based specification, identification and analysis of perioperative risks //Journal of biomedical semantics. - 2017. - V.8. - No.1. - P.36.
- [15] **Yu Lin, Jie Zheng and Yongqun He.** VICO: Ontology-based representation and integrative analysis of Vaccination Informed Consent forms // Journal of Biomedical Semantics (2016) 7:20 – DOI: 10.1186/s13326-016-0062-4.
- [16] **Bouamrane MM., Rector A., Hurrell M.** Using OWL ontologies for adaptive patient information modelling and preoperative clinical decision support // Knowl Inf Syst (2011) 29:405–418. – DOI: 10.1007/s10115-010-0351-7.
- [17] **Abidi S. et al.** Using OWL ontologies for clinical guidelines based comorbid decision support //System Science (HICSS), 2012 45th Hawaii International Conference on. – IEEE, 2012. – С.3030-3038. – DOI: 10.1109/HICSS.2012.629.
- [18] **Gribova V.V., Kleshchev A.S., Moskalenko F.M., Timchenko V.A.** A Two-level Model of Information Units with Complex Structure that Correspond to the Questioning Metaphor // Automatic Documentation and Mathematical Linguistics. 2015. Vol.49. No.5. - P.172-181.
- [19] **Gribova V.V., Kleshchev A.S., Moskalenko F.M., Timchenko V.A.** A Model for Generation of Directed Graphs of Information by the Directed Graph of Metainformation for a Two-Level Model of Information Units with a Complex Structure // Automatic Documentation and Mathematical Linguistics. 2015. Vol.49, No.6. - P.221-231.
- [20] **Клещёв, А.С.** Модель онтологии предметной области «Медицинская диагностика». Часть 1. Неформальное описание и определение базовых терминов / А.С. Клещев, М.Ю. Черняховская, Ф.М. Москаленко // Журнал НТИ - Серия 2. - 2005. - № 12. - С.1-7.
- [21] **Москаленко Ф.М., Черняховская М.Ю.** Формальное описание заболевания Хронический панкреатит // Информатика и системы управления. 2012. №4. - С.97-106.
- [22] **Петряева, М.В.** Язвенная болезнь в формальном представлении / М.В. Петряева, Д.Б. Окунь // Научная дискуссия: вопросы медицины: сб. ст. по материалам XLVII Международной научно-практической конференции «Научная дискуссия: вопросы медицины». – № 3 (34). – М., Изд. «Интернаука», 2016. – С.62-66.
- [23] **Черняховская, М.Ю.** Формальное представление знаний о конъюнктивитах (издание второе, исправленное и дополненное) / М.Ю. Черняховская, Ф.М. Москаленко, В. Я. Мельников, В. И. Негода, Л. П. Догадова. - Владивосток: ИАПУ ДВО РАН, 2009. – 56 с.
- [24] **Петрова, О.В.** Референсные интервалы количества лейкоцитов в крови и лейкоцитарной формулы у взрослого населения при применении автоматического гематологического анализатора Sysmex xt 2000i / О.В. Петрова, Г.Р. Шабанова, Т.Г. Егорова // Гематология и трансфузиология. 2016. №3. - <https://cyberleninka.ru/article/n/referentsnye-interval-y-kolichestva-leykotsitov-v-krovi-i-leykotsitarnoy-formuly-u-vzroslogo-naseleniya-pri-primenenii-avtomaticheskogo>.
- [25] **Казакова, М.С.** Референсные значения показателей общего анализа крови взрослого работающего населения / М.С. Казакова, С.А. Луговская, В.В. Долгов // Клиническая лабораторная диагностика. 2012. №6. - <https://cyberleninka.ru/article/n/referentsnye-znacheniya-pokazateley-obshego-analiza-krovi-vzroslogo-rabotayuschego-naseleniya>.
- [26] **Бавыкина Л.** Референсные значения тестостерона в крови у мужчин // Актуальная эндокринология. 2015. №4. - <https://cyberleninka.ru/article/n/referentsnye-znacheniya-testosterona-v-krovi-u-muzhchin>.
- [27] **Иголина, Н.А. и др.** Уровень холестерина в популяции взрослого населения РФ 20-70 лет // Атеросклероз и дислипидемии. 2012. №4. - <https://cyberleninka.ru/article/n/uroven-holesterina-v-populyatsii-vzroslogo-naseleniya-rf-20-70-let>.
- [28] **Miller R.A., Myers J., Pople H.** INTERNIST-1: An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine. N Engl J Med. 1982. vol. 307. - P.468-476.

- [29] **Клещёв, А.С.** Постановки практически полезных задач интеллектуальной деятельности / А.С. Клещёв, Е.А. Шалфеева // Дальневосточный математический журнал. 2016. Том 1. - С.44–61.
- [30] **Жмудяк, М.Л.** Автоматизированная система медицинской диагностики заболеваний с учётом их динамики / М.Л. Жмудяк, А.Н. Повалихин, А.В. Стребуков, А.Л. Жмудяк, Г.Г. Устинов // Ползуновский вестник. № 1. 2006. – С.95-106.
- [31] **Москаленко, Ф.М.** База терминов для интеллектуальных медицинских сервисов// Материалы X международной научной конференции "Системный анализ в медицине" (САМ 2016) / Ф.М. Москаленко, Д.Б. Окунь, М.В. Петряева // под общ. ред. В.П.Колосова. 22-23 сентября 2016 г. - Благовещенск: ДНЦ ФПД, 2016. С. 155-158.

## MEDICAL DIAGNOSIS ONTOLOGY FOR INTELLIGENT DECISION SUPPORT SYSTEMS

V.V. Gribova<sup>1</sup>, M.V. Petryaeva<sup>2</sup>, D.B. Okun<sup>3</sup>, E.A. Shalfeeva<sup>4</sup>

*Institute of Automation Control Processes of the FEB RAS, Vladivostok, Russia*  
<sup>1</sup>*gribova@iacp.dvo.ru, <sup>2</sup>margaret@iacp.dvo.ru, <sup>3</sup>okdm@dvo.ru, <sup>4</sup>shalf@iacp.dvo.ru*

### Abstract

An ontology for representation of knowledge about the diagnosis of diseases and syndromes is described. It allows experts to represent knowledge about the diagnosis of a wide range of diseases. Knowledge of the disease diagnosis including its form, etiology, pathogenesis, variants of the clinical course is represented according to their clinical classification: a set of diagnostic features, alternative symptom complexes with pre-morbid biological, personality and other factors, dynamic signs of symptoms and clinical features altered by the impact of events. The ranking of symptoms by specificity is possible, as well as a description of the conditions predisposing or contributing to the disease, which allows to attribute the disease with a set of specific features. The article includes an informal description of the ontology, as well as its model with a description of the main terms, knowledge, situations and limitations of their integrity in the form of ontological agreements. An example of the use of the ontology for the formation of a knowledge base on diagnosis and differential diagnosis of diseases from the group "Diseases of the digestive organs" is given in the article. The ontology is implemented on the cloud platform of IACPaaS and now is actively used by experts to create knowledge bases in various fields of medicine.

**Keywords:** ontology, knowledge base, diagnosis of diseases, syndrome, symptom complex, medical intelligent systems, decision support system.

**Citation:** Gribova VV, Petryaeva MV, Okun DB, Shalfeeva EA. Medical diagnosis ontology for intelligent decision support systems [In Russian]. *Ontology of designing*. 2018; 8(1): 58-73. - DOI: 10.18287/2223-9537-2018-8-1-58-73.

### References

- [1] **Galanova GI.** Vrachebnye oshibki – problema ne tol'ko vracha // Menedzher zdavoohranenie. [In Russian].– 2014; 8: 49-52.
- [2] Problemnye voprosy kontrolya kachestva i bezopasnosti medicinskoj deyatelnosti na sovremennom ehtape. [In Russian]. - <http://rusmedunion.ru/?p=2222/>.
- [3] Rasporyazhenie Pravitel'stva RF №1662-r ot 17.11.08. [In Russian].
- [4] Iskusstvennyj intellekt dlya Mediciny – obzor. - [In Russian]. - [http://www.livemd.ru/tags/iskusstvennyj\\_intellekt/](http://www.livemd.ru/tags/iskusstvennyj_intellekt/).
- [5] **Gribova VV, Kleshchev AS, Krylov DA. et al.** // Proekt IACPaaS. Kompleks dlya intellektual'nyh sistem na osnove oblachnyh vychislenij / Iskusstvennyj intellekt i prinyatie reshenij. [In Russian]. 2011; 1: 27–35.
- [6] **Soltan RA, Soltan RA, Rashad MZ, El-Desouky B.** Diagnosis of Some Diseases in Medicine via computerized Experts System // International Journal of Computer Science & Information Technology. 2013; 5(5): 79-90.
- [7] **Zagorulko YuA.** Semantic technology for development of intelligent systems oriented on experts in subject domain [In Russian]. *Ontology of Designing*. 2015; 5(1): 30-46.
- [8] **Mel'nik KV, Ershova SI.** Problemy i osnovnye podhody k resheniyu zadachi medicinskoj diagnostiki //Sistemi obrobotki informacii, 2011; 2 (92). [In Ukraine].

- [9] **Le NV, Kamaev VA, Panchenko DP, Trushkina OA.** Model' predstavleniya znaniy pri sozdaniy meditsinskoj ehkspertnoj sistemy differencial'noj diagnostiki // *Izvestiya Volgogradskogo gosudarstvennogo tekhnicheskogo universiteta.* – 2014; 20(6): 42-50. [In Russian].
- [10] **Le NV, Kamaev VA, Panchenko DP, Trushkina OA.** Obzor podhodov k proektirovaniyu meditsinskoj sistemy differencial'noj diagnostiki // *Izvestiya Volgogradskogo gosudarstvennogo tekhnicheskogo universiteta.* 2014; 20(6): 50–58. [In Russian].
- [11] **ZHarkova OS, SHaropin KA, Seidova AS, Berestneva EV, Osadchaya IA.** Postroenie sistem podderzhki prinyatiya reshenij v medicine na osnove derev'ev reshenij // *Sovremennyye naukoemkie tekhnologii.* – 2016; 6-1: 33-37. [In Russian].
- [12] **Haug et al** An ontology-driven, diagnostic modeling system // *Journal of the American Medical Informatics Association.* – 2013; 20(e1): e102-e110.
- [13] **Patil RS, Szolovits P, Schwartz WB.** Modelling Knowledge of the Patient in Aced-base and Electrolyte Disorders // in Szolovits, P. (Ed.) *Artificial Intelligence in Medicine,* - Boulder, CO: Westview Press, 1982. - P.191-226.
- [14] **Uciteli A et al.** Ontology-based specification, identification and analysis of perioperative risks // *Journal of biomedical semantics.* – 2017; 8(1): 36.
- [15] **Yu Lin, Jie Zheng and Yongqun He** VICO: Ontology-based representation and integrative analysis of Vaccination Informed Consent forms // *Journal of Biomedical Semantics* (2016) 7:20 - DOI: 10.1186/s13326-016-0062-4.
- [16] **Bouamrane MM, Rector A, Hurrell M.** Using OWL ontologies for adaptive patient information modelling and preoperative clinical decision support // *Knowl Inf Syst* (2011) 29:405–418. – DOI: 10.1007/s10115-010-0351-7
- [17] **Abidi S. et al.** Using OWL ontologies for clinical guidelines based comorbid decision support // *System Science (HICSS), 2012 45th Hawaii International Conference on.* – IEEE, 2012. – P.3030-3038. - DOI: 10.1109/HICSS.2012.629.
- [18] **Gribova VV, Kleshchev AS, Moskalenko FM, Timchenko VA.** A Two-level Model of Information Units with Complex Structure that Correspond to the Questioning Metaphor [In Russian] // *Automatic Documentation and Mathematical Linguistics.* 2015; 49(5): 172-181.
- [19] **Gribova VV, Kleshchev AS, Moskalenko FM, Timchenko VA.** A Model for Generation of Directed Graphs of Information by the Directed Graph of Metainformation for a Two-Level Model of Information Units with a Complex Structure [In Russian] // *Automatic Documentation and Mathematical Linguistics.* 2015; 49(6): 221-231.
- [20] **Kleshchev AS, CHernyahovskaya MYU, Moskalenko FM.** Model' ontologii predmetnoj oblasti «Medicinskaya diagnostika». CHast' 1. Neformal'noe opisaniye i opredeleniye bazovykh terminov [In Russian] // *ZHurnal NTI - Seriya 2.* – 2005; 12: 1-7.
- [21] **Moskalenko FM, CHernyahovskaya MYU.** Formal'noe opisaniye zabolevaniya Hronicheskij pankreatit // *Informatika i sistemy upravleniya.* 2012; 4: 97-106. [In Russian]
- [22] **Petryaeva MV, Okun DB.** YAzvennaya bolezn' v formal'nom predstavlenii // *Nauchnaya diskussiya: voprosy mediciny: sb. st. po materialam XLVII Mezhdunarodnoj nauchno-prakticheskoy konferencii «Nauchnaya diskussiya: voprosy mediciny».* – № 3 (34). – M., Izd. «Internauka», 2016. – Pp. 62-66. [In Russian]
- [23] **CHernyahovskaya MY, Moskalenko FM, Melnikov VYA, Negoda VI, Dogadova LP.** Formal'noe predstavlenie znaniy o kon'yunktivitah (izdaniye vtoroje, ispravlennoje i dopolnennoje). - Vladivostok: IAPU DVO RAN, 2009. – 56 p. [In Russian].
- [24] **Petrova OV, SHabanova GR, Egorova TG.** Referensnyye intervaly kolichestva lejkocitov v krovi i lej-kocitarnoy formuly u vzroslogo naseleniya pri primenenii avtomaticheskogo gematologicheskogo analizatora Sysmex xt 2000i // *Gematologiya i transfuziologiya.* 2016. №3. [In Russian]. - <https://cyberleninka.ru/article/n/referentsnyye-intervaly-kolichestva-leykotsitov-v-krovi-i-leykotsitarnoy-formuly-u-vzroslogo-naseleniya-pri-primenenii-avtomaticheskogo>.
- [25] **Kazakova MS, Lugovskaya SA, Dolgov VV.** Referensnyye znacheniya pokazatelej obshchego analiza krovi vzroslogo rabotayushchego naseleniya // *Klinicheskaya laboratornaya diagnostika.* 2012. №6. [In Russian]. - <https://cyberleninka.ru/article/n/referentsnyye-znacheniya-pokazateley-obshchego-analiza-krovi-vzroslogo-rabotayushchego-naseleniya>.
- [26] **Bavykina L.** Referensnyye znacheniya testosterona v krovi u muzhchin // *Aktual'naya ehndokrinologiya.* 2015. No4. [In Russian]. - <https://cyberleninka.ru/article/n/referentsnyye-znacheniya-testosterona-v-krovi-u-muzhchin>.
- [27] **Igonina NA et al.** Uroven' holesterina v populyacii vzroslogo naseleniya RF 20-70 let // *Ateroskleroz i dislipidemii.* 2012. No4. [In Russian]. - <https://cyberleninka.ru/article/n/uroven-holesterina-v-populyatsii-vzroslogo-naseleniya-rf-20-70-let>.
- [28] **Miller RA, Myers J, Pople H.** INTERNIST-1: An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine. *N Engl J Med.* 1982; 307: 468-476.
- [29] **Kleshchev AS, Shalfeeva EA.** Postanovki prakticheski poleznykh zadach intellektual'noj deyatel'nosti // *Dal'nevostochnyj matematicheskij zhurnal.* 2016; 1: 44–61 [In Russian].
- [30] **ZHmudiyak ML, Povalihin AN, Strebukov AV, ZHmudiyak AL, Ustinov GG.** Avtomatizirovannaya sistema



medicinskoj diagnostiki zabolevanij s uchetom ih dinamiki // Polzunovskij vestnik. 2006; 1: 95-106 [In Russian].

- [31] *Moskalenko FM, Okun DB, Petryaeva MV*. Baza terminov dlya intellektual'nyh medicinskih servisov// Materialy X mezhdunarodnoj nauchnoj konferencii "Sistemnyj analiz v medicine" (SAM 2016) / pod obshch. red. V.P. Kolosova. 22-23.09.2016, Blagoveshchensk. Blagoveshchensk: DNC FPD, 2016. - P.155-158 [In Russian].

## Сведения об авторах



**Грибова Валерия Викторовна**, 1965 г. рождения. Окончила Ленинградский политехнический институт (по специальности "прикладная математика") в 1989 г., д.т.н. (2007). Заведующий лабораторией интеллектуальных систем Института автоматизации и процессов управления Дальневосточного отделения РАН, заместитель директора по научной работе, вице-президент Российской ассоциации искусственного интеллекта. В списке научных трудов более 200 работ в области искусственного интеллекта, проблемно-ориентированных систем, основанных на знаниях, специализированных программных моделей и систем.

**Valeriya Victorovna Gribova** (b.1965) graduated from the Leningrad Polytechnic University in 1989, Professor's degree (2007). She is a Head of lab. of intellectual systems in the Institute for Automation & Control Processes of the FEB RAS, Research Deputy Director, an expert of Analytic Center in Government of Russian Federation, a Vice-President of Russian Association of Artificial Intelligence.

He is co-author of more than 200 publications in the fields of AI, informatics, program models, technologies and systems.

**Петряева Маргарита Вячеславовна**, 1965 г. рождения. Окончила Тихоокеанский государственный медицинский университет по специальности «лечебное дело» в 1989 г., к.м.н. (2001). Врач кибернетик, эксперт в области медицинской диагностики. В списке научных трудов более 90 работ в области формализации медицинских знаний, разработки интеллектуальных систем поддержки принятия решений.

**Margarita Vyacheslavovna Petryaeva** (b.1965) graduated from the Pacific State Medical University (Vladivostok) in 1989, Ph.D in Medical sciences (2001). Research associate of the laboratory of Intelligent systems at the Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences, is co-author of more than 90 publications.



**Окунь Дмитрий Борисович**, 1973 г. рождения. Окончил Владивостокский государственный медицинский университет по специальности: «лечебное дело» в 1996 году, к.м.н. (2000). Научный сотрудник лаборатории Интеллектуальных систем Института автоматизации и процессов управления Дальневосточного отделения РАН.

**Dmitry B. Okun** (b.1973) graduated from the Vladivostok State Medical University in 1996, PhD in Medical sciences (2000). Research associate of the laboratory of Intelligent systems at the Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences.



**Шалфеева Елена Арефьевна**, 1967 г. рождения. Окончила Дальневосточный государственный университет по специальности "прикладная математика" в 1989 г., к.т.н. (2000). Старший науч-

ный сотрудник лаборатории Интеллектуальных систем Института автоматизации и процессов управления Дальневосточного отделения РАН, доцент по специальности. В списке научных трудов более 80 работ.

**Elena Arefjevna Shalfeeva** (b.1967) graduated from the Far Eastern State University (Vladivostok) in 1989, PhD (2000). She is Senior Researcher at lab. of intelligent systems in the Institute for Automation & Control Processes of the FEB RAS, lecture. She is co-author of more than 80 publications in the fields of Program models and systems and AI.



УДК 004.912

## ОНТОЛОГИЧЕСКИЕ РЕСУРСЫ И ИНФОРМИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА В ПРЕДМЕТНОЙ ОБЛАСТИ «БЕЗОПАСНОСТЬ»

Н.В. Лукашевич<sup>1</sup>, Б.В. Добров<sup>2</sup>, А.М. Павлов<sup>3</sup>, С.В. Штернов<sup>4</sup>

Научно-исследовательский вычислительный центр МГУ имени М.В. Ломоносова, Москва, Россия  
<sup>1</sup>louk\_nat@mail.ru, <sup>2</sup>dobrov\_bv@mail.ru, <sup>3</sup>pavlov.andrew.m@gmail.com, <sup>4</sup>shternov@gmail.com

### Аннотация

В статье рассматривается подход к описанию широкой области национальной безопасности как тезауруса (лингвистической онтологии) для автоматической обработки документов. Созданный Тезаурус по безопасности имеет модель представления тезауруса RuТез и не имеет аналогов в мире в такой широкой предметной области. Тезаурус по безопасности используется в специализированной информационно-аналитической системе, в том числе для автоматической текстовой классификации документов в соответствии с несколькими рубриками, включая рубрику угроз, рубрику ценностей, рубрику региональных проблем и др. Используемая информационно-поисковая система NearIdx обеспечивает стандартные функции информационного поиска, а также даёт возможность задания запросов и поиска информации с использованием специализированных ресурсов, включая онтологию и рубрики. Аналитический компонент NearIdx предоставляет возможности фасетного анализа, спектрально-фасетного анализа временных рядов, построения когнитивных схем и аналитических справок, в которых могут использоваться созданные лексико-терминологические ресурсы.

**Ключевые слова:** тезаурус, онтология, национальная безопасность, информационный поиск, рубрику, автоматическая классификация, текстовая аналитика.

**Цитирование:** Лукашевич, Н.В. Онтологические ресурсы и информационно-аналитическая система в предметной области «Безопасность» / Н.В. Лукашевич, Б.В. Добров, А.М. Павлов, С.В. Штернов // Онтология проектирования. – 2018. – Т. 8, №1(27). – С.74-95. – DOI: 10.18287/2223-9537-2018-8-1-74-95.

### Введение

Настоящее время характеризуется огромными объёмами электронной текстовой информации, пополняемой ежедневно, включая новости и газетные публикации, аналитические отчёты, научные статьи, а также сообщения в социальных сетях. Такого рода электронные текстовые ресурсы дают новые возможности для выявления происходящих процессов и возникающих трендов в разных сферах деятельности [1-4].

Одной из важных сфер применения мониторинга текстовой информации является сфера национальной безопасности, для которой важно отслеживание и предупреждение реальных и потенциальных угроз. Термин «национальная безопасность» определяется в Стратегии национальной безопасности Российской Федерации<sup>1</sup> как состояние защищённости личности, общества и государства от внутренних и внешних угроз, при котором обеспечиваются реализация конституционных прав и свобод граждан Российской Федерации, достойные качество и уровень их жизни, суверенитет, независимость, государственная и территориальная целостность, устойчивое социально-экономическое развитие Российской Федерации. Национальная безопасность включает в себя оборону страны и все виды безопасности, предусмотренные

---

<sup>1</sup>[http://www.consultant.ru/document/cons\\_doc\\_LAW\\_191669/61a97f7ab0f2f3757fe034d11011c763bc2e593f/](http://www.consultant.ru/document/cons_doc_LAW_191669/61a97f7ab0f2f3757fe034d11011c763bc2e593f/)

Конституцией Российской Федерации и законодательством Российской Федерации, прежде всего государственную, общественную, информационную, экологическую, экономическую [5], транспортную, энергетическую безопасность, безопасность личности.

В данной статье рассматриваются возможности для определения угроз и трендов в разных сферах предметной области «Безопасность» на основе специализированных лексико-терминологических ресурсов: лингвистических онтологий и специализированных рубрикаторов. Применяемая лингвистическая онтология представляет собой расширенную версию тезауруса RuТез [6]. Созданные рубрикаторы содержат в себе категории существующих ценностей, т.е. того, что нужно защищать, и предполагаемых угроз, выявленных как по нормативным документам, так и на основе анализа текстов средств массовой информации.

Созданные лексико-терминологические ресурсы используются для автоматической классификации документов, а также в рамках специализированной информационно-аналитической системы. В статье рассматриваются инструменты, которые входят в информационно-аналитическую систему и дают дополнительные возможности для проведения мониторинга поступающей информации.

## 1 Методы автоматической обработки документов в сфере безопасности

В литературе представлено несколько направлений обработки текстов на естественном языке в сфере национальной и международной безопасности.

Многие работы посвящены анализу экстремистских сообщений в социальных сетях. Часть исследований проявлений терроризма и экстремизма в Twitter посвящена деятельности ИГИЛ (Исламское государство Ирака и Леванта). Известно, что эта террористическая организация активно работает с множеством социальных сетей [7-8]. В частности, ИГИЛ и связанные с ним организации поддерживают большое количество учётных записей Twitter на многих языках для распространения своих идей. В работе [9] дан анализ сообщений в социальных сетях людей, которые присоединились или попытались присоединиться к ИГИЛ. В большинстве случаев они выражали сильные антиамериканские и антизападные настроения задолго для присоединения к этой организации.

В статье [10] описывается корпус, содержащий 100 текстов, написанных исламистами. В частности, из собраний хадисов (исламских религиозных текстов) были извлечены фрагменты о войне, неверующих и наказаниях (64 текста). Кроме того, корпус содержит сообщения из исламских блогов, а также статьи из исламистского журнала Inspire. Эти тексты были размечены на нескольких уровнях, включая синтаксические, временные и другие аннотации.

Поскольку одним из предпосылок поддержки терроризма является резко негативное отношение к определённым явлениям или группам людей, то так называемые заявления «ненависти» (“hate”) требуют особого внимания. В работе [11] сообщения в блогах классифицируются не только на классы выраженных эмоций (положительные, отрицательные, нейтральные), но и в зависимости от типа обсуждаемого действия (отрицательный - *кровопролитие, жестокость*, или положительный – *помощь, поддержка*). Квок и Ван [12] описывают собранный сбалансированный набор из 24,5 тыс. твитов, которые классифицируются как расистские и нормальные. В [13] представлен корпус из 16 000 твитов, в которых 3,3 тыс. твитов обозначены как сексистские, а 1,9 тыс. – расистские.

В работе [14] указывается, что автоматическое обнаружение заявлений, которые разжигают ненависть по отношению к некоторым группам населения (что является одним из признаков экстремизма), осложняется следующими факторами:

- сообщение не всегда может быть найдено на основе простого набора ключевых слов, потому что некоторые слова намеренно искажены (чтобы избежать обнаружения);

- любые фиксированные списки оскорбительных слов постоянно требуют добавления;
- заявления о ненависти могут быть написаны на прекрасном литературном языке;
- преступление и ненависть могут пересекать границы предложения, когда объект, на который направлено высказывание, находится в другом предложении;
- использование сарказма.

Шмидт и Виганд [15] предоставляют обзор существующих подходов к обнаружению сообщений о ненависти. Отмечается, что такие подходы основаны на применении методов машинного обучения на основе нескольких групп признаков.

В сфере автоматической обработки текстов по информационной безопасности Лим и др. [16] обсуждают создание базы данных для аннотирования текстов описаний вредоносных программ. Структура аннотаций вводится на основе словаря МАЕС для определения характеристик вредоносного программного обеспечения [17], а также базы данных. Авторы планируют использовать базу данных для создания моделей, которые потенциально могут помочь исследователям кибербезопасности в своих усилиях по сбору и анализу данных. Горохов и др. [18] изучают применение свёрточной нейронной сети для обнаружения аномалий в данных электронной почты.

Несколько проектов посвящены глобальному мониторингу происходящих в мире событий, что необходимо для понимания существующих проблем и формирования подходов к их разрешению, например, Международная система раннего предупреждения о кризисах (ICEWS), поддерживаемая Lockheed Martin и Global Data on Events Language and Tone (GDELT) [19-20]. Для прогнозирования эскалации конфликтов [19] использовались ручные и автоматически собранные данные о событиях. В ICEWS используются также статистические и агентные модели и утверждается, что точность прогнозирования составляет 80%. Система GDELT была использована для отслеживания сообщений, вызывающих ненависть, после голосования по поводу отделения Великобритании от Евросоюза [20].

## 2 Лингвистическая онтология в области безопасности

Для представления знаний в сфере безопасности используется модель комплекса лингвистических онтологий РуТез, которые представляют собой формализованные информационно-поисковые тезаурусы, ориентированные на автоматическую обработку больших текстовых коллекций и поддержку решения задач текстовой аналитики [6, 21]. В модели учитываются три парадигмы описания знаний в широких предметных областях: информационно-поисковые тезаурусы, тезаурусы типа WordNet, формальные онтологии.

В настоящее время существует несколько крупных русскоязычных лингвистических онтологий (тезаурусов).

- Тезаурус РуТез, содержащий слова и фразы литературного русского языка вместе с терминами так называемой общественно-политической области [22].
- Тезаурус РуТез-lite, опубликованная версия РуТез<sup>1</sup>, который исследователи могут получить бесплатно для некоммерческого использования [23].
- Общественно-политический тезаурус, содержащий слова и выражения литературного языка и термины из общественно-политической области. Общественно-политическая область – это широкая область современных общественных отношений, представляющая повседневную жизнь современного общества и объединяющая многие профессиональные сферы, такие как политика, право, экономика, международные отношения, финансы, военные дела, искусство и другие. Особенностью этой предметной области является то, что

---

<sup>1</sup> <http://www.labinform.ru/pub/ruthes/index.htm>

термины этой предметной области обычно известны не только профессионалам, но и обычными людям [24], поскольку это сфера пересечения профессиональных и общих знаний о мире. Общественно-политический тезаурус может использоваться отдельно для обработки документов. В то же время он входит в состав трех крупных тезаурусов: РуТез, онтологии ОЕНТ и Тезауруса по безопасности.

- Онтология по естественным наукам и технологиям (ОЕНТ), которая включает в себя термины математики, физики, химии, геологии, астрономии и т. д., термины инженерных предметных областей (нефть и газ, электростанции, космические технологии, летательные аппараты и т. д.). ОЕНТ также включает Общественно-политический тезаурус, поскольку научно-технические проблемы могут обсуждаться вместе с политическими, экономическими, промышленными и другими вопросами [25].
- Тезаурус безопасности, который является расширением тезауруса РуТез и включает терминологию, связанную с социальными, национальными и религиозными конфликтами, экстремизмом и терроризмом, информационной безопасностью.

В таблице 1 приведены количественные характеристики вышеперечисленных ресурсов.

Таблица 1 - Лингвистико-терминологические ресурсы в формате тезауруса РуТез

Тезаурус	Число понятий	Число текстовых входов	Число отношений
РуТез	55275	170130	226743
РуТез-lite	31540	111559	128866
Общественно-политический тезаурус	41426	121292	161523
Онтология ОЕНТ	94103	262955	376223
Тезаурус по безопасности	66810	236321	271297

Тезаурус по безопасности представляет собой лингвистическую онтологию, т.е. онтологию, понятия которой опираются на значения существующих в языке слов и выражений. Каждое понятие онтологии имеет уникальное и однозначное имя. Каждое понятие онтологии связано с набором слов и выражений, посредством которых данное понятие может выражаться в тексте – текстовые входы понятия. Набор текстовых входов понятия может включать собственно синонимы, слова разных частей речи (так называемые дериваты), устойчивые словосочетания и другие типы выражений.

На рисунках 1-3 показан интерфейс разработки тезауруса. Верхняя левая форма содержит список понятий, нижняя левая форма показывает текстовые входы для выделенного понятия. Правая верхняя форма представляет отношения выделенного понятия, а нижняя правая форма показывает текстовые входы для соответствующих понятий. Можно видеть, что понятия снабжены многочисленными текстовыми вариантами, извлеченными из реальных текстов, например, понятие *импортная зависимость* (рисунок 1) может быть выражено в тексте как *зависимость от импорта* или *зависимость от импортных товаров*. На рисунке 3 показаны варианты языковых выражений, используемых для выражения понятий *атака отказа обслуживания* и *распределенная DOS-атака* в сфере компьютерной безопасности.

Как и в РуТез, в Тезаурусе по безопасности имеется четыре основных типа отношений [24]. Первый тип отношений – родовидовое отношение *ниже-выше*, представляет собой отношение класс-подкласс, обладает свойствами транзитивности и наследования.

Второй тип отношений – отношение *часть-целое*. Отношение используется не только для описания физических частей, но и для других внутренних сущностей понятия, таких как свойства или роли для ситуаций. Важным условием при установлении этого отношения является то, что понятия-части должны быть жестко связаны со своим целым, то есть каждый пример понятия-части должен в течение всего времени своего существования являться ча-

стью для понятия-целого и не относиться к чему-либо другому. В этих условиях удаётся выполнить свойство транзитивности введённого таким образом отношения часть-целое, что очень важно для автоматического вывода в процессе автоматической обработки текстов. Кроме того, в таких условиях при появлении в тексте упоминания части можно автоматически выводить, что текст имеет отношение и к целому.

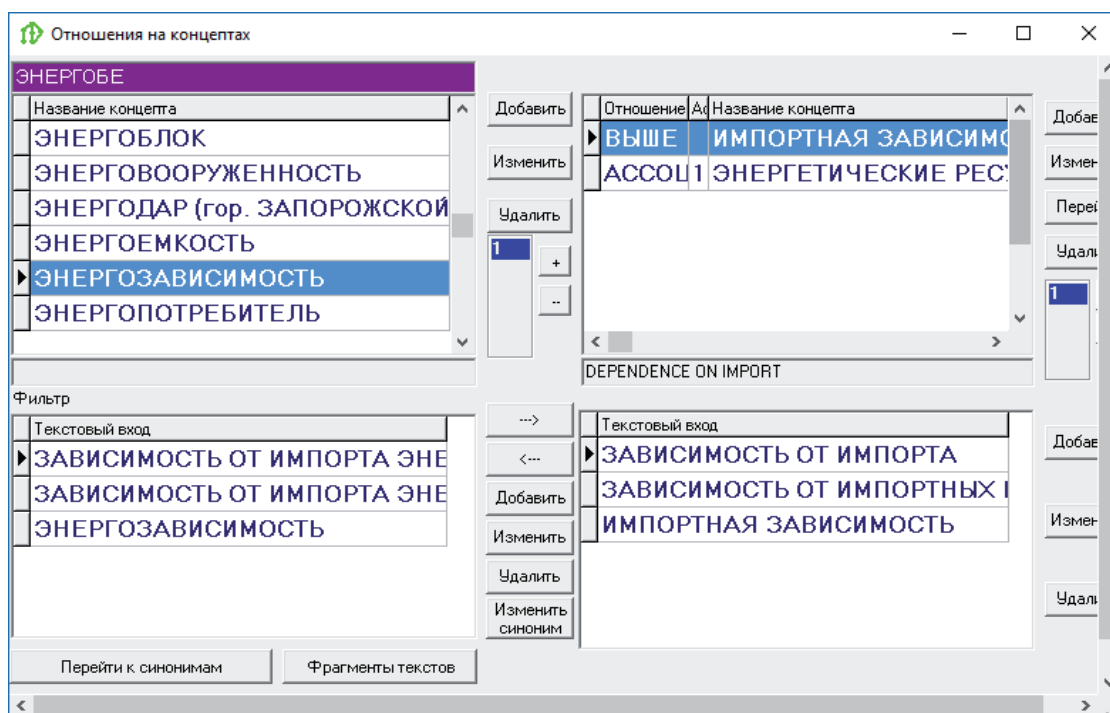


Рисунок 1 - Текстовые входы и отношения понятия *энергозависимость* в сфере экономической безопасности

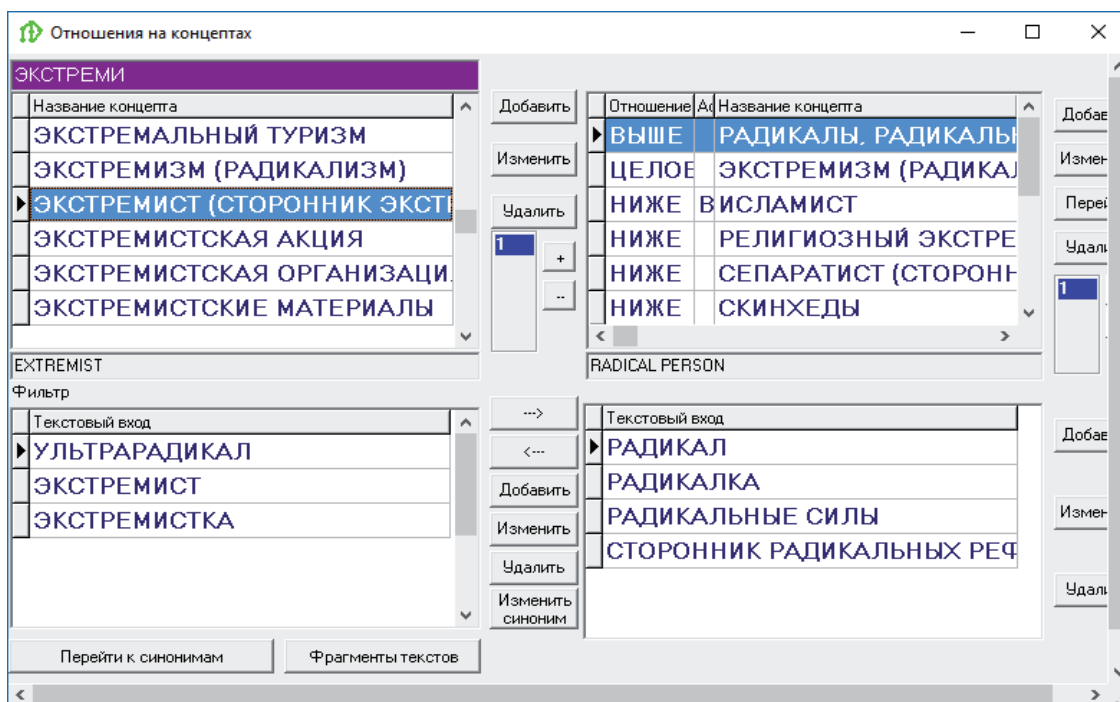


Рисунок 2 - Текстовые входы и отношения для понятия *экстремист* в сфере борьбы с экстремизмом

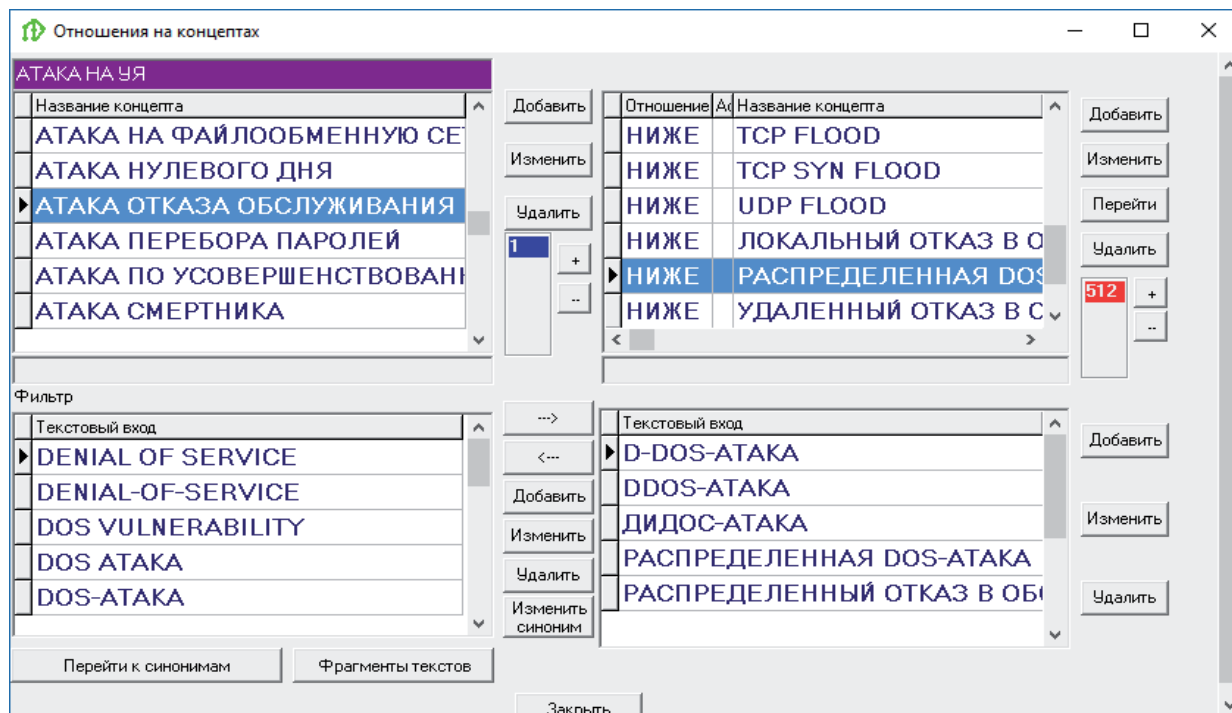


Рисунок 3 - Текстовые входы и отношения для понятия атака отказа обслуживания в сфере компьютерной безопасности

Ещё один тип отношения, называемого несимметричной ассоциацией  $асц_2 - асц_1$ , связывает два понятия, которые не находятся в отношениях часть-целое или класс-подкласс, и существование одного из понятий требует существования другого понятия. На рисунке 1 показан пример такой зависимости понятия *энергетическая зависимость* от понятия *энергоресурсы*. Также понятие *замещение импорта* появляется, когда существует понятие *импортной зависимости*. В сфере компьютерной безопасности можно привести пример понятия *атака на уязвимость*, существование которого зависит от понятия *уязвимость в программе*.

Последний тип отношений - симметричная ассоциация  $асц$  связывает, например, понятия, очень близкие по смыслу, но которые разработчики не решились соединить в одно понятие (предсинонимия).

Таким образом, система отношений в тезаурусах типа РуТез описывает наиболее существенные отношения понятий. Рисунок 4 показывает фрагмент иерархии понятий, относящихся к понятию *опасность, угроза*.

Тезаурус по безопасности как отдельный ресурс создаётся в течение последних пяти лет. Он пополняется на основе нескольких источников.

Во-первых, анализируется существующая нормативно-справочная литература, содержащая термины, относящиеся к сфере безопасности. Такие термины вносятся в тезаурус.

Во-вторых, создаются специализированные текстовые коллекции по конкретным подобластям сферы безопасности, из которых производится автоматическое извлечение кандидатов в термины на основе синтаксических, лексических и синтаксических признаков [5]. Извлечённые слова и выражения просматриваются терминологами и вносятся в тезаурус в виде новых понятий или как текстовые входы к существующим понятиям.

В-третьих, поступающие новости из средств массовой информации по тематике безопасности обрабатываются на основе текущей версии тезауруса (см. раздел 3 и рисунок 5), при этом может быть выявлена нехватка терминов, которые добавляются в тезаурус.

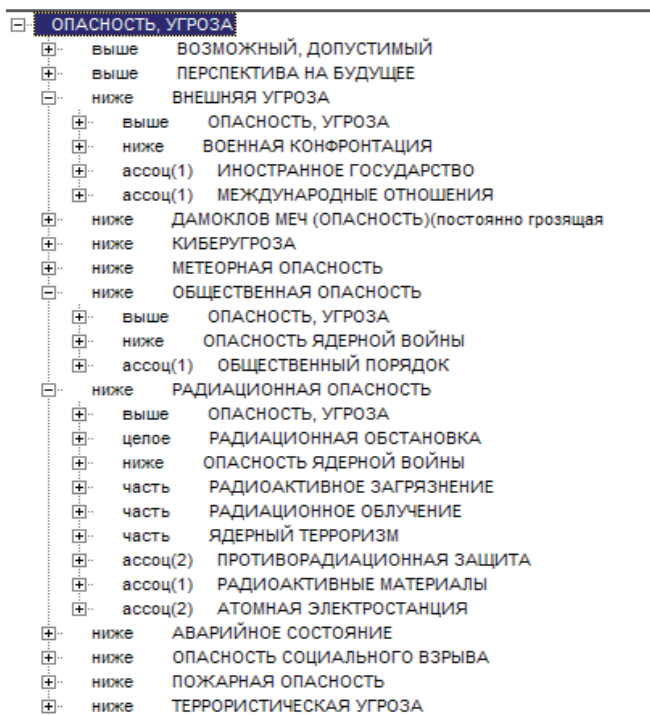


Рисунок 4 - Фрагмент иерархии понятий, относящихся к понятию *опасность, угроза*

Сантимент     NECRF     Подсвечивать фон    [Справка](#)    [Новая обработка](#)

СПИСОК НАЙДЕННЫХ РУБРИК  
 СПИСОК НАЙДЕННЫХ РУБРИК САНТИМЕНТА  
 ТЕМАТИЧЕСКАЯ АННОТАЦИЯ  
 АННОТАЦИЯ  
 ОБРАБОТАННЫЙ ТЕКСТ

**«Выход из ситуации не просматривается»**  
 Дмитрий Дризе — о **противостоянии России** и Запада  
 26.02.2018, 09:37

Дональд Трамп назвал позором действия России и Ирана в Сирии. Так президент США прокомментировал боевые действия в пригороде Дамаска — Восточной Гуте. Ранее Совет безопасности ООН единогласно принял резолюцию о гуманитарном перемирии в Сирии сроком на 30 дней. Однако в Восточной Гуте продолжились вооруженные столкновения. Политический обозреватель «Коммерсантъ» Дмитрий Дризе считает, что Россия ждет новое серьезное испытание на международной арене.

Совет безопасности ООН призвал стороны конфликта в Сирии прекратить боевые действия на всей территории страны на срок не менее 30 дней. Документ также призывает стороны конфликта прекратить огонь над населением, которое страдает от бомбежек и обстрелов. Документ также призывает стороны конфликта прекратить огонь над населением, которое страдает от бомбежек и обстрелов.

Не секрет, что резолюция ООН была инициирована в ответ на операцию в Восточной Гуте, где армия Асада при поддержке Ирана и, возможно, России ведет боевые действия против засевших там вооруженных формирований.

По уже сложившейся традиции любое продвижение правительственных войск, с кем бы они ни боролись, встречает праведный гнев западной общественности, в частности, указывается на страдания мирного населения, чего никогда не случается при ударах коалиции во главе с США.

Рисунок 5 - Пример сопоставления текста с тезаурусом

### 3 Автоматическая обработка текстов на основе тезауруса

Обработка текстов на основе тезауруса включают в себя следующие этапы.

- Графематический и морфологический анализ, который переводит разные формы слова к единой словарной форме (лемме).



- Терминологический анализ документа, состоящий в автоматическом сопоставлении текста с тезаурусом на основе последовательностей лемм. На рисунке 5 показан результат терминологического анализа текста «О противостоянии России и Запада»<sup>1</sup>. Выделенные слова и словосочетания найдены в Тезаурусе по безопасности.
- Автоматическое разрешение многозначности слов. Фрагменты текста, выделенные на рисунке 5 коричневым и синим цветом, означают обнаруженные многозначные термины, т.е. термины, которые соответствуют разным понятиям тезауруса. Так, имя *Дамаск* может относиться к столице Сирии или сирийской провинции. Разрешение многозначности слов производится на основе сопоставления контекста неоднозначного слова в документе и списка близких по тезаурусу слов и словосочетаний [6]. Так, в тексте примера разметка "Т\_М" означает, что выбрано значение *Дамаска* как столицы.
- Автоматическое построение тематического представления текста заключается в том, что близкие по смыслу понятия группируются в так называемые тематические узлы, которые затем распределяются по их значимости для текста: основные и локальные. Значимость тематического узла учитывается в формируемом весе понятия для формирования концептуального индекса документа. На рисунке 6 показаны извлечённые из вышеупомянутого текста тематические узлы, где каждый следующий узел показан с отступом по отношению к предыдущему. Видно, что в один из крупных узлов собрались различные понятия, связанные с демографической обстановкой.

☑ СПИСОК НАЙДЕННЫХ РУБРИК Справка Новая обработка

Угрозы

U060200000	Религиозная напряженность	67
U300103000	Внешняя агрессия	64
U100500000	Незаконные вооруженные формирования	56
U100400000	Международный терроризм	56

■ СПИСОК НАЙДЕННЫХ РУБРИК САНТИМЕНТА

☑ ТЕМАТИЧЕСКАЯ АННОТАЦИЯ

\*\*\*\* РОССИЙСКАЯ ФЕДЕРАЦИЯ; ГОСУДАРСТВО-УЧАСТНИК; РОССИЯНЕ; ГОСУДАРСТВО; МОСКВА;

\*\*\*\* ● ИРАН; ТЕГЕРАН;

\*\*\*\* ● ● СИРИЯ; СИРИЙСКАЯ АРАБСКАЯ АРМИЯ; ДЕЙР-ЭЗ-ЗОР (ГОРОД); БАШАР АСАД; ДАМАСК;

\*\*\*\* ● ● ● ОРГАНИЗАЦИЯ ОБЪЕДИНЕННЫХ НАЦИЙ; РЕЗОЛЮЦИЯ СОВЕТА БЕЗОПАСНОСТИ ООН;

\*\*\*\* ● ● ● ● БОЕВЫЕ ДЕЙСТВИЯ; ГУМАНИТАРНАЯ ПАУЗА; АТАКА, ВОЕННЫЙ УДАР; ПЕРЕМИРИЕ; БОМБАРДИРОВКА; ВОЕННЫЕ ДЕЙСТВИЯ;

\*\*\*\* ● ● ● ● ● КОНФЛИКТ; ВООРУЖЕННАЯ АГРЕССИЯ; ГРАЖДАНСКАЯ ВОЙНА В СИРИИ; ГРАЖДАНСКАЯ ВОЙНА;

\*\*\*\* ● ● ● ● ● ● ТРАМП, ДОНАЛЬД; ПРЕЗИДЕНТ США; УПРАВЛЯТЬ, РУКОВОДИТЬ; США; ПОЛИТИЧЕСКАЯ ДЕЯТЕЛЬНОСТЬ; НАСЕЛЕНИЕ; ПОЛИТИЧЕСКИЙ ДЕЯТЕЛЬ; ОРГАНИЗАЦИЯ, УЧРЕЖДЕНИЕ;

■ АННОТАЦИЯ

▼ ☑ ОБРАБОТАННЫЙ ТЕКСТ

«Выход из ситуации не просматривается»  
 Дмитрий Дризе — о **противостоянии России** и Запада  
 26.02.2018, 09:37

**Дональд Трамп** назвал позором **действия России** и **Ирана** в **Сирии**. Так **президент США** прокомментировал **боевые действия** в пригороде **Дамаска** — Восточной Гуте. Ранее Совет безопасности **ООН** единогласно принял резолюцию о гуманитарном перемирии в **Сирии** сроком на 30 дней. Однако в Восточной Гуте продолжились **вооруженные столкновения**. **Политический обозреватель** «Коммерсантъ FM» Дмитрий Дризе считает, что **Россию** ждет новое серьезное **испытание** на **международной арене**.

**Совет безопасности ООН** требует от всех **участников конфликта в Сирии** прекратить **боевые действия** на всей **территории страны** на срок не менее 30 дней — это необходимо для **оказания помощи населению**, которое страдает от **бомбежек** и обстрелов. **Документ**

Рисунок 6 - Пример автоматически проставленных рубрик и тематических узлов

- Формирование концептуального индекса документа. Концептуальный индекс документа состоит из понятий, найденных в документе, и их весов. Вес понятия учитывает значимость соответствующего узла темы и частоту понятия в документе. В тексте примера понятие *гуманитарная пауза* было явно упомянуто только один раз в тексте, и соответствующее понятие может получить слишком низкий вес, если учитывать только частоту

<sup>1</sup> <https://www.kommersant.ru/doc/3558551>

упоминания, но при поддержке основного тематического узла «боевые действия» вес понятия *гуманитарная пауза* становится значительно выше.

- Автоматическая рубрикация по заданному рубрикатору. На рисунке 6 показаны рубрики, которые были автоматически получены для текста примера: *Религиозная напряженность, Внешняя агрессия, Международный терроризм* и др.
- Автоматическое аннотирование и др.
- Результаты обработки документа, включая пословный индекс, концептуальный индекс, вычисленные рубрики и т.п. загружаются в информационно-аналитическую систему.

Результаты обработки отдельного текста используются для построения компонентов различных информационных систем – информационного поиска, агрегирования новостных потоков (кластеризация), составления аналитических отчетов (различные типы обзорных рефератов) и т.д.

#### 4 Автоматическая рубрикация текстов на основе тематического представления

Основной современной технологией автоматической классификации текстов является подход на основе машинного обучения. Этот подход предполагает наличие текстовой коллекции достаточного объема для обучения алгоритмов. Однако в новой сложной задаче рубрикации текстов даже собственно рубрикатор, т.е. система категорий для рубрикации, может отсутствовать и должен быть создан с нуля или с использованием существующих похожих рубрикаторов.

В таких условиях более приемлемыми являются методы автоматической рубрикации, основанные на знаниях, т.е. на ручных правилах присвоения рубрик. Создавая рубрикатор и правила вывода рубрик в широкой предметной области, необходимо использовать поддержку тезауруса в написании правил, потому что тезаурус позволяет работать не с отдельными словами и выражениями, а с понятиями и подструктурами тезауруса [24].

Используемая процедура автоматической рубрикации текстов базируется на автоматически построенном тематическом представлении документов, которое моделирует основную тему и подтемы документа наборами (тематическими узлами) близких по смыслу понятий, упомянутых в документе [6, 24]. Такая основа рубрикации даёт возможность обрабатывать тексты разных типов и размеров: нормативные акты, газетные статьи, новостные сообщения, научные публикации в области гуманитарных наук, социологические опросы.

При создании лингвистического профиля рубрикатора каждая рубрика  $R$  описывается (см. рисунок 7) дизъюнкцией альтернатив, каждый дизъюнкт представляет собой конъюнкцию  $R = \bigcup_i D_i$  ;  $D_i = \bigcap_j K_{ij}$ . Конъюнкты в свою очередь описываются экспертами с по-

мощью так называемых «опорных» понятий. Для каждого опорного понятия задаётся правило его расширения  $f(\cdot)$ , определяющее, каким образом вместе с опорным понятием учитывать подчинённые ему по иерархии понятия: без расширения (обозначается символом «N»), полное расширение по дереву иерархии (символ «E»), расширение только по родовидовым связям (символ «L»), расширение по всем видам отношений на один уровень иерархии (символ «W»), расширение на один уровень иерархии, не включая отношения *ниже* (символ «V»).

Например, один из компонентов описания рубрики «Религиозная напряжённость» представляет собою конъюнкцию, первый элемент которой понятие *конфликт* без расширения (N), а второй элемент - понятия *религия* и *верующий* с расширением по видам (L), т.е. данная рубрика будет выводиться, если в тексте встречается упоминание понятия *конфликт* с любым понятием, относящимся к религиям, включая различные конфессии, верующих, религи-

озные догмы, ритуалы и т.п. Таким образом, с помощью тезауруса преодолевается проблема трудоёмкости инженерных методов рубрикации: если тезаурус уже существует, то описание рубрик, вывод которых основан на большом количестве разных слов, делается достаточно быстро за счёт иерархической организации тезауруса. Описание эксперимента по измерению времени создания описаний для рубрик рубрикатора и оценки качества получившейся системы рубрикации приводится в [24].

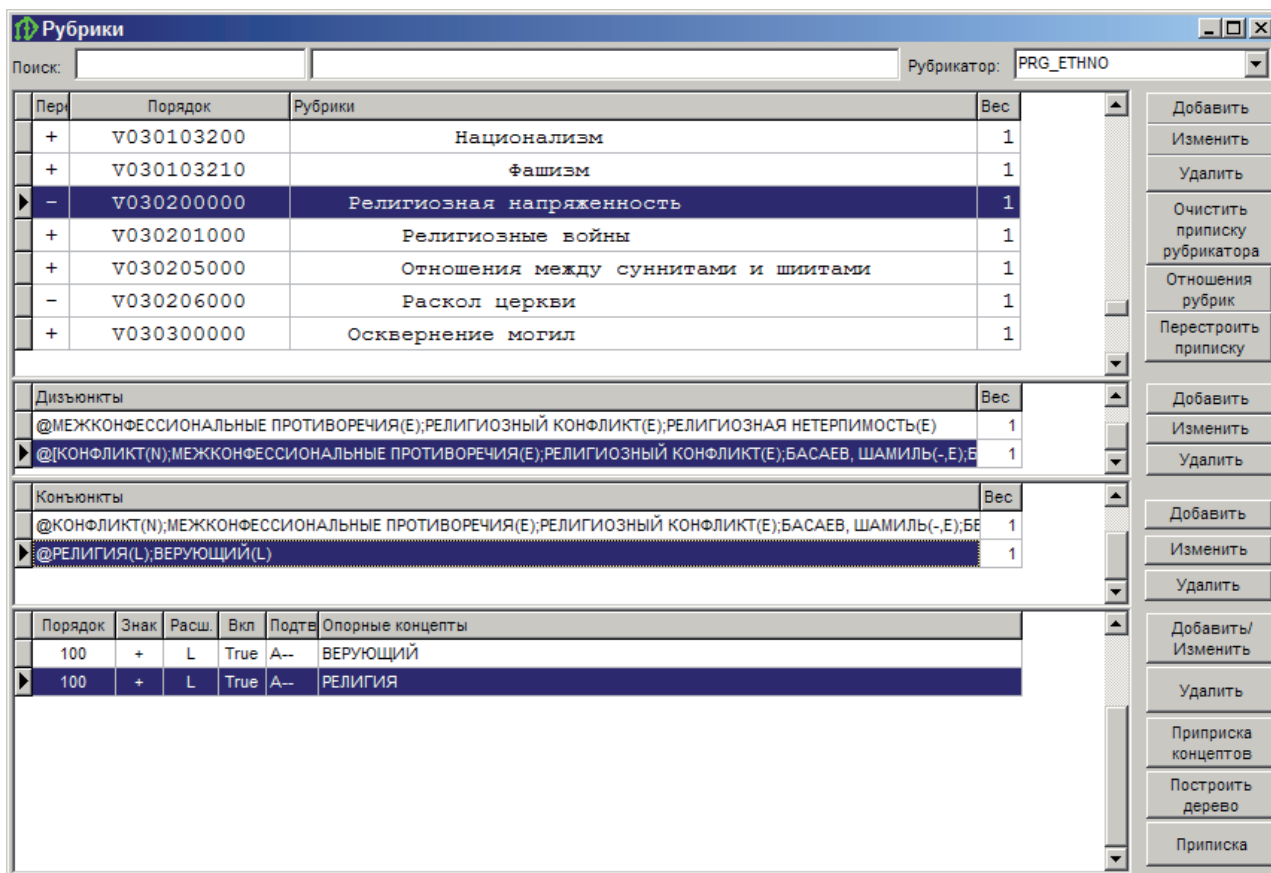


Рисунок 7 - Интерфейс описания содержания рубрик с помощью понятий тезауруса. Представление рубрики «Религиозная напряженность»

Вес конъюнкта зависит от максимального веса входящего в него понятия онтологии. Вес дизъюнкта предназначен учитывать не только сумму весов составляющих его конъюнктов, но и меру близости конъюнктов в тексте. Вес рубрики представляет собой максимум весов входящих в описание рубрики альтернатив. В случае имеющихся иерархических связей между рубриками оценка релевантности нижестоящих рубрик переносится на вышестоящие, так что при запросе по вышестоящей рубрике будут выходить и документы, к которым были приписаны нижестоящие рубрики.

Алгоритм рубрицирования работает следующим образом. Для всех понятий тезауруса, найденных в тексте, определяется множество рубрик, которые могут быть определены в тексте. В результирующем множестве остаются рубрики, вес которых превосходит задаваемый заранее для коллекции порог.

Для предметной области «Безопасность» разработано несколько рубрикаторов:

- рубрикатор угроз, описывающий существующие угрозы национальной безопасности (188 категорий, 5-уровневая иерархия);

- рубрикатор ценностей, представляющий индивидуальные, социальные и государственные ценности, например, свобода, демократия, права человека, семейные ценности и т. д. (109 категорий, иерархия 4 уровня);
- рубрикатор этно-конфессиональных отношений (94 категории, иерархия 4 уровня);
- рубрикатор региональных проблем (84 категории, 2-уровневая иерархия);
- рубрикатор регионов (325 регионов: субъекты Российской Федерации и иностранные государства).

## 5 Информационно-аналитическая система

Основным инструментом анализа больших коллекций текстовых документов является информационно-аналитическая система. В результате обработки отдельного текста выделяется большое количество объектов разных типов, которые традиционно привязываются к словопозициям (порядковому номеру слова в тексте). Результаты обработки загружаются в поисковые индексы информационно-поисковой системы.

Использование онтологий позволяет организовать работу аналитика более эффективно, как в части задания информационных потребностей, так и в части интерпретации получаемых от поисковой машины результатов.

### 5.1 Базовый поиск и организация поисковых индексов

В качестве поисковой машины авторы используют noSQL базу данных NearIdx 9.0, развиваемую Лабораторией информационных исследований совместно с Научно-исследовательским вычислительным центром МГУ. Традиционно используются два вида поисковых индексов: обратный и прямой, которые хранятся в специальных noSQL базах данных, поскольку накладные расходы на перестроение поисковых индексов для текстовых коллекций делают использование реляционных баз данных неэффективным.

Обратный (инверсный) индекс хранит информацию, организованную от элемента словаря данных к документу. Для каждого элемента данных  $E_i$  поддерживаются структуры данных по документам  $d_j$ , оценка релевантности документа  $\text{rank}(E_i, d_j)$  и вектор словопозиций  $[positions_{ijk}]$ .

$$[ E_i [d_j, \text{rank}(E_i, d_j), [positions_{ijk}] ] ]$$

Обратный индекс используется для поиска документов. В запросе пользователя к системе определяются элементы словарей данных. Для каждого элемента с жёсткого диска в оперативную память извлекаются указанные структуры, которые пересекаются между собой, образуя список документов, куда входят либо все элементы запроса при «строгом» поиске, либо не все при задании условий «нежёсткого/нечёткого» поиска.

Прямой индекс хранит ту же информацию, но организованную по документам:

$$[ d_j [E_i, \text{rank}(E_i, d_j), [positions_{ijk}] ] ]$$

Прямой индекс используется, например, для подсветки релевантных запросу фрагментов документов, что позволяет ускорить процесс отбора нужных пользователю документов. Также прямой индекс является основным средством поддержки для продвинутых средств анализа поисковой выборки, когда для первоначального запроса получена группа документов и требуется выбрать наиболее значимые элементы в данной выборке.

Для ускорения поддержки анализа временных рядов используются вспомогательные индексы по датировке публикации документов:

$$[ E_i [datetime_j, \text{doc\_count}(E_i, \text{rank}(E_i, d_j) > r_0) ] ]$$

Особенностью NearIdx является тесная интеграция с результатами обработки автоматической лингвистической обработки текстов (АЛОТ). Поддерживаются специальные поисковые индексы по словопозициям вхождения текстовых входов понятий лингвистических онтологий, в том числе с учётом иерархии понятий, а также пословные индексы, ассоциированные с описанием используемых рубрикаторов. Также поддерживаются индексы по выделённым именованным сущностям, тонально окрашенным словам и выражениям, прямой и косвенной речи и т.п.

Отметим, что поисковый индекс по понятиям с иерархией всего примерно в четыре раза превосходит по объёму индекс по понятиям без учёта иерархии (поиск по синонимам) и примерно равен по размеру пословному индексу с учётом морфологии.

Условия поиска задаются либо словами на обычном (естественном) языке, либо в следующем формате: /Словарь\_данных= «Элемент данных». Например, запись запроса "Наркотики в Российской Федерации" по данным публикаций октября 2017 года с использованием понятий онтологии будет выглядеть следующим образом (регистр не важен):

/Термин\_расш=«НАРКОТИК»

/Термин\_расш=«РОССИЙСКАЯ ФЕДЕРАЦИЯ»

/Дата\_док=«01.01.2017-31.10.2017»

Данный запрос позволит найти документы с упоминанием всех видов наркотиков и относящихся к наркотикам понятий, указанных в тезаурусе, с учётом всех регионов Российской Федерации.

При написании запроса допустимы логические операторы (отсутствие оператора эквивалентно логическому «И»), скобки для задания очередности обработки условий запроса, а также дополнительные контекстные операторы (поиск внутри одного или нескольких предложений, с начала или в конце текста, в окне заданного размера и т.д.).

Ключевым вопросом при обработке больших текстовых коллекций является время обработки запросов, которое при указанной организации определяется временем поиска и считывания информации с жёсткого диска. Время поиска по простым запросам по обратному индексу для коллекции 10 миллионов документов на стандартных современных компьютерах составляет 1-2 секунды.

Для поисковой машины недостаточно просто определить список релевантных документов. Необходимо сформировать выдачу, желательно с подсветкой релевантных фрагментов в найденных документах. Для сложно организованных поисковых индексов, например, для подсветки по рубрикам, требуется обращение к прямому индексу.

При этом интегральная (аналитическая) обработка первых 200 документов поисковой выдачи по прямому индексу составляет 0.3 секунды, 5 секунд для первых 2000 документов, а для анализа больших поисковых выборок может потребоваться значительное время. Поэтому стандартным решением для современных поисковых машин, в том числе для NearIdx, является очень быстрое получение именно первой страницы поисковой выдачи, но не всех страниц выдачи. Часто получение следующей страницы поисковой выдачи требует исполнения отдельного запроса.

## 5.2 Фасетный анализ

Для выборки документов доступны средства фасетного анализа по словарям данных. Можно установить количество наиболее релевантных запросу документов, по которым будет рассчитываться «информер» – специальная информационная панель, содержащая наиболее частотные или характерные для документов выборки элементы словаря данных. При этом возможны информеры по онтологии (см. рисунок 8) или по различным классификаторам (см. рисунок 9).

The screenshot shows a search interface with the following elements:

- Search bar: "Запрос (2) док. 50 Точность: точно (0.00) зона <Основная>"
- Filters: "отчет", "искать в найденном", "XML-граф", "отчет по сентименту", "расширенный поиск по кластерам", "демографические угрозы"
- Buttons: "Найти [Ctrl + Enter]", "ранжировать по релевантности ↑", "<ПО ВСЕМ>"
- Results summary: "Показано документов: 50 из 805 Время поиска: 00:00:00.073 (инфо)"
- Search results table:
 

Ранг/ИД	Местоположение / Снippet (еще поля)
0.85	<a href="#">Заголовки утра. 20 ноября</a> « Иммиграционный зов »: эксперты ЦСР назвали главные демографические «народные» облигации банков, рассказывается в материале «Купоны в поддержку продаж алкоголя в интернете». С нового года в сеть попадет разведут по-новому»:...
550815034	<b>Описание:</b> Канцлер британского казначейства Филип Хэммонд допустил финансовые предложения, связанные с отсоединением от ЕС, еще до гл... <i>The Guardian. В Британии нет безработных, неуместно заявил Хэммонд says Hammond in TV gaffe). Reuters: Американский ядерный генерал заявил, что он воспротивился бы "незаконному" приказу Трампа нанести удар» (U.S. nuclear general says would resist 'illegal' Trump strike order).</i>
0.85	<a href="#">Иммиграционный зов</a> Иммиграционный зов Иммиграционный зов Эксперты ЦСР назвали главные демографические угрозы России Эксперты Центра стратегических разработок Алексея Кудрина сформулировали основные демографические вызовы в России. Чтобы преодолеть их, члены экспертного совета предлагают привлекать еще больше мигрантов и давать им гражданство Фото: Владимир...
550795021	<b>Описание:</b> Иммиграционный зов Иммиграционный зов Эксперты ЦСР назвали главные демографические угрозы России Эксперты Центра стратегических разработок Алексея Кудрина сформулировали основные демографические вызовы в России. Чтобы решить их, нужно привлечь больше мигрантов, а россиянам придется повысить порог социальной терпимости, заключили аналитики. При этом авторы доклада отмечают, что прекращение роста населения является
- TERMIN list:
 

+	-	+t	-t	ТЕРМИН
+	-	+t	-t	ДЕМОГРАФИЧЕСКАЯ ОБСТАНОВКА
+	-	+t	-t	УБЫЛЬ НАСЕЛЕНИЯ
+	-	+t	-t	АБОРТ
+	-	+t	-t	РОЖДАЕМОСТЬ
+	-	+t	-t	ПРИРОСТ НАСЕЛЕНИЯ
+	-	+t	-t	ДЕМОГРАФИЧЕСКИЙ ПОКАЗАТЕЛЬ
+	-	+t	-t	ДЕМОГРАФИЧЕСКИЙ ПРОГНОЗ
+	-	+t	-t	СОКРАЩЕНИЕ РОЖДАЕМОСТИ
+	-	+t	-t	ЧИСЛЕННОСТЬ НАСЕЛЕНИЯ
+	-	+t	-t	ВОЕННАЯ КОНФРОНТАЦИЯ
+	-	+t	-t	МИЗУЛИНА, ЕЛЕНА БОРИСОВНА
+	-	+t	-t	ДЕМОГРАФ
+	-	+t	-t	ГЕОПОЛИТИКА
+	-	+t	-t	НАЦИОНАЛЬНАЯ БЕЗОПАСНОСТЬ
+	-	+t	-t	ЕСТЕСТВЕННАЯ УБЫЛЬ НАСЕЛЕНИЯ
+	-	+t	-t	ИНСТИТУТ СЕМЬИ
- ИНФОРМЕР: 16

Рисунок 8 - Информер показывает наиболее характерные понятия онтологии для текущего запроса «Демографические угрозы»

The figure consists of three tables:

РЕГИОНЫ	УГРОЗЫ	ЦЕННОСТИ
+	+	+
-	-	-
[796] СТРАНЫ МИРА	[209] СОЦИАЛЬНАЯ НАПРЯЖЕННОСТЬ, РАСКОЛ В	[77] ГОСУДАРСТВЕННОЕ УСТРОЙСТВО
+	+	+
-	-	-
[756] РОССИЙСКАЯ ФЕДЕРАЦИЯ	[149] СЛАБОСТЬ ВЛАСТЕЙ	[57] ЭКОНОМИЧЕСКАЯ БЕЗОПАСНОСТЬ
+	+	+
-	-	-
[653] СУБЪЕКТЫ РФ	[134] КРИМИНАЛЬНАЯ ОБСТАНОВКА	[53] ДУХОВНЫЕ ЦЕННОСТИ
+	+	+
-	-	-
[352] МОСКВА	[132] ПРОТЕСТНАЯ АКТИВНОСТЬ	[49] НРАВСТВЕННЫЕ ЦЕННОСТИ
+	+	+
-	-	-
[290] КИТАЙ	[122] ЭКОНОМИЧЕСКИЕ УГРОЗЫ	[42] ЭКОНОМИЧЕСКИЙ РОСТ
+	+	+
-	-	-
[279] США	[121] ПОЛИТИЧЕСКИЙ КРИЗИС	[38] ПАТРИОТИЗМ
+	+	+
-	-	-
[234] УКРАИНА	[110] КСЕНОФОБИЯ	[34] ГОСУДАРСТВЕННЫЙ СУВЕРЕНИТЕТ
+	+	+
-	-	-
[197] ГЕРМАНИЯ	[109] ЭКСТРЕМИЗМ	[32] ПРАВА ЧЕЛОВЕКА
+	+	+
-	-	-
[197] КАЗАХСТАН	[104] НЕЗАКОННАЯ МИГРАЦИЯ	[31] ГОСУДАРСТВЕННЫЕ ИНТЕРЕСЫ
+	+	+
-	-	-
[169] ВЕЛИКОБРИТАНИЯ	[101] ОРГАНИЗОВАННАЯ ПРЕСТУПНОСТЬ	[29] ЕВРОПЕЙСКИЕ(ЗАПАДНЫЕ) ЦЕННОСТИ
+	+	+
-	-	-
[162] САНКТ-ПЕТЕРБУРГ	[95] СОЦИАЛЬНАЯ ДИФФЕРЕНЦИАЦИЯ НАСЕЛЕНИЯ	[25] ГЕОПОЛИТИЧЕСКИЕ ИНТЕРЕСЫ
+	+	+
-	-	-
[161] УЗБЕКИСТАН	[89] ДИСКРИМИНАЦИЯ	[20] ЗДОРОВЬЕ
+	+	+
-	-	-
[153] ФРАНЦИЯ	[88] НАРКОВБИЗНЕС	[20] ЛИЧНЫЕ И СОЦИАЛЬНЫЕ ЦЕННОСТИ
+	+	+
-	-	-
[145] БЕЛОРУССИЯ	[84] ВНЕШНИЕ УГРОЗЫ	[20] ЕВРОПЕЙСКИЕ ЦЕННОСТИ. СОЦИАЛЬНАЯ СП
+	+	+
-	-	-
[142] ТАДЖИКИСТАН	[80] МАССОВЫЕ ВОЛНЕНИЯ, ДРАКИ	[18] ПОЛИТИЧЕСКИЕ СВОБОДЫ
+	+	+
-	-	-
[141] СИРИЯ	[74] ЧРЕЗВЫЧАЙНЫЕ СИТУАЦИИ	[14] КУЛЬТУРНО-ИСТОРИЧЕСКИЕ ЦЕННОСТИ
+	+	+
-	-	-

а) Информер по упоминаемым регионам      б) Информер по упоминаемым угрозам      в) Информер по упоминаемым ценностям

Рисунок 9 - Информеры по разным классификаторам для запроса «Демографические угрозы»

На рисунке 9 показаны наиболее характерные категории классификаторов для запроса «Демографические угрозы»: регионы (а), угрозы (б), ценности (в). Кнопки «+» и «-» позволяют «одним кликом мыши» добавить соответствующее условие в запрос.

Обеспечение полноты поиска является фундаментальной проблемой из-за естественного разнообразия языка. Поиск по понятиям онтологии равносителен поиску с учётом синонимии, поиск по понятиям онтологии с учётом расширения по иерархии (когда релевантными считаются документы, содержащие синонимы выбранного понятия или его подчинённых по иерархическим связям) может быть весьма эффективным: запрос выглядит компактно, а результаты полны. Однако опыт эксплуатации различных информационно-поисковых систем выявил проблемы пользователей, которым трудно подобрать нужное понятие в большой понятийно-терминологической сети.

Реализация информера по понятиям онтологии (информер «Термины») позволяет эффективно решать данную проблему в динамике общения с информационно-поисковой системой. Действительно, пользователь может быстро исполнить запрос на естественном языке по интересующей его теме. Результаты поиска и информеры получаются очень быстро. При этом нужные понятия содержатся в информере. Далее достаточно одного-двух кликов «мыши» для выбора нужных понятий, которые и обеспечивают полноту поиска.

### 5.3 Спектрально-фасетный анализ временных рядов

Спектрально-фасетный анализ временных рядов темпорального распределения документов поисковой выдачи реализуется программным обеспечением SpectralView. Результаты выводятся в виде графика количества публикаций по разным запросам в зависимости от времени, при необходимости с требуемым уровнем агрегации (по часам/ дням/ неделям/ месяцам).

Автоматически, аналогично XYZ-статистикам, рассчитываются наиболее значимые факторы, графики которых наносятся на основной график. Для потоков текстовых документов задача определения XYZ-статистики естественным образом аналогична задаче выявления основных факторов, которые действуют на всём времени выборки документов или, наоборот, объясняют появление тех или иных пиков на временном графике.

Имеется возможность наложить несколько иерархий лингвистической онтологии, чтобы осуществлять выбор статистик только из заданных иерархий. Этим удобно пользоваться, например, чтобы выделять наиболее значимые именованные сущности: регионы, города и т.д. Для аналитика также полезна фильтрация факторов по иерархиям абстрактных понятий, например, по видам преступлений, видам болезней, видам зданий и т.п. Для удобства некоторые линии графиков можно выделять, другие, наоборот, затенять.

На рисунке 10 приведены данные анализа выборки (по корпусу российских новостей за 2017 год, 9 миллионов документов) по запросу: /Рубрика=«ЭКОЛОГИЧЕСКИЕ УГРОЗЫ». Как нетрудно видеть, резонанс обсуждения по теме в значительной мере формируется текстами, в которых обсуждаются пожары и лесные пожары.

Для любой точки графика можно, быстро исполнив дополнительный запрос, получить список документов, релевантный исходному запросу, выбранному элементу словаря данных и опубликованному в выбранном временном интервале. В частности, легко узнать, что пик на рисунке 10 в конце апреля объясняется публикациями по годовщине аварии на Чернобыльской АЭС 26.04.1986 г.

На рисунке 11 нанесены графики распределения по субъектам РФ публикаций с упоминанием тематики лесных пожаров (запрос /Термин\_расш=«ЛЕСНОЙ ПОЖАР»). Согласно этим данным наибольший резонанс в марте-августе 2017 года вызывали лесные пожары в восточной Сибири – Иркутской области, Бурятии, Забайкальском крае.

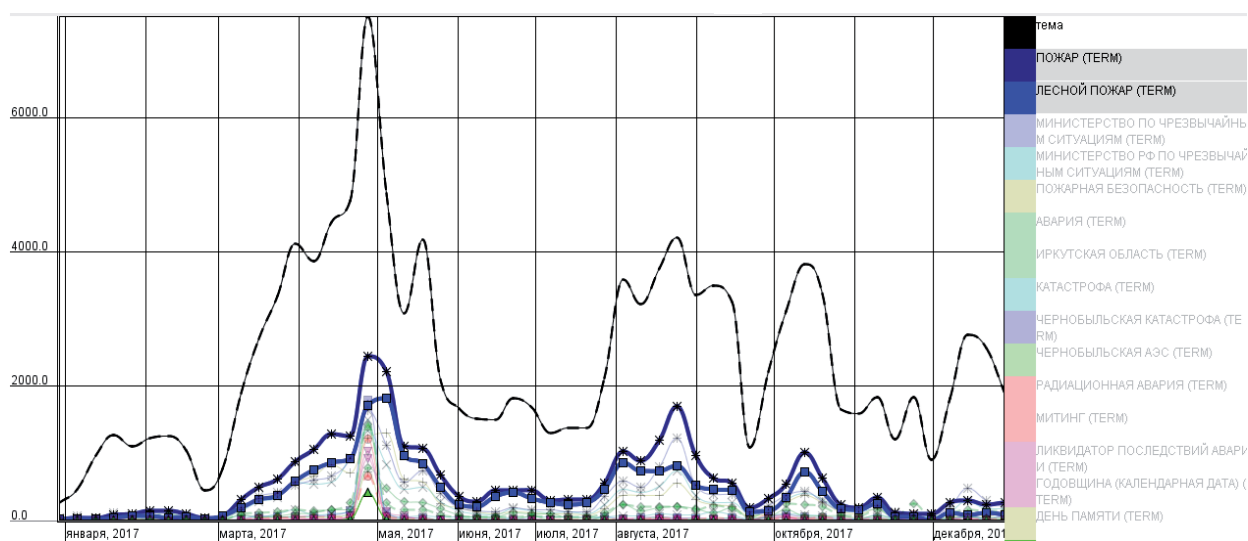


Рисунок 10 - Анализ запроса /Рубрика=«ЭКОЛОГИЧЕСКИЕ УГРОЗЫ»

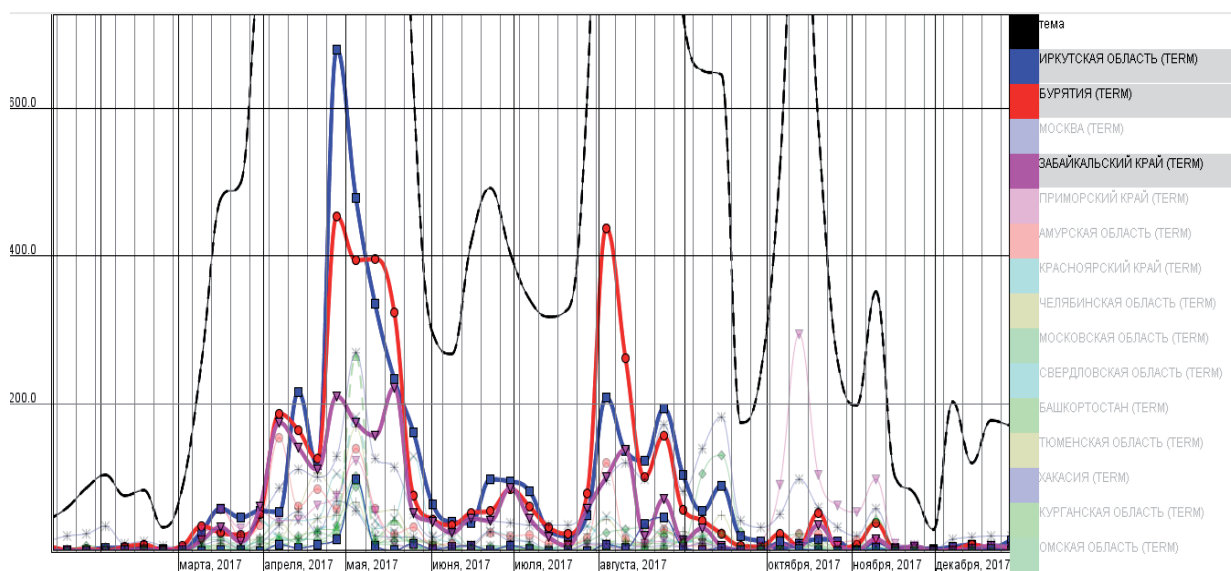


Рисунок 11 - Распределение по субъектам РФ упоминания в новостях 2017 года лесных пожаров

#### 5.4 Когнитивные схемы

Фасеты позволяют выявить основные элементы словарей данных, наиболее характерные для поисковых выборок. Спектрально-фасетный анализ позволяет анализировать вклад в содержимое выборок разных элементов словарей данных в динамике.

Между факторами, определяющими закономерности развития исследуемого процесса или явления могут существовать достаточно сложные зависимости. Для описания сложных зависимостей полезно использовать графовое представление. В среде аналитиков широко известен программный продукт IBM i2<sup>1</sup>, предназначенный для исследования взаимосвязей между именованными сущностями. При этом сила связи между сущностями является монотонной функцией от количества документов, в тексте которых упоминаются обе сущности.

<sup>1</sup> <https://www.ibm.com/ru-ru/marketplace/analysts-notebook>



В рамках данной информационно-аналитической системы используется программное обеспечение визуализации графов GraphView, тесно увязанное с поисковой машиной NearIdx (см. рисунок 12). В NearIdx хранится значительно больше информации, чем простые индексы по именованным сущностям. Соответственно в GraphView граф может включать элементы разных словарей данных.

Граф на основе выборки формируется в NearIdx, при этом имеется возможность выбора различных фасетов, а также любых других сущностей словарей данных, других запросов для отображения на графе. После чего граф разрисовывается и модифицируется в GraphView.

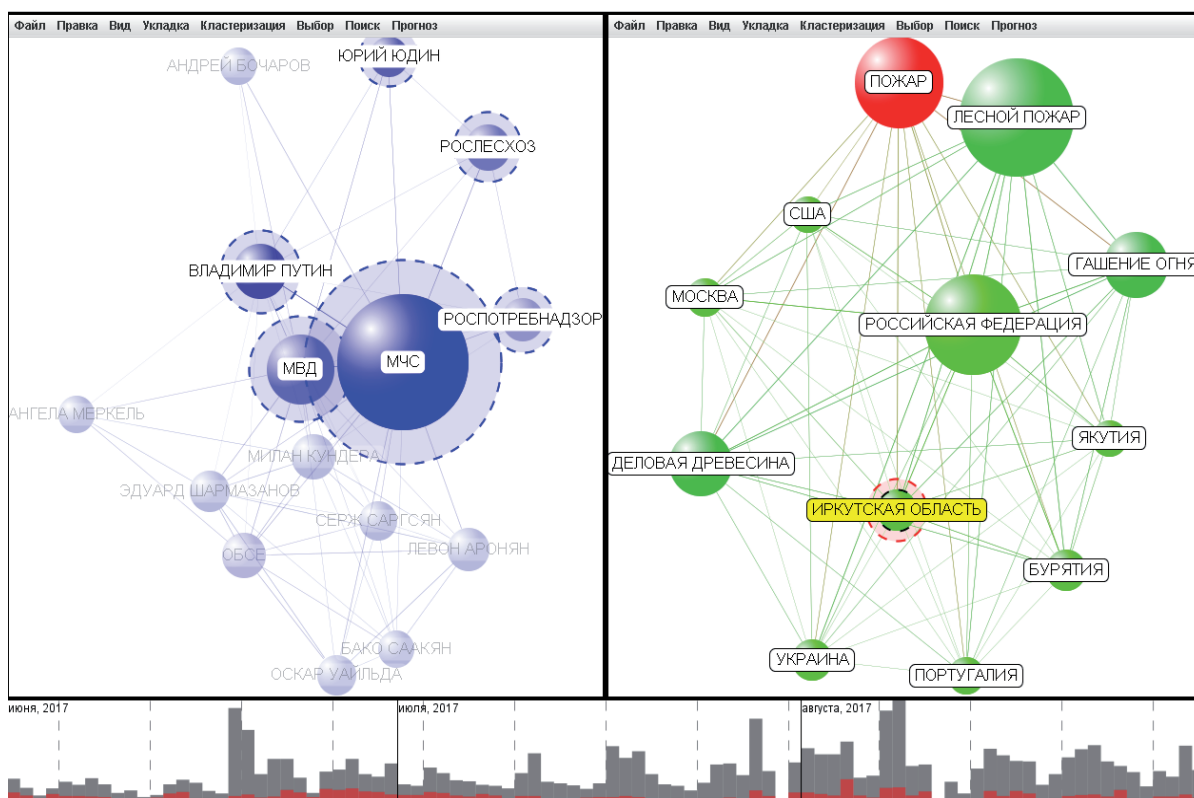


Рисунок 12 - Представление содержания документов по тематике лесных пожаров.

На левой панели представлены упоминаемые персоны и организации, на правой панели – понятия онтологии

GraphView разрабатывался для поддержки труда аналитиков, для которого характерны отработка задач выдвижения гипотез и их быстрой проверки. Так как одной из проблем при аналитическом исследовании графов является наличие большого количества связей между объектами, то в GraphView для лучшего восприятия информации содержимое графа разделено на две панели. Рёбра между объектами на разных панелях не отображаются, но информация о них хранится, и взаимосвязи между объектами на разных панелях визуализируются выделением яркостью цвета между связанными узлами и затемнением несвязанных узлов графа. Таким образом, поддерживается исходный «полный» граф, который виртуально разделяется на несколько отображаемых визуальными панелями. При этом узлы графа можно перемещать/копировать с одного подграфа на другой, что полезно, например, для анализа сравнительно небольшого подграфа.

Ниже двух панелей приводится временной ряд количества документов выборки, на котором визуализируется количество документов по выбранным сущностям.

Опыт показывает, что при анализе по именованным сущностям и лингвистическим онтологиям удобно размещать именованные сущности на одном подграфе, а понятия онтологии -

на другом. Тогда один подграф «отвечает» на вопрос «о ком говорилось в текстах выборки», а второй подграф отвечает на вопрос «что говорилось в текстах выборки». При этом выбор одного или нескольких объектов на одном подграфе приводит к обозначению связанных объектов на обоих подграфах на фоне затеняемых несвязанных объектов (рисунок 12).

## 5.5 Аналитические справки

Изучение стандартных аналитических отчетов, а также отчетов по мониторингу прессы, показывает, что они имеют схожую, достаточно простую структуру. Отчеты посвящены одной теме и, при необходимости, структурируются по частям, относящимся к другим темам. При этом единицей информации являются фрагменты исходных документов. Такая структура может быть смоделирована с использованием качественной поисковой машины.

Действительно, если рассмотреть заглавный запрос  $Q_0$  и множество структурирующих запросов  $Q_i$  (будем называть их «подрубрикатором»), тогда общий запрос вида  $Q_0 \& Q_i$  должен вернуть все релевантные документы (фрагменты). После этого остаётся только избавиться от повторов (например, кластеризовав фрагменты) и расположить фрагменты в соответствии с заданной последовательностью дополнительных запросов.

Рассмотрим в качестве примера запрос по рубрике (рисунок 13):

/УГРОЗЫ= «Демографические угрозы»

В данном автоматически построенном отчете полученные фрагменты (релевантными считаются фрагменты длиной не более 3 предложений), сгруппированы по сходству и выведены по убыванию даты публикации.

### Отчет по запросу: /УГРОЗЫ="ДЕМОГРАФИЧЕСКИЕ УГРОЗЫ"

Дата отчета: 2017-12-18 20:39:24

Формирование отчета: по документам

Сортировка: по убыванию даты

#### (0.74) 28.11.2017 22:02:22 На выплаты первенцам государство потратит за три года 144 млрд рублей [Полит.ру]

На выплаты первенцам государство потратит за три года 144 млрд рублей На выплаты первенцам государство потратит за три года 144 млрд рублей Младенцы Советник Института современного развития Никита Масленников заявил РИА Новости, что объявленная президентом РФ Владимиром Путиным перезагрузка в демографической политике отвечает чаяниям россиян и поможет исправить ситуацию с рождаемостью в стране. Речь в частности, идет о предложении российского лидера ввести ежемесячные социальные выплаты на первенцев. «В целом эти инициативы абсолютно оправданы, во-первых, потому, что они просто необходимы по жизни, надо думать о том, как мы будем вылезать из этой демографической ямы, как стабилизировать численность населения страны на ближайшие 20 лет.

#### (0.69) 28.11.2017 19:13:05 Эксперт оценил пользу социальных предложений Путина [РИА Новости]

Эксперт оценил пользу социальных предложений Путина МОСКВА, 28 ноя — РИА Новости. Инициативы в социальной сфере, озвученные президентом РФ Владимиром Путиным во вторник, помогут России выбраться из демографической "ямы" и ответить на социальные ожидания населения в начале нового политического цикла, считает советник Института современного развития Никита Масленников. Президент РФ Владимир Путин во вторник озвучил ряд инициатив в социально-демографической сфере. Тем более что есть там и переходящие деньги на следующий год, и определенный резерв на различного рода меры социальной и иной поддержки заложены и на 2018 год и очевидно на следующий год", — рассказал эксперт. "Поэтому здесь, я думаю, с источниками должно быть все нормально, тем более понятна целевая функция этих мер, потому что мы уже в следующем году сталкиваемся с достаточно серьезной ситуацией на рынке труда, когда примерно 300 тысяч человек с него уйдет и дальше до 2024 года по меньшей мере у нас должны постоянно сокращения с рынка труда, от 300 до 600 тысяч. Поэтому естественно, демографическая проблема требует повышенного внимания с учетом того, что нам нужно стимулировать деторождаемость и рост населения с тем, чтобы на рубеже 2030-2035 годов более-менее выровнять ситуацию", — отметил Масленников.

#### (0.69) 28.11.2017 11:06:17 Население Омской области вымирает [КоммерческиеВести]

Население Омской области вымирает 0 75 С января по октябрь в регионе число умерших было на 2296 человек больше, чем родившихся. В прошлом году отмечался естественный прирост населения на 521 человека. Омкстат обнародовал данные о естественном движении населения региона в январе-октябре 2017 года. **+ 1 фрагм.**

#### (0.69) 27.11.2017 20:06:25 Каждый второй нижегородец умирает от болезни системы кровообращения [IA REGNUM]

Каждый второй нижегородец умирает от болезни системы кровообращения Эдвард Мунк. У смертного одра. 1893 Нижний Новгород, 27 ноября 2017. 19:41 — REGNUM Болезни систем кровообращения остаются самыми частыми причинами смерти среди жителей

Рисунок 13 – Отчёт по запросу «Демографические угрозы»

На рисунке 14 представлен пример формы задания параметров построения аналитической справки с заданным подрубрикатором по регионам Сибирского федерального округа. Для задания разделов подрубрикатора указывается порядок, название раздела и запрос, который определяет наполнение данного раздела. На рисунке 15 представлен результат построения аналитической справки, структурированной по регионам.

Таким образом, созданная информационно-аналитическая система NearIdx обеспечивает стандартные функции информационного поиска, а также даёт возможность задания запросов и поиска информации с использованием специализированных лексико-терминологических ресурсов (онтологий, рубрикаторов). Аналитический компонент NearIdx предоставляет возможности фасетного анализа, спектрально-фасетного анализа, построения когнитивных схем и аналитических справок, в которых также могут использоваться созданные ресурсы.

The screenshot shows the 'Запрос (?)' (Query) configuration page. At the top, there are fields for 'док.' (50), 'Точность:' (точно (0.00)), 'зона' (<Основная>), 'HTML', 'инф.' (100), and 'м.в.' (↑). Below this is a search input field containing '0'. A row of checkboxes includes 'отчет' (checked), 'расширение запроса', 'параметры', and 'справка по кластерам', with a 'Меню' button to the right. The main query field contains '/УГРОЗЫ="ДЕМОГРАФИЧЕСКИЕ УГРОЗЫ"'. The 'Параметры отчета' (Report Parameters) section includes: 'Первые N док.' (500), 'Макс. окно' (150), 'Предложений' (3), 'Макс. фрагментов' (1000), 'Сбор по предложениям в случае отсутствия окна' (unchecked), 'Использовать рубрику "Прочее"' (unchecked), 'Сдвиг первого предложения не дальше' (10000), 'словопозиций', 'Мин. вес' (0.01), 'Не искать фрагменты (использовать полный текст документа)' (unchecked), 'Подсветка в отчете' (checked), 'Кластеризация фрагментов' (checked), 'Без учета запроса' (checked), and 'Сохранить фрагменты без построения' (unchecked). There are radio buttons for 'Классификатор из списка' (selected) and 'Классификатор в виде текста'. The list shows: '03100 КЕМЕРОВСКАЯ ОБЛАСТЬ=/ТЕРМИН\_расш="КЕМЕРОВСКАЯ ОБЛАСТЬ"', '03900 КРАСНОЯРСКИЙ КРАЙ=/ТЕРМИН\_расш="КРАСНОЯРСКИЙ КРАЙ"', '05200 НОВОСИБИРСКАЯ ОБЛАСТЬ=/ТЕРМИН\_расш="НОВОСИБИРСКАЯ ОБЛАСТЬ"', '05500 ОМСКАЯ ОБЛАСТЬ=/ТЕРМИН\_расш="ОМСКАЯ ОБЛАСТЬ"', '07800 ТОМСКАЯ ОБЛАСТЬ=/ТЕРМИН\_расш="ТОМСКАЯ ОБЛАСТЬ"', and '08000 РЕСПУБЛИКА ТЫВА=/ТЕРМИН\_расш="ТЫВА /ТУВА/"'. Below this is 'Классификатор по тезаурусу' with 'Разделов' (10) and 'Расш.' (E). A 'Все' button is present. The 'Результат:' section has radio buttons for 'по кластерам' and 'по документам' (selected), and a 'Сортировка' dropdown set to 'по убыванию даты'. The 'Дополнительный запрос' field contains '+/- предложений' (2). At the bottom, there is a 'Найти [Ctrl + Enter]' button, a 'ранжировать' dropdown set to 'по релевантности ↑', and a 'Выбрать' button.

Рисунок 14 – Пример формы задания исходного запроса: рубрика «Демографические угрозы» из рубрикатора угроз (в качестве подрубрик указаны некоторые субъекты Российской Федерации)

## Заключение

В статье рассмотрен подход к описанию широкой области национальной безопасности как тезауруса для автоматической обработки документов. Созданный Тезаурус по безопасности имеет модель представления тезауруса RuТез, используется в специализированной информационно-аналитической системе и для автоматической текстовой классификации документов в соответствии с несколькими рубрикаторами, включая рубрикатор угроз, рубрикатор ценностей, рубрикатор региональных проблем и др.

Созданная информационно-аналитическая система NearIdx обеспечивает стандартные функции информационного поиска, а также даёт возможность задания запросов и поиска информации с использованием специализированных ресурсов, включая онтологию и рубри-

каторы. Аналитический компонент NearIdx предоставляет возможности фасетного анализа, спектрально-фасетного анализа, построения когнитивных схем и аналитических справок, в которых также могут использоваться созданные лексико-терминологические ресурсы.

## Отчет по запросу: /УГРОЗЫ="ДЕМОГРАФИЧЕСКИЕ УГРОЗЫ"

Дата отчета: 2017-12-18 20:47:09

Формирование отчета: по документам

Сортировка: по убыванию даты

### 00600 РЕСПУБЛИКА АЛТАЙ

**(0.56)** 18.10.2017 06:01:11 Республика Алтай вошла в пятерку регионов-лидеров по рождаемости [БезФормата.Ru Республика Алтай Горно-Алтайск]

Республика Алтай вошла в пятерку регионов-лидеров по рождаемости Фото: [www.gorno-altaisk.info](http://www.gorno-altaisk.info) На фоне всеобщего снижения показателей рождаемости в целом по стране Республика Алтай продолжает лидировать по этому показателю и входит в пятерку регионов-лидеров, рассказал министр здравоохранения региона Владимир Пелеганчук. Он отметил, что показатель рождаемости по итогам восьми месяцев 2017 года составил 15,9 на 1000 населения, в абсолютных цифрах – 2299 человек. «В этом году у нас, как и в целом по России, отмечается снижение рождаемости.

### 00700 АЛТАЙСКИЙ КРАЙ

**(0.57)** 01.03.2017 08:17:54 Барнаул впервые за пять лет пережил сокращение численности населения [Информационное агентство АМИТЕЛ]

Барнаул впервые за пять лет пережил сокращение численности населения По итогам 2016 года количество жителей краевой столицы уменьшилось на 2,1 тысячи человек Барнаул в 2016 году впервые за последние пять лет пережил сокращение численности населения, передает "Интерфакс" со ссылкой на данные Росстата. Отмечается, что за минувший год количество барнаульцев уменьшилось на 2,1 тысячи человек — до 633,5 тысячи горожан. Ранее портал Amic.ru также сообщал, что отрицательная динамика в Барнауле наблюдалась и в разрезе миграционных процессов.

**(0.5)** 15.02.2017 20:20:16 Население Алтайского края сократилось на один район в 2016 году [Сетевое издание "ВладТайм"]

Население Алтайского края сократилось на один район в 2016 году В Алтайском крае за прошедший год населения стало меньше на 6472 человека. Приблизительно такое же число людей проживает в Ельцовском и Суевском районах региона. В 2016 году из области выехало 38 713 граждан, заехало 32 241 человек.

### 01400 РЕСПУБЛИКА БУРЯТИЯ

**(0.51)** 28.09.2017 16:13:04 В Бурятии работает каждый пятый пенсионер [Байкал-Daily]

В Бурятии работает каждый пятый пенсионер Одновременно с процессом старения населения в Бурятии растёт численность пенсионеров

Рисунок 15 - Отчет по запросу «Демографические угрозы» по нескольким регионам России

## Благодарности

Работа частично поддержана грантом РФФИ (проект 16-29-09606) и контрактом Министерства образования и науки Российской Федерации (№ 14.601.21.0018).

## Список источников

- [1] Abbas, A., Zhang, L., Khan, S. (2014). A literature review on the state-of-the-art in patent analysis / A. Abbas, L., Zhang, L., S. Khan. // World Patent Information. – 2014. – V. 37. – P.3-13.
- [2] Efimenko, I. Peaks, Slopes, Canyons and Plateaus: Identifying Technology Trends Throughout the Life Cycle / I. Efimenko, V Khoroshevsky // International Journal of Innovation and Technology Management. – 2017. – 14(02).
- [3] Ena, O. A methodology for technology trend monitoring: the case of semantic technologies / O. Ena, N. Mikova, O, Saritas, A. Sokolova // Scientometrics. – 2016. – 108 (3). – P.1013-1041.
- [4] Nassirtoussi, A.K. Text mining for market prediction: A systematic review / A.K. Nassirtoussi, S, Aghabozorgi, T.Y. Wah, D.C, Ngo. // Expert Systems with Applications. – 2014. – V. 41 (16). – P.7653-7670.
- [5] Трошин, Д.В. Основы концептуальной модели источников угроз экономической безопасности на национальном уровне / Д.В. Трошин // Онтология проектирования. – 2017. – Т. 7, № 4(26). – С.410-422. – DOI: 10.18287/2223-9537-2017-7-4-410-422.

- [6] *Лукашевич, Н.В.* Тезаурусы в задачах информационного поиска / Н.В. Лукашевич. – М.: Издательство МГУ, 2011.
- [7] *Shane, S.* Isis displaying a deft command of varied media / *S. Shane, B. Hubbard* // New York Times 31. – 2014.
- [8] *Weimann, G.* New Terrorism and New Media / *G. Weimann* // Wilson Center Common Labs. – 2014.
- [9] *Kohlmann, E.* Profiles of foreign fighters in Syria and Iraq / *E. Kohlmann, L. Alkhouri* // CTC Sentinel, September. – 2014. – V. 29.
- [10] *Finlayson, M.A.* The N2 corpus: A semantically annotated collection of Islamist extremist stories / *M.A. Finlayson, J.R. Halverson, S.R. Corma*, // In LREC-2014. – 2014. – P.896-902.
- [11] *Sela, S.* Changes in the discourse of online hate blogs: The effect of Barack Obama's election in 2008 / *S. Sela, T. Kuflik, G.S. Mesch*. // First Monday. – 2012. – V.17, N.11.
- [12] *Kwok, I.* 2013. Locate the hate: Detecting tweets against blacks / *I. Kwok, Y. Wang* // In AAAI-2013. – 2013.
- [13] *Zeeraq, W.* Hateful symbols or hateful people? predictive features for hate speech detection on twitter / *W. Zeeraq, D. Hovy* // Proceedings of NAACL-HLT-2016. – 2016. – P.88-93.
- [14] *Nobata, Ch.* Abusive language detection in online user content / *Ch. Nobata, J. Tetreault, A. Thomas, M. Yashar, Yi Chang* //, in Proc. of the 25th International Conference on World Wide Web. – 2016. – P.145-153.
- [15] *Schmidt, A.* A survey on hate speech detection using natural language processing / *A. Schmidt, M. Wiegand* // SocialNLP-2017. – 2017.
- [16] *Lim, S.* MalwareTextDB: A Database for Annotated Malware Articles / *S. Lim, A. Muis, W. Lu, C. Ong* // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. – 2017. – V.1.
- [17] *Kirillov, I.* 2010. Malware Attribute Enumeration and Characterization / *I. Kirillov, D. Beck, P. Chase, R. Martin* // The MITRE Corporation, Tech. Rep. – 2010.
- [18] *Gorokhov, O.* Convolutional Neural Networks for Unsupervised Anomaly Detection in Text Data / *O. Gorokhov, M. Petrovskiy, I. Mashechkin* // International Conference on Intelligent Data Engineering and Automated Learning. – Springer, Cham, 2017. – P.500-507.
- [19] *O'brien, S.P.* Crisis early warning and decision support: Contemporary approaches and thoughts on future research / *S.P. O'brien*. // International studies review. – 2010. – 12(1). – P.87-104.
- [20] *Wang, W.* Growing pains for global monitoring of societal events. / *W. Wang, R. Kennedy, D. Lazer, N. Ramakrishnan* // Science. – 2016. – 353(6307). – P.1502-1503.
- [21] *Лукашевич, Н.В.* Проектирование лингвистических онтологий для информационных систем в широких предметных областях / Н.В. Лукашевич, Б.В. Добров // Онтология проектирования. – 2015. – Т. 5. – №. 1 (15) – С.47-69.
- [22] *Loukachevitch, N.* RuThes linguistic ontology vs. Russian wordnets / *N. Loukachevitch, B. Dobrov* // Proceedings of Global WordNet Conference GWC-2014. 2014.
- [23] *Loukachevitch, N.* RuThes-Lite, a publicly available version of Thesaurus of Russian language RuThes / *N. Loukachevitch, B. Dobrov, I. Chetviorkin* // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference в Dialogue-2014. – 2014. – P.340-349.
- [24] *Loukachevitch, N.* The Sociopolitical Thesaurus as a resource for automatic document processing in Russian / *N. Loukachevitch, B. Dobrov* // Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication. – 2015. – V. 21, N.2. – P.237-262.
- [25] *Dobrov, B.* Development of Linguistic Ontology on Natural Sciences and Technology / *B. Dobrov, N. Loukachevitch* // Proceedings of Linguistic Resources and Evaluation Conference LREC-2006. – 2006.

## ONTOLOGICAL RESOURCES AND INFORMATION-ANALYTICAL SYSTEM IN SECURITY DOMAIN

N.V. Loukachevitch<sup>1</sup>, B.V. Dobrov<sup>2</sup>, A.M. Pavlov<sup>3</sup>, S.V. Shternov<sup>4</sup>

Research Computing Center of Lomonosov Moscow State University, Moscow, Russia

<sup>1</sup>louk\_nat@mail.ru, <sup>2</sup>dobrov\_bv@mail.ru, <sup>3</sup>pavlov.andrew.m@gmail.com, <sup>4</sup>shternov@gmail.com

### Abstract

The article considers an approach to describing a broad area of national security as a thesaurus for automatic document processing. The created thesaurus of safety has a model of representation of the thesaurus RuThes. The security thesaurus is used in a specialized information and analytical system and for automatic text classification of documents in ac-

cordance with several systems of subject headings, including the Threat categories, the Value, Regional problem categories, etc. The created Information and Analytical System NearIdx provides standard information search functions, and it also the ability to specify queries and search for information using specialized resources, including ontology and subject headings. The analytical component NearIdx provides the possibilities of facet analysis, spectral-facet analysis, construction of cognitive schemes and analytical references, in which the ontological resources can also be used.

**Key words:** *thesaurus, ontology, national security, information retrieval, subject headings, automatic text categorization.*

**Citation:** *Loukachevitch NV, Dobrov BV, Pavlov AM, Shternov SV. Ontological Resources and Information-Analytical System in Security Domain [In Russian]. *Ontology of designing*. 2018; 8(1): 74-95. - DOI: 10.18287/2223-9537-2018-8-1-74-95.*

## Acknowledgment

The work was partially supported by the RFBR grant (project 16-29-09606) and the contract of the Ministry of Education and Science of the Russian Federation (No. 14.601.21.0018).

## References

- [1] **Abbas A, Zhang L, Khan S.** A literature review on the state-of-the-art in patent analysis. *World Patent Information*; 2014; 37: 3-13.
- [2] **Efimenko I, Khoroshevsky V.** Peaks, Slopes, Canyons and Plateaus: Identifying Technology Trends Throughout the Life Cycle. *International Journal of Innovation and Technology Management*; 2017; 14(02).
- [3] **Ena O, Mikova N, Saritas O, Sokolova A.** A methodology for technology trend monitoring: the case of semantic technologies. *Scientometrics*; 2016; 108(3): 1013-1041.
- [4] **Nassirtoussi AK, Aghabozorgi S, Wah TY, Ngo DC.** Text mining for market prediction: A systematic review. *Expert Systems with Applications*; 2014; 41(16): 7653-7670.
- [5] **Troshin DV.** Foundations of a conceptual model of economic security threat sources at the national level [In Russian]. *Ontology of designing*. 2017; 7(4): 410-422. - DOI: 10.18287/2223-9537-2017-7-4-410-422.
- [6] **Lukashevich NV.** Thesauri in information-retrieval tasks [in Russian]. – M.: MSU, 2011.
- [7] **Shane S, Hubbard B.** Isis displaying a deft command of varied media. *New York Times*; 2014; 31.
- [8] **Weimann G.** *New Terrorism and New Media*. Wilson Center Common Labs; 2014.
- [9] **Kohlmann E, Alkhoury L.** Profiles of foreign fighters in Syria and Iraq. *CTC Sentinel*, September; 2014; 29.
- [10] **Finlayson MA, Halverson JR, Corman SR.** The N2 corpus: A semantically annotated collection of Islamist extremist stories. In *LREC-2014*; 2014: 896-902.
- [11] **Sela S, Kuflik T, Mesch GS.** Changes in the discourse of online hate blogs: The effect of Barack Obama's election in 2008 // *First Monday*; 2012; 17 (11).
- [12] **Kwok I, Wang Y.** Locate the hate: Detecting tweets against blacks. In *AAAI-2013*; 2013.
- [13] **Zeerak W, Hovy D.** Hateful symbols or hateful people? predictive features for hate speech detection on twitter. *Proceedings of NAACL-HLT-2016*; 2016: 88-93.
- [14] **Nobata Ch, Tetreault J, Thomas A, Yashar M, Chang Yi.** Abusive language detection in online user content, in *Proc. of the 25th International Conference on World Wide Web*; 2016: 145–153.
- [15] **Schmidt A, Wiegand M.** A survey on hate speech detection using natural language processing. *Proceedings of SocialNLP*; 2017.
- [16] **Lim S, Muis A, Lu W, Ong C.** MalwareTextDB: A Database for Annotated Malware Articles. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1; 2017.
- [17] **Kirillov I, Beck D, Chase P, Martin R.** Malware Attribute Enumeration and Characterization. *The MITRE Corporation, Tech. Rep.*; 2010.
- [18] **Gorokhov O, Petrovskiy M, Mashechkin I.** Convolutional Neural Networks for Unsupervised Anomaly Detection in Text Data. *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, Cham; 2017: 500-507.
- [19] **O'Brien SP.** Crisis early warning and decision support: Contemporary approaches and thoughts on future research // *International studies review*; 2010; 12(1): 87-104.
- [20] **Wang W, Kennedy R, Lazer D, Ramakrishnan N.** Growing pains for global monitoring of societal events. *Science*; 2016; 353(6307): 1502-1503.

- [21] **Lukashevich NV, Dobrov BV.** Designing linguistic ontologies for information systems in broad domains [in Russian]. *Ontology of designing*. 2015; 5(1): 47-69.
- [22] **Loukachevitch N, Dobrov B.** RuThes linguistic ontology vs. Russian wordnets. *Proceedings of Global WordNet Conference GWC-2014*; 2014.
- [23] **Loukachevitch N, Dobrov B, Chetviorkin I.** RuThes-Lite, a publicly available version of Thesaurus of Russian language RuThes. *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference в Dialogue, Bekasovo, Russia*; 2014: 340-349.
- [24] **Loukachevitch N, Dobrov B.** The Sociopolitical Thesaurus as a resource for automatic document processing in Russian. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*; 2015; 21 (2): 237-262.
- [25] **Dobrov B, Loukachevitch N.** Development of Linguistic Ontology on Natural Sciences and Technology. *Proceedings of Linguistic Resources and Evaluation Conference LREC-2006*. – 2006.

## Сведения об авторах



**Лукашевич Наталья Валентиновна**, 1964 г. рождения. Окончила факультет вычислительной математики и кибернетики МГУ имени М.В. Ломоносова в 1986 г., к.ф.-м.н. (1989), д.т.н. (2016). Ведущий научный сотрудник НИВЦ МГУ имени М.В. Ломоносова. В списке научных трудов более 180 работ в области автоматической обработки текстов, представления знаний.

**Natalia Valentinovna Loukachevitch** (b.1964) graduated from Lomonosov Moscow State University in 1986, PhD (1989), Doctor of Sciences (2016). She is a leading researcher in Research Computing Center of Lomonosov Moscow State University. She is an author of more 180 scientific papers in the field of natural language processing, knowledge representation.

**Добров Борис Викторович**, 1963 г. рождения. Окончил факультет вычислительной математики и кибернетики МГУ имени М.В. Ломоносова в 1985 г., к.ф.-м.н. (1988). Заведующий лабораторией НИВЦ МГУ имени М.В. Ломоносова. В списке научных трудов более 120 работ в области информационного поиска, онтологий

**Boris Viktorovich Dobrov** (b. 1963) graduated from Lomonosov Moscow State University in 1985, PhD (1988). Chief of laboratory in Research Computing Center of Lomonosov Moscow State University. He is an author of more than 120 publications in the field of information retrieval, ontologies.



**Павлов Андрей Михайлович**, 1986 г. рождения. Окончил факультет вычислительной математики и кибернетики МГУ имени М.В. Ломоносова в 2009 г. Программист НИВЦ МГУ имени М.В. Ломоносова. Имеет публикации по тематике кластеризации новостного потока.

**Andrew Mikhailovich Pavlov** (b. 1986) graduated from Lomonosov Moscow State University in 2009. Programmer of Research Computing Center of Lomonosov Moscow State University. He is an author of publications in the field of news clusterization.

**Штернов Сергей Владимирович**, 1980 г. рождения. Окончил Московский авиационный институт в 2003 г. Программист НИВЦ МГУ имени М.В. Ломоносова. Имеет публикации по тематике информационного поиска и автоматической обработки текстов.

**Sergey Vladimirovich Shternov** (b. 1980) graduated from Moscow Aviation Institute in 2003. Programmer of Research Computing Center of Lomonosov Moscow State University. He is an author of publications in the fields of information retrieval, natural language processing.



УДК 004.043

## ОНТОЛОГИЯ ПРОЕКТИРОВАНИЯ, ПРИМЕНЕНИЯ И СОПРОВОЖДЕНИЯ ПОРТАЛОВ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ

М.А. Навроцкий<sup>1,a</sup>, Н.А. Жукова<sup>1,2,b</sup>, Д.И. Муромцев<sup>1,c</sup>

<sup>1</sup> Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, Россия

<sup>a</sup> [m.navrotskiy@gmail.com](mailto:m.navrotskiy@gmail.com), <sup>c</sup> [d.muromtsev@gmail.com](mailto:d.muromtsev@gmail.com)

<sup>2</sup> Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), Санкт-Петербург, Россия

<sup>b</sup> [nazhukova@mail.ru](mailto:nazhukova@mail.ru)

### Аннотация

В статье рассматриваются вопросы разработки онтологической модели для семантических научных порталов на основе технологии открытых данных. Такие порталы представляют новый класс семантических порталов, которые ориентированы на работу с научными и образовательными знаниями. Семантические научные порталы позволяют поддерживать образовательные процессы и процессы получения научных знаний в прикладных предметных областях. Для описания предметной области портала предложены проблемно-ориентированные модели. Процесс разработки онтологической модели включает следующие этапы: определение конечных пользователей; определение сценариев применения онтологии; определение требований к модели; определение возможности повторного использования существующих онтологий; верификация онтологической модели на основе компетентностных вопросов. При разработке использовались модели: VIVO, TEACH, FOAF, VIVO. Разработанная онтология состоит из модулей: модуль описания пользователя портала; модуль описания источников данных; модуль описания образовательного ресурса; модуль описания поисковой выдачи. Внедрение предложенной модели позволяет реализовать подход, при котором знания предметной области извлекаются из открытых источников с учётом интересов пользователей портала, и сохраняются в онтологическую модель для повторного использования и анализа, при этом происходит порождение новых знаний. Также, онтология позволяет реализовать единый интерфейс для работы с открытыми данными различных предметных областей. С использованием предложенной онтологической модели разработан семантический научный портал для поддержки образовательного процесса в Санкт-Петербургском национальном исследовательском университете информационных технологий, механики и оптики.

**Ключевые слова:** онтология, открытые связанные данные, интеграция открытых данных, поиск в открытых данных.

**Цитирование:** Навроцкий, М.А. Онтология проектирования, применения и сопровождения порталов научно-технической информации / М.А. Навроцкий, Н.А. Жукова, Д.И. Муромцев // Онтология проектирования. – 2018. – Т. 8, №1(27). – С.96-109. – DOI: 10.18287/2223-9537-2018-8-1-96-109.

### Введение

С каждым годом возрастает количество публикуемой научной и образовательной информации в сети. Это образовательные курсы, научные публикации, результаты исследований и другие. Такие данные образуют Сеть науки или Научный Веб [1]. Это данные высокого уровня качества и доверия. Существует сообщество европейских университетов LinkedUniversities [2], которые разрабатывают онтологии описания этих данных. Эти университеты также публикуют свои научные и образовательные данные на собственных LOD (Связанные Открытые Данные – LinkedOpenData) порталах. При публикации научной и об-



разовательной информации организации, являющиеся поставщиками данных, могут использовать разные форматы от стандартных HTML-страниц до семантических наборов данных.

Отдельное место здесь занимают научные порталы [3], обеспечивающие систематизацию знаний и информационных ресурсов, а также их интеграцию в единое информационное пространство. Такое направление является достаточно развитым и актуальным, однако недостатком таких порталов является их направленность на некоторую одну предметную область (ПрО) [4, 5]. С другой стороны, такие порталы могут являться поставщиками научных и образовательных данных весьма высокого качества и уровня доверия.

Научный Веб содержит большой объём образовательной и научной информации, которая может использоваться в различных образовательных процессах. Это не только процессы обучения в учебных учреждениях, но и повышение квалификации, переобучение на предприятиях и другие. Использование Научного Веба позволит решить несколько проблем, в частности, обеспечить индивидуализацию образовательных процессов, повысить их информативность, существенно снизить стоимость разработки и поддержки. Предлагается использовать его для поддержки индивидуальных образовательных траекторий.

Основным препятствием к широкому применению Научного Веба является отсутствие связанных формализованных моделей представления открытых научных данных. Для решения этой проблемы необходимо определить проблемно-ориентированные модели, на основе которых могут быть построены онтологические модели.

## 1 Проблемно-ориентированные модели

Анализируя ПрО образовательных процессов, можно выделить основные модели, с помощью которых её можно описать:

- модель образовательного курса;
- модель обучаемого;
- модель источника данных;
- модель научных интересов пользователя портала (обучаемого);
- модель подразделения университета;
- модели требований.

Модель образовательного курса [6] позволяет описывать образовательный курс, для которого планируется использовать портал научно-технической информации.

$$(1) \text{CResource} = \langle \text{ContentType}, \text{Type}, \text{Verbosity}, \text{Role}, \text{Language}, \text{Complexity}, \text{Prerequisites}, \text{Results} \rangle,$$

где:

- *ContentType* – тип ресурса;
- *Type* = {*Practice*, *Theory*} – характеристика теоретической, либо практической направленности данного образовательного ресурса;
- *Verbosity* = {*Hig*, *Medium*, *Low*} – характеристика подробности изложения материала;
- *Language* – язык материала;
- *Complexity* – сложность изложения;
- *Prerequisites* – требуемые для изучения знания;
- *Results* – знания, получаемые в процессе обучения.

Модель обучаемого (2) позволяет представить обучаемого с учётом его навыков, интересов.

$$(2) \text{Student} = \langle \text{MetaProperties}, \text{Knowledge}, \text{Preferences} \rangle,$$

где:

- $MetaProperties = \{MetaProperty_1, MetaProperty_2, \dots, MetaProperty_n\}$  – набор метаданных пользователя, включающих ФИО и прочую информацию;
- $Knowledge$  – модель текущего поля знаний обучаемого;
- $Preferences$  – модель персональных предпочтений пользователя.

Модель источника данных (3), позволяет описать ресурсы, с которых загружаются данные с учётом ПрО:

$$(3) \quad DataSource = \langle Url, Name, License, Langs, Upd, Stb, Crdb, RA, Datasets \rangle,$$

где:

- $Url$  – адрес источника;
- $Name$  – название источника;
- $Type = \{sparql, json, xml\}$  - формат доступа к данным;
- $License = \{MIT, GNUv3, \dots, None\}$  - лицензия публикуемых данных;
- $Langs = \{RUS, ENG, \dots\}$  - поддерживаемые языки представления наборов данных;
- $Upd = \{Never, Rarely, Often\}$  – обновляемость данных в источнике;
- $Stb = \{Low, Normal, Hig\}$  – стабильность источника данных;
- $Crdb = True/False$  - авторитетность публикуемых данных;
- $RA$  - множество ПрО, к которым относятся наборы данных в источнике;
- $Datasets = \{Description, People, Articles, Projects, Courses, Links\}$  - поддерживаемые наборы данных (описание термина, люди, публикации, проекты, образовательные курсы, темы, ссылки).

Модель научных интересов пользователя (4) представляет собой простое множество ключевых слов:

$$(4) \quad Prefences = \{Keyword_1, Keyword_2, \dots, Keyword_m\}.$$

Модель подразделения организации (5) описывает научное или образовательное подразделение организации. Для университетов такими подразделениями являются лаборатории, кафедры, факультеты и т.д.

$$(5) \quad Department = \langle Url, Name, University, Projects, People, Publs, RA \rangle,$$

где:

- $Url$  - адрес подразделения;
- $Name$  - название подразделения;
- $University$  - университет подразделения;
- $Projects$  - множество проектов подразделения;
- $People$  - множество сотрудников подразделения;
- $Publs$  - множество публикаций подразделения;
- $RA$  - множество ПрО, в которых осуществляет деятельность подразделение.

Модель (6) описывает требования, предъявляемые к источникам и наборам данных:

$$(6) \quad Reqs = \langle StdR, StR, BR, UnR \rangle,$$

которую составляют перечисляемые далее компоненты.

$StdR$  - модель обучаемого, описывающая пользователя портала (7):

$$(7) \quad StdR = \langle Skills, Knowledg, Interests \rangle$$

где:

- $Skills$  - массив получаемых навыков;
- $Knowledg$  - массив получаемых знаний;
- $Interests$  - массив интересов пользователя.

$StR$  - модель требований стандартов - представляет собой множество компетенций (8):

$$(8) \quad StR = \{Competence_1, \dots, Competence_N\}$$

BR - модель требований потребностей бизнеса - описывается как множество навыков (9):

$$(9) \quad BR = \{Skill_1, \dots, Skill_N\}.$$

UnR - модель требований обучающего.

## 2 Разработка онтологической модели

При разработке онтологии учитываются принципы, определённые в [7]:

- Не существует одного единственного правильного способа моделирования ПрО, всегда существуют жизнеспособные альтернативы. Наилучшее решение почти всегда зависит от его применения и тех расширений, которые ожидаются.
- Разработка онтологий – это неизбежно итеративный процесс.
- Концепты в онтологии должны быть как можно более близки к объектам (физическим или логическим) и связям в рассматриваемой ПрО. Они наиболее вероятно являются существительными (объекты) и глаголами (связи) в предложениях, описывающих рассматриваемую ПрО.

К наиболее распространённым методологиям разработки онтологий относятся METHONTOLOGY, On-To-Knowledge, DILIGENT и NeOn. В рамках данной работы использована методология NeOn [7], выбор которой обусловлен следующими факторами:

- наиболее современная методология;
- учитывает наличие большого количества имеющихся онтологий и предусматривает повторное использование существующих онтологий;
- предлагает несколько сценариев разработки, которые выбираются в зависимости от входных требований.

С учётом методологии NeOn, разработка онтологической модели включает в себя следующие шаги:

- 1) идентификация цели, области действия и языка описания онтологии;
- 2) идентификация предполагаемых конечных пользователей;
- 3) идентификация предполагаемых сценариев применения;
- 4) идентификация функциональных и нефункциональных требований;
- 5) группирование функциональных требований (компетентностных вопросов);
- 6) валидация требований;
- 7) определение приоритетов требований;
- 8) извлечение терминологии и оценка её частоты;
- 9) обзор существующих онтологических ресурсов (онтологий, тезаурусов, словарей и т.д.) как в формате RDFS или OWL, так любых других для повторного использования в разрабатываемой модели;
- 10) непосредственное структурирование или кодирование модели на языке RDFS или OWL;
- 11) оценка соответствия разработанной модели требованиям, сформулированным на первом шаге.

Главным сценарием использования онтологии является представление доменных данных и знаний в семантическом научном портале. Это представление включает в себя:

- модель для хранения результатов поиска;
- построение логического вывода на онтологии для определения источников данных для портала при его первоначальной настройке;
- предоставление сохранённых результатов поиска для анализа использования портала и уточнения ПрО портала.

Сохранённые результаты поиска представляют собой новые знания: связь между интересами пользователя, его навыками и поисковыми результатами.

Для описания онтологии требуется использовать язык описания RDF-схем - RDF Schema (RDFS) - или язык описания онтологии – Web Ontology Language (OWL).

### 3 Предполагаемые конечные пользователи и сценарии применения онтологии

Для разрабатываемой онтологии выделяются следующие группы пользователей:

- пользователь 1 - специалист, разворачивающий портал в организации (осуществляющий первичную настройку системы);
- пользователь 2 - пользователь системы, вводящий запросы для поиска;
- пользователь 3 - специалист, добавляющий источники данных в онтологию;
- пользователь 4 - разработчик портала, использующий онтологию при разработке портала. Предлагаемую онтологию можно использовать в следующих сценариях:
- получение списка источников данных, которые будут использоваться для организации поиска (выполняется пользователем 1);
- сохранение результатов поиска (выполняется пользователем 2);
- добавление новых источников данных в онтологию (выполняется пользователем 3);
- обработка данных, описанных онтологической моделью, с целью реализации модулей портала (выполняется пользователем 4).

### 4 Нефункциональные и функциональные требования

Разрабатываемая онтология должна отвечать следующим функциональным требованиям (ФТ):

- ФТ1: обеспечить получение списка источников данных по определённым требованиям;
- ФТ2: обеспечить выдачу списка пользователей портала;
- ФТ3: обеспечить выдачу поисковых интересов пользователей за выбранный промежуток времени. Данное требование включает в себя получение:
  - определений (на русском и английском языках) ключевого слова;
  - списка специалистов в исследуемой ПрО;
  - списка научных публикаций;
  - списка исследовательских проектов.

Разрабатываемая онтология должна соответствовать следующим нефункциональным требованиям:

- наименование концептов онтологии должно производиться на английском языке; наименование локальных имен концептов должно соответствовать UpperCamelCase-стилю для классов и индивидов, и lowerCamelCase стилю<sup>1</sup> для связей;
- онтология не должна напрямую импортировать ни одну из существующих онтологий, чтобы не вносить прямые зависимости от внешних онтологий;
- онтология должна использовать только конструкции подмножества языка RDF Schema.

На основе описанных функциональных требований можно выделить следующее множество основных терминов: источник данных, пользователь, лицензия распространения, тип

<sup>1</sup>UpperCamelCase (с англ. — «ВерблюжийРегистр» или «ГорбатыйРегистр») — стиль написания составных слов, при котором несколько слов пишется слитно без пробелов, при этом каждое слово внутри фразы пишется с заглавной буквы. В стиле lowerCamelCase внутри фразы с заглавной пишутся все слова, кроме первого. *Прим. ред.*

доступа к данным, набор данных, язык публикации данных, ПрО, термин, перевод термина, значение термина.

## 5 Определение возможности повторного использования существующих онтологий

Одной из рекомендаций по разработке новых онтологий является повторное использование концептов из существующих онтологий. На данный момент разработано много онтологий, с помощью которых можно описать проблемно ориентированную модель. В связи с этим необходимо провести анализ и определить, какие из существующих онтологий могут быть использованы [8]. Анализ онтологий проводится по следующим параметрам:

- предназначение (для чего данная онтология была разработана?);
- пространство имен онтологии;
- концепты для повторного использования (какие из концептов, определённых в данной онтологии целесообразно использовать повторно?).

*Онтология VIVO* [9]. Онтологическая модель университета как организации включает в себя персонал, корпуса и документооборот. Модель разработана организацией AKSWGGroup. Пространства имен: <http://vivoplus.aksw.org/ontology#> и <http://vivoweb.org/ontology/core#>. Предлагается использовать данную модель для описания университета и ПрО наборов данных и знаний. В онтологии используется префиксы «vivoplus#» и «vivoweb#». Онтология ориентирована на группы требований ФТ2 (описание интересов пользователя и место его работы или обучения), ФТ3 (описание ПрО наборов данных в поиске).

*Онтология TEACH* [10], ориентирована на обучение и охватывает организационные аспекты (аудитория, корпус, преподаватель, студент). Пространство имен: <http://linkedscience.org/teach/ns#>. Предлагается использовать классы и свойства, которые содержатся в модели для описания моделей пользователя системы (учащихся) и образовательных ресурсов. В онтологии используется префикс «teach#». Онтология ориентирована на группы требований ФТ2 (описание пользователя).

*Онтология FOAF* [11]. Распространённая модель, которая может использоваться во многих ПрО. Онтология для описания домашних страниц, людей и социальных сетей. Пространство имен: <http://xmlns.com/foaf/0.1/>. Предлагается использовать для описания базовой модели пользователя системы, для описания научно-исследовательских проектов и образовательных учреждений. В онтологии используется префикс «foaf#». Онтология ориентирована на группы требований ФТ2 (описание метаданных пользователя – места работы или учебы), ФТ3 (описание проектов в поиске).

*Онтология BIBO* [12], используется для описания библиографии. Содержит базовые концепты и свойства, предназначенные для описания цитат и библиографических ссылок. Пространство имен: <http://purl.org/ontology/bibo/>. В проекте предлагается использовать для описания научных публикаций: статей, книг и других. Онтология ориентирована на группы требований ФТ3 (описание публикаций в поиске). Онтология состоит из модулей: модуль описания пользователя системы; модуль описания источников данных; модель словаря ПрО; модуль описания поисковой выдачи. Основные концепты модулей изображаются в виде диаграмм, нарисованных в соответствии со спецификацией VOWL.

*Модуль описания пользователя системы.* На рисунке 1 представлен фрагмент структуры модуля. Используются классы существующих онтологий:

- vivoplus:ResearchArea - область исследований подразделения;
- vivoplus:University - университет, в котором обучается пользователь;
- foaf:Department - подразделение, в котором работает (учится) пользователь;

- foaf:Person - пользователь системы;
- foaf:Project - научно-исследовательские проекты;
- teach:Student - студент;
- bibo:Publication – научные публикации;
- Preference - персональные предпочтения пользователя;
- Knowledge - текущее поле знаний пользователя.

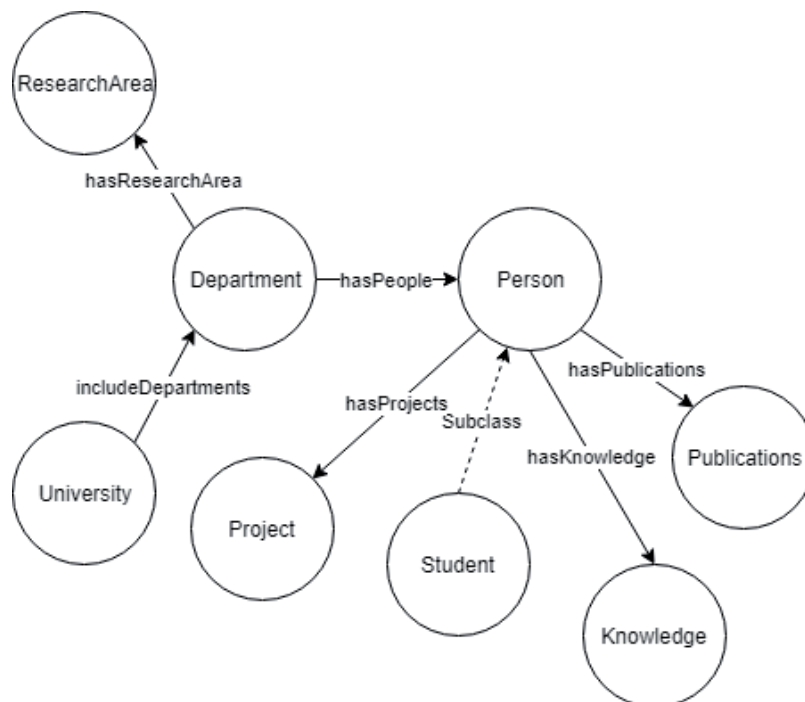


Рисунок 1 - Структура модуля описания пользователя (фрагмент)

*Модуль описания источников данных.* Структура модуля представлена на рисунке 2. Используются следующие классы:

- vivoplus:ResearchArea – ПрО (область исследований);
- DataSource - источник данных;
- Language - язык;
- License - лицензии распространения данных;
- Updatability - обновляемость данных;
- SourceType - тип формата доступа к данным;
- Stability - стабильность источника данных;
- Dataset – наборы данных, которые поддерживает источник данных;
- Credibility - авторитетность публикуемых данных.

*Модуль описания поисковой выдачи* представлен на рисунке 3. Используются следующие классы:

- bibo:Article - научная статья;
- foaf:Person - личность (ученого), который работал в ПрО запроса пользователя;
- foaf:Project - научно-исследовательский проект;
- SearchData - данные поиска (запрос пользователя);
- Keyword - термин (поиск по термину);
- Wikidata - данные с ресурса wikidata;

- Link - ссылки по ПрО (поиску пользователя);
- Subject - связанные темы для ПрО (поиску пользователя);

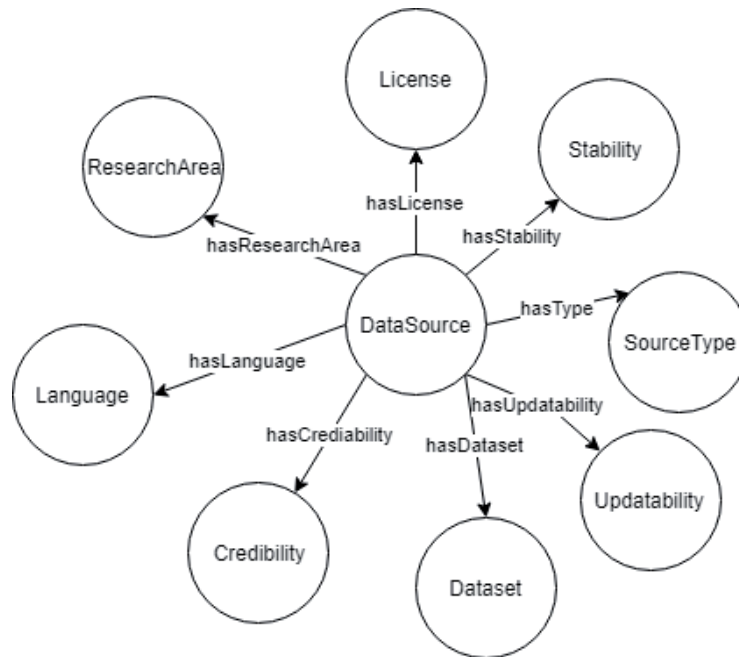


Рисунок 2 - Структура модуля описания источников данных

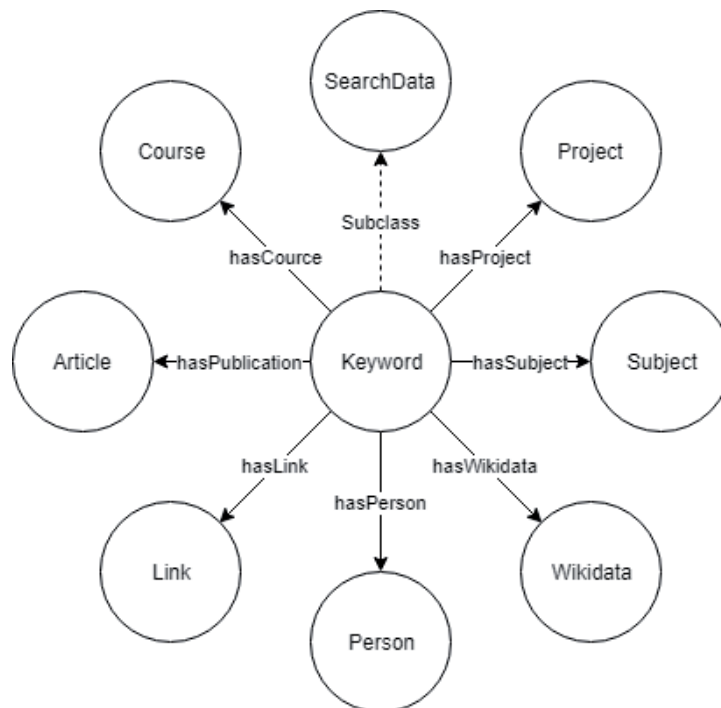


Рисунок 3 - Структура модуля описания поисковой выдачи

- teach:Resource - образовательный ресурс, курса (см. рисунок 4);
- Type - характеристика образовательного ресурса (курса);
- Prerequisites - требуемые для изучения знания;
- Result - знания, получаемые в процессе обучения;

- ContentType - характеристика данного ресурса (курса);
- Complexity - сложность изложения;
- Verbosity – характеристика подробности изложения материала.

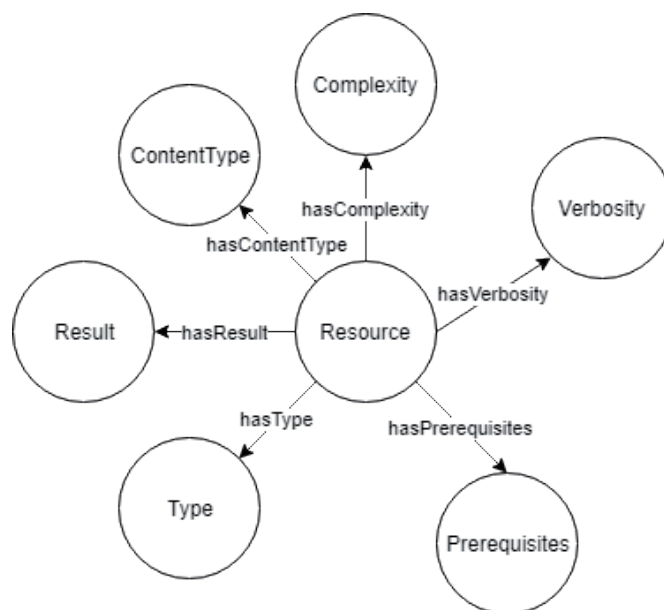


Рисунок 4 – Структура модуля описания учебного ресурса

## 6 Верификация онтологической модели

Метод верификации онтологической модели заключается в представлении компетентностных вопросов в запросы на языке SPARQL, которые впоследствии выполняются над данными аннотированными проверяемой онтологией.

Оценка соответствия онтологической модели каждому из требований включает в себя следующие шаги:

- 1) подготовка набора данных, содержащего описание данных;
- 2) запись компетентностного вопроса в виде SPARQL-запроса;
- 3) выполнение данного запроса и сравнение его результатов с ожидаемыми данными.

Первая группа функциональных требований представлена в виде вопроса (ФТ1) «Получить источники данных на русском языке, доступ к которым определяется свободной лицензией». Компетентностный вопрос в виде SPARQL-запроса имеет вид:

```
SELECT ?source
WHERE {
    ?source rdf:type lodifmo:DataSource .
    ?source lodifmo:hasLicense:FREE .
    ?source lodifmo:hasLanguage lodifmo:RUS
}
```

Ожидается, что из всех добавленных источников данных вернутся: DBpedia, Wikidata, LOD-IFMO. Только эти источники выдают наборы данных на русском языке и по свободной лицензии.

Результат выполнения SPARQL-запроса приведён в таблице 1. Сравнивая фактический и ожидаемый результаты запроса, приходим к заключению, что требование выполнено.



Таблица 1 – Результат выполнение SPARQL-запроса

№	source
1	DBpedia
2	LOD-IFMO
3	Wikidata

Вторая группа функциональных требований (ФТ2) представлена в виде вопроса «Получить список всех пользователей портала». Ожидается, что будут получены все пользователи портала. Компетентный вопрос в виде SPARQL-запроса имеет вид:

```
SELECT ?user
WHERE {
    ?user rdf:type:Student
}
```

Результат выполнения SPARQL-запроса приведён в таблице 2. Сравнивая полученные результаты, приходим к заключению, что требование выполнено.

Таблица 2 – Результат выполнения SPARQL-запроса (Фрагмент)

№	user
1	Navrotskiy_MA
2	Navrotskiy_RA
3	Mouromtsev_DI

Третья группа функциональных требований (ФТ3) представлена в виде вопроса «Получение поисковых интересов (в виде ключевых слов) пользователя за выбранный промежуток». Ожидается, что будут получены поисковые интересы: SemanticWeb, Ontology, OWL, LinkedOpenData. Компетентный вопрос в виде SPARQL-запроса имеет вид:

```
SELECTDISTINCT ?search
WHERE {
    ?search rdf:type:lodifmo:Keyword .
    ?search lodifmo:createdAt ?date .
    ?user rdf:label ?name .
    ?search lodifmo:hasPerson ?user .
    FILTER (?date > "2018-01-09T00:00:00+03:00"^^xsd:dateTime) .
    FILTER (?name = "Navrotskiy"^^xsd:string)
}
```

Результат выполнения SPARQL-запроса приведён в таблице 3. Сравнивая полученные результаты, приходим к заключению, что требование выполнено.

Таблица 3 – Результат выполнения SPARQL-запроса

№	source
1	SemanticWeb
2	Ontology
3	OWL20
4	LinkedOpenData

## 7 Применение онтологической модели

Разработанная онтологическая модель применялась при создании научного портала для поддержки образовательного процесса в университете ИТМО на кафедре информатики и прикладной математики.

Ниже приведён пример *входных данных*.

Пользователь - студент университета 4 курса, начинающий работу над своей диссертацией в области семантического веба. Описать его можно так:

$User = \langle MetaProperties, Knowledge, Preferences \rangle$ ,

где:

- *MetaProperties* - описание его метаданных: ФИО (Навроцкий Михаил Александрович), учебная группа (3400), электронная почта (m.navrotskiy@gmail.com), направление обучения (Программная инженерия).
- *Knowledge* = {*Knowledge*<sub>1</sub>, *Knowledge*<sub>2</sub>, ..., *ledge*<sub>n</sub>} - множество знаний пользователя: алгоритмы и структуры данных; основы программирования; базы данных и т.д.
- *Preferences* = {*Pference*<sub>1</sub>, *Preference*<sub>2</sub>, *Preference*<sub>3</sub>} – множество научных интересов пользователя:
  - *Preference*<sub>1</sub> = «Web»;
  - *Preference*<sub>2</sub> = «LOD»;
  - *Preference*<sub>3</sub> = «Ontology».

В данном случае запросом пользователя будет «SemanticWeb». Пользователь получает наборы данных по своему запросу (ключевому слову):

- английское определение;
- русское определение;
- список специалистов ПрО;
- список научных публикаций;
- список исследовательских проектов;
- список ссылок;
- список ближайших тем и определений.

Примером *выходных данных* для разработанного портала является HTML-страница, фрагмент которой представлен на рисунке 5. Страница отображает следующие данные:

- определение термина на двух языках: русский (RUS) и английский (ENG);
- список публикаций по ПрО поиска (Publications);
- список научно-исследовательских проектов по ПрО поиска;
- список известных специалистов предметной области (People);
- список ведущих университетов ПрО;
- связанные ПрО и близкие термины.

### Заключение

Проблемно-ориентированные модели, разработанные на основе существующих моделей ПрО, позволяют описывать ее с учётом требований к образовательным процессам и использовать несколько источников данных. Разработанные онтологические модели представляют общую онтологию моделирования открытых научно-технических данных для различных ПрО. Её отличие от существующих заключается в том, что вместо моделирования всех возможных концептов онтология ориентирована на расширение с использованием уже существующих онтологий с помощью подмножества языка RDFS и единого интерфейса для работы с открытыми данными различных ПрО.

## Semantic Web

Data
Data graph
Structure graph

ENG
RUS

The Semantic Web is an extension of the Web through standards by the World Wide Web Consortium (W3C). The standards promote common data formats and exchange protocols on the Web, most fundamentally the Resource Description Framework (RDF). According to the W3C, "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries". The term was coined by Tim Berners-Lee for a web of data that can be processed by machines. While its critics have questioned its feasibility, proponents argue that applications in industry, biology and human sciences research have already proven the validity of the original concept. The 2001 Scientific American article by Berners-Lee, Hendler, and Lassila described an expected evolution of the existing Web to a Semantic Web. In 2006, Berners-Lee and colleagues stated that: "This simple idea...remains largely unrealized". In 2013, more than four million Web domains contained Semantic Web markup.

dbpedia
Wikipedia
W3C

**Publications**

- Marco-Ruiz, Luis; Pedrinaci, Carlos ; Maldonado, J.A.; Panziera, Luca; Chen, Rong and Bellika, J. Gustav (2016). Publication, discovery and interoperability of Clinical Decision Support Systems: A Linked Data approach. *Journal of Biomedical Informatics*, 62 pp. 243–264. [↗](#)
- d'Aquin, Mathieu and Motta, Enrico (2016). The Epistemology of Intelligent Semantic Web Systems. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 6 (1). Morgan & Claypool. [↗](#)
- de Ribaupierre, Héliène ; Osborne, Francesco and Motta, Enrico (2016). Combining NLP and Semantics for Mining Software Technologies from Research Publications. In: *WWW '16 Companion: Proceedings of the 25th International Conference Companion on World Wide Web, International World Wide Web Conferences Steering Committee*, pp. 23–24. [↗](#)
- Distinto, Isabella; d'Aquin, Mathieu and Motta, Enrico (2016). LOTED2: An Ontology of European Public Procurement Notices. *Semantic Web Journal*, 7(3) [↗](#)
- Osborne, Francesco ; Salatino, Angelo ; Birukou, Aliaksandr and Motta, Enrico (2016). Automatic Classification

**People**

- Tim Finin (University of Illinois at Urbana-Champaign)
- Deborah McGuinness
- Nigel Shadbolt (University of Edinburgh)
- Joseph G. Davis (Indian Institute of Management Ahmedabad)
- Carole Goble (University of Manchester)
- Tim Finin (Massachusetts Institute of Technology)
- Farshad Fotouhi (Michigan State University)
- Ian Horrocks (University of Manchester)
- Marc Twagirimukiza (University of Rwanda)
- Frank van Harmelen (University of Edinburgh)
- James Hendler (Brown University)
- Peter Fox (professor) (Monash University)
- Wendy Hall (City University London)
- David Karger (Harvard University)
- David Karger (Stanford University)
- Wendy Hall (University of Southampton)
- Rudi Studer (University of Stuttgart)
- Nigel Shadbolt (Newcastle University)
- Joseph G. Davis (University of Calicut)
- David De Roure
- Marc Twagirimukiza (Ghent University)
- Joseph G. Davis (University of Pittsburgh)

dbpedia

Рисунок 5 – Пример выходных данных (фрагмент)

### Список источников

- [1] *Муромцев, Д.И.* Исследование актуальных способов публикации открытых научных данных в сети / Д.И. Муромцев, Й. Леманн, И.А. Семерханов, М.А. Навроцкий, И.С. Ермилов // *Научно-технический вестник информационных технологий, механики и оптики*. - 2015. - Т. 15, №6. - С. 1081–1087. - DOI: 10.17586/2226-1494-2015-15-6-1081-1087.
- [2] *Halac, T.G.* et al. Publishing and linking university data considering the dynamism of data sources // *Proceedings of the 9<sup>th</sup> International Conference on Semantic Systems*. – ACM, 2013. – P.140-145.
- [3] *Загоруйко, Ю.А.* Подход к построению порталов научных знаний / Ю.А. Загоруйко, О.И. Боровикова // *Автоматрия*. – 2008. - №1. – С. 100-110.
- [4] *Андреева, О.А.* Об организации порталов знаний по археологии на основе онтологий / О.А. Андреева, О.И. Боровикова, С.В. Булгаков, Ю.А. Загоруйко, Е.А. Сидорова, Ю.П. Холушкин, Б.Г. Циркин // *Вестник НГУ*. – 2009. - №5.
- [5] *Загоруйко, Ю.А.* Подход к построению предметной онтологии для портала знаний по компьютерной лингвистике / Ю.А. Загоруйко, О.И. Боровикова, И.С. Кононенко, Е.А. Сидорова // *Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог–2006»*. – С. 148-151.
- [6] *Ужва, А.Ю.* Автоматизированная разработка онтологической модели предметной области для поиска образовательных ресурсов с использованием анализа текстов рабочих программ / А.Ю. Ужва // *Современные проблемы науки и образования*. – 2013. – №. 1. – <http://www.science-education.ru/ru/article/view?id=8324>.
- [7] *Suárez-Figueroa, M.* The NeOn methodology for ontology engineering / M. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López // *Ontology engineering in a networked world*. – Springer Berlin Heidelberg, 2012. – P. 9-34.
- [8] *Janowicz K.* et al. Five stars of linked data vocabulary use. Editorial // *Semantic Web 0(2014)*. IOS Press.4 p.
- [9] *Ding, Y.* Extending the VIVO ontology to iSchools: Enabling networking of information scientists / Y. Ding, E. Yan, A. Ghazinejad, H. Jia // *iConference 2013 Proceedings* – P. 905-908. - DOI: 10.9776/13460. - <https://www.ideals.illinois.edu/bitstream/handle/2142/42085/460.pdf?sequence=2>.
- [10] *Chuprina, S.* Ontology based data access methods to teach students to transform traditional information systems and simplify decision making process / S. Chuprina, I. Postanogov, O. Nasraoui // *Proceedings Computer Science*. – 2016. – V. 80. – P. 1801-1811.

- [11] **Golbeck, J.** Linking Social Networks on the Web with FOAF: A Semantic Web Case Study / J. Golbeck, M. Rothstein // AAAI. – 2008. – V. 8. – P.1138-1143.
- [12] **Ruiz-Iniesta, A.** A review of ontologies for describing scholarly and scientific documents / A. Ruiz-Iniesta, O. Corcho // Proceedings of the 4th Workshop on Semantic Publishing, co-located with the 11th Extended Semantic Web Conference (ESWC 2014). Anissaras, Greece, May 25th, 2014. – 12 p. - <http://ceur-ws.org/Vol-1155/paper-07.pdf>.

## ONTOLOGY FOR DESIGN, APPLICATION AND SUPPORT OF SCIENTIFIC AND ENGINEERING DATA PORTALS

**M.A. Navrotskiy<sup>1,a</sup>, N.A. Zhukova<sup>1,2,b</sup>, D.I. Mouromtsev<sup>1,c</sup>**

<sup>1</sup>ITMO University, Saint Petersburg, Russia

<sup>a</sup> [m.navrotskiy@gmail.com](mailto:m.navrotskiy@gmail.com), <sup>c</sup> [d.muromtsev@gmail.com](mailto:d.muromtsev@gmail.com)

<sup>2</sup>Saint Petersburg Electro Technical University "LETI", Saint Petersburg, Russia

<sup>b</sup> [nazhukova@mail.ru](mailto:nazhukova@mail.ru)

### Abstract

In the article the description of the development of an ontological model for semantic research portals is given. They are a new class of semantic portals focused on working with scientific and educational knowledge. Such portals allow supporting educational processes and processes of obtaining scientific knowledge in applied domains. The domain has been described by using problem-oriented models. Developing ontology includes: defining the end-user and application scenarios of ontology; the definition of requirements to the model; determination of the possibility of reusing existing ontologies; verification of the ontological model on the basis of competence issues. The existing models: VIVO, TEACH, FOAF, BIBO were utilized in the development of the described ontology model. The developed ontology model consists of several modules: module description of the user; a module for description of data sources; a module of the training resource; module for description of search results. A novelty of the approach lies in the way that domain knowledge is derived from open sources, taking into account the interests of the users of the portal and stored in the ontological model. The possibility of new portals is shown by the example of the portal developed for the ITMO University.

**Key words:** *ontology, linked open data, open data integration, open data search, LOD.*

**Citation:** *Navrotskiy MA, Zhukova NA, Mouromtsev DI. Ontology for design, application and support of scientific and engineering data portals [In Russian]. Ontology of designing. 2018; 8(1): 96-109. - DOI: 10.18287/2223-9537-2018-8-1-96-109.*

### References

- [1] **Mouromtsev DI, Lehmann J, Semerkhanov IA, Navrotskiy MA, Ermilov IS.** Study of current approaches for Web publishing of open scientific data. [In Russian]. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2015, 15(6): 1081–1087. - DOI: 10.17586/2226-1494-2015-15-6-1081-1087.
- [2] **Halac TG.** et al. Publishing and linking university data considering the dynamism of data sources // Proceedings of the 9th International Conference on Semantic Systems. – ACM, 2013. – P.140-145.
- [3] **Zagorulko YuA, Borovikova OI.** Approach to building portals of scientific knowledge [In Russian]. Avtometriya. – 2008. - No1. – P.100-110.
- [4] **Andreeva OA, Borovikova IO, Bulgakov SV, Zagorulko YuA, Sidorova EA, Chelushkin P, Tsirkin BG.** On the organization of the portals of knowledge on the archaeology of ontology-based [In Russian] // Vestnik Novosibirsk State University. – 2009. - No5.
- [5] **Zagorulko YuA, Borovikova OI, Kononenko IS, Sidorova EA.** Approach to construction of subject domain ontologies for the knowledge portal on computational linguistics // Computational Linguistics and Intellectual Technologies. – 2006. – P.148-151.
- [6] **Uzhva AY.** Automatic development of ontology model for case-based reasoning in search of educational resources using analysis of educational programs [In Russian]. Modern problems of science and education. – 2013. – No1. – <http://www.science-education.ru/ru/article/view?id=8324>.

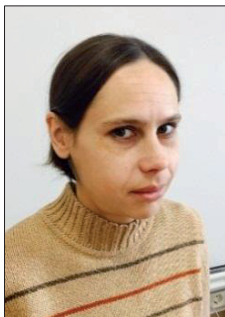
- [7] *Suárez-Figueroa MC, Gómez-Pérez A, Fernández-López M.* The NeOn methodology for ontology engineering //Ontology engineering in a networked world. – Springer Berlin Heidelberg, 2012. – P.9-34.
- [8] *Janowicz K.* et al. Five stars of linked data vocabulary use // Semantic Web. – 2014; 5(3): 173-176.
- [9] *Ding Y, Yan E, Ghazinejad A, Jia H.* Extending the VIVO ontology to iSchools: Enabling networking of information scientists // iConference 2013 Proceedings – P. 905-908. - DOI: 10.9776/13460. - <https://www.ideals.illinois.edu/bitstream/handle/2142/42085/460.pdf?sequence=2>.
- [10] *Chuprina S, Postanogov I, Nasraoui O.* Ontology based data access methods to teach students to transform traditional information systems and simplify decision making process //Proceedings Computer Science. – 2016; 80: 1801-1811.
- [11] *Golbeck, J.* Linking Social Networks on the Web with FOAF: A Semantic Web Case Study / J. Golbeck, M. Rothstein // AAAI. – 2008. – V. 8. – P.1138-1143.
- [12] *Ruiz-Iniesta A, Corcho O.* A review of ontologies for describing scholarly and scientific documents // Proceedings of the 4th Workshop on Semantic Publishing, co-located with the 11th Extended Semantic Web Conference (ESWC 2014). Anissaras, Greece, May 25th, 2014. – 12 p. - <http://ceur-ws.org/Vol-1155/paper-07.pdf>.

## Сведения об авторах



*Навроцкий Михаил Александрович*, 1991 г. рождения. Окончил Волгоградский технический университет в 2014 г. Ассистент кафедры информатики и прикладной математики Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики (ИТМО). В списке научных трудов более 5 работ в области технологий семантического Web.

*Navrotsky Mikhail Alexandrovich* (b. 1991) graduated from the Volgograd State Technical University in 2014. He is an assistant at Informatics and Applied Mathematics Department, ITMO University. He is co-author of more than 5 publications in the field of semantic web technologies.



*Жукова Наталья Александровна*, 1983 г. рождения. Окончила Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина) (СПбГЭТУ ЛЭТИ) в 2006 году, к.т.н. (2008). Доцент кафедры математического обеспечения ЭВМ СПбГЭТУ ЛЭТИ и кафедры информатики и прикладной математики Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики (ИТМО). В списке научных трудов более 20 работ в области обработки и анализа данных.

*Zhukova Natalya Alexandrovna* (b. 1983) graduated from the Saint Petersburg Electrotechnical University "LETI" (ETU) in 2006. Ph.D (2008), Associate Professor, ETU and ITMO University. She is co-author of more than 20 publications in the field of data processing and analyses.



*Муромцев Дмитрий Ильич*, 1976 г. рождения. Окончил факультет технической кибернетики Санкт-Петербургского государственного политехнического университета (1999). Получил степень кандидата технических наук в области компьютерных наук в Санкт-Петербургском национальном исследовательском университете информационных технологий, механики и оптики (ИТМО) в 2003 г. Заведующий кафедрой информатики и прикладной математики Университета ИТМО. Является автором и соавтором более 100 научных и учебно-методических публикаций и 4 книг.

*Mouromtsev Dmitriy Ilyich* (b. 1976) received his BSc (1997) and MSc (1999) degrees in Computer System Design from Saint-Petersburg Polytechnical University (Russia). He received a PhD degree in Computer Science from ITMO University in 2003. Currently he is a Chair at

the Department of Computer Science and Applied Mathematics of ITMO University. He is an author and co-author of more than 100 research and educational publications and 4 books.

УДК 004.032.26

## ОБЪЕКТНО-ОРИЕНТИРОВАННАЯ ДЕКОМПОЗИЦИЯ ПРОГРАММНОЙ ЛОГИКИ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Р.В. Гирин<sup>1</sup>, С.П. Орлов<sup>2</sup>

<sup>1</sup> ООО «Интеллект Софт», Самара, Россия  
romangirin@gmail.com

<sup>2</sup> Самарский государственный технический университет, Самара, Россия  
orlovsp1946@gmail.com

### Аннотация

В статье описана декомпозиция логики программной реализации искусственных нейронных сетей с целью её последующей реализации в виде набора слабосвязанных доменных классов. Рассмотрены структурная и функциональная декомпозиция логики. Методика декомпозиции программной логики иллюстрируется UML-диаграммами. Приведено краткое описание практической реализации программной логики искусственной нейронной сети на языке C#, выполненной на основе объектно-ориентированного подхода. Подобная реализация использована при проектировании нейронных сетей различной конфигурации с целью их последующего обучения, эксплуатации, а также проведения экспериментов. Предлагаемое разделение программной логики между классами и обеспечение слабой связанности между ними упрощает процесс поиска ошибок в программном коде, делает код более управляемым, повышает продуктивность разработчика. Приведённые в статье примеры демонстрируют, как рациональная декомпозиция программной логики нейронной сети на семантически независимые между собой блоки в сочетании с применением концепции «Внедрение зависимостей», способствует большей структурированности кода. Новым результатом является применение объектно-ориентированной декомпозиции к программной логике нейронной сети, что позволило значительно упростить процесс проектирования программного кода. Приведён пример тестирования декомпозированной логики нейронной сети. Предлагаемый подход к декомпозиции программной логики искусственных нейронных сетей можно применить к широкому спектру различных нейронных сетей, таких, как конволюционные и деконволюционные сети, сети, содержащие одновременно конволюционные и полносвязанные слои нейронов. При этом единообразие реализации сетей различных видов обуславливает упрощение их понимания, снижает трудовые затраты на сопровождение и развитие реализующего их кода.

**Ключевые слова:** искусственная нейронная сеть, системный анализ, декомпозиция, объектно-ориентированный анализ.

**Цитирование:** Гирин, Р.В. Объектно-ориентированная декомпозиция программной логики искусственных нейронных сетей / Р.В. Гирин, С.П. Орлов // Онтология проектирования. – 2018. – Т. 8, №1(27). – С.110-123. – DOI:10.18287/2223-9537-2018-8-1-110-123.

### Введение

Использование искусственных нейронных сетей (ИНС) стало прорывом во многих областях машинного обучения и искусственного интеллекта [1-3]. В этой связи актуальна разработка различных подходов к проектированию ИНС. Во многих проектах необходимо наличие программной библиотеки, предоставляющей ИНС таким образом, чтобы можно было гибко варьировать и расширять её функциональность. Это, в свою очередь, достижимо только в случае рационального проектирования программной логики ИНС. В этом плане задачи реализации ИНС не отличаются от какого-либо другого проекта, связанного с разработкой программного обеспечения [4].

Известно, что при выполнении проектирования программного обеспечения значительное упрощение достигается путем декомпозиции всей совокупности реализуемой функциональности на её составляющие. В некоторых источниках подобную выделенную функциональность, подлежащую реализации в виде отдельного слабо связанного компонента, условно называют логикой [4], например, логика доступа к данным (англ. data access logic) и т.п. В данной статье под программной логикой ИНС понимается вся совокупность функциональностей, выполняющих обработку данных на слоях нейронной сети. Следует отметить, что программная логика – это не логическое программирование, парадигма которого основана на автоматическом доказательстве теорем.

В настоящее время разработано большое число программных реализаций нейронных сетей. При этом во многих реализациях получены эффективные решения, которые целесообразно использовать в дальнейших разработках. В тоже время отсутствует методологический подход, который бы опирался на эффективные приёмы объектно-ориентированного проектирования и использовал накопленные знания в области ИНС.

В предлагаемой статье развивается объектно-ориентированный подход к анализу программной логики ИНС [5-7]. Совокупность реализаций ИНС, доступных в исходных кодах (open source) на различных языках, можно условно разделить на две большие группы. К одной группе относятся реализации специализированной платформы для алгоритмов машинного обучения, в их числе и ИНС. В другую группу входят библиотеки, которые не используют какие-либо сторонние платформы, и именно они рассматриваются в данной публикации. Выполненная авторами имплементация (практическая реализация) ИНС также не использует сторонние платформы для алгоритмов машинного обучения, и выполнение кода осуществляется непосредственно на CLR (*Common Language Runtime* — общезыковая исполняющая среда) [8].

В силу актуальности тем, связанных с использованием ИНС, в настоящее время существует большое число примеров их исходных кодов. Многие из программных библиотек, доступных в исходных кодах, например [4, 9], выделяют класс нейронной сети, на который возложено хранение перечня слоев. Этот класс инкапсулирует логику взаимодействия данных слоев для обеспечения прямого и обратного прохождения сигнала. Некоторые варианты реализации добавляют в логику нейронной сети функциональность, обеспечивающую сбор и обработку основной статистики, накапливаемой при обучении нейронной сети. Интересно отметить, что с определённой точки зрения подобное совмещение может противоречить принципу единственности ответственности.

Реализация ИНС в программной библиотеке *Caffe (Convolution Architecture For Feature Extraction* - свёрточная архитектура для распознавания признаков) [10], доступная в исходном коде на языке C++, содержит богатую палитру различных имплементаций слоёв нейронов ИНС. Эти слои удобно взаимно заменять, получая новую функциональность ИНС, обеспеченную набором слоёв с заданными характеристиками, такими как: функция активации, принцип, по которому сигнал корректируется весом синоптической связи, и др.

Известно, что нейронные сети широко варьируются по своей структуре, принципу передачи сигнала между слоями, применяемым функциям активации и другим характеристикам [11]. Программная библиотека, описанная в [12], обеспечивает реализации ИНС на языке Java при едином технологическом подходе к сетям разной структуры.

## 1 Структурная и функциональная декомпозиция программной логики ИНС

Рассматривая всю совокупность логики ИНС, можно выделить следующие основные структурные компоненты.

- *Функция активации.* Она может быть представлена в виде сущности, инкапсулирующей логику, связанную с функцией активации. Эта сущность содержит данные о параметрах функции активации и логику вычисления значений функции и её производной.
- *Функция ошибки.* Эта сущность применяется при обучении ИНС. Для тренировки ИНС применяется подход, который в области машинного обучения именуется «Обучение с учителем». В ходе этого процесса вместе с входным сигналом доступны сведения об ожидаемом выходном сигнале, который можно сравнить с фактически полученным сигналом. Оценку величины расхождения эталонного и фактического результата оценивают с использованием функции ошибки для определения величины, на которую необходимо скорректировать веса синаптических связей.
- *Слой нейронов ИНС.* Он может быть представлен в виде сущности, инкапсулирующей логику работы совокупности искусственных нейронов, находящихся на одном слое ИНС. Он обеспечивает хранение значений весов синаптических связей и содержит логику вычисления поля индукцирования (передаваемого функции активации) на основании входного сигнала. Поскольку данная логика использует функцию активации, каждый слой содержит хотя бы один объект функции активации.
- *Нейронная сеть.* Сеть может быть представлена в виде совокупности одного или нескольких слоев искусственных нейронов вместе с логикой их взаимодействия (передача сигнала в процессе его прямого и обратного прохождения, хранение значений предыдущих сигналов при необходимости и т.п.). Нейронная сеть для своего функционирования должна содержать хотя бы один слой нейронов.

В то же время, рассматривая функциональную организацию логики ИНС, можно выделить следующие уровни.

- *Логика прямого прохождения сигнала.* Она представляет собой основной режим работы нейронной сети, когда сигнал, поданный на вход ИНС, передается на вход функции активации по синаптическим связям с корректировкой на их веса. При этом функция активации, в свою очередь, формирует выходной сигнал отдельно взятого слоя нейронов или ИНС в целом. В конволюционных сетях [13] все нейроны, расположенные на одном слое, используют одинаковые синаптические связи. Это необходимо учитывать и в логике прямого прохождения сигнала, и в логике корректировки весов при обучении.
- *Логика обратного прохождения сигнала ошибки.* Эта функциональность применяется для обучения ИНС [13]. При обратном прохождении сигнала выходные значения ИНС сравниваются с эталонным сигналом (который доступен при обучении с учителем) и их расхождение оценивается с использованием функции ошибки. Полученная ошибка на выходном слое передается на предшествующий слой и далее до входного слоя.
- *Логика корректировки весов и сдвигов нейронной сети.* Подобная корректировка необходима для обучения нейронной сети. Удобно, когда эта логика гибкая и взаимозаменяемая. При этом частная производная функции ошибки используется для определения непосредственно величины, на которую корректируется вес связи или сдвиг. При необходимости имплементация логики ИНС должна выполнять замену используемой функции ошибки.
- *Вспомогательная логика.* К данной функциональной части можно отнести логику, выполняющую матричные операции, к которым сводится большинство численных преобразований, выполняемых внутри нейронной сети. Например, при использовании концепции глубокого обучения [14] необходимо осуществить «предобучение» каждого отдельного слоя заданной нейронной сети на базе «Обучение без учителя», с последующим обучением всей ИНС на основе «Обучение с учителем». На стадии предобучения требуется логика, выполняющая анализ структуры заданной нейронной сети, генерацию автоэнкодеров



для предобучения каждого слоя и само предобучение со сбором статистики. С целью соблюдения принципа единственности ответственности эта логика инкапсулирована в отдельном классе.

Дополняя описанную выше функциональную декомпозицию программной логики работы ИНС, рассмотрим UML - диаграммы последовательности, приведённые на рисунках 1 и 2.

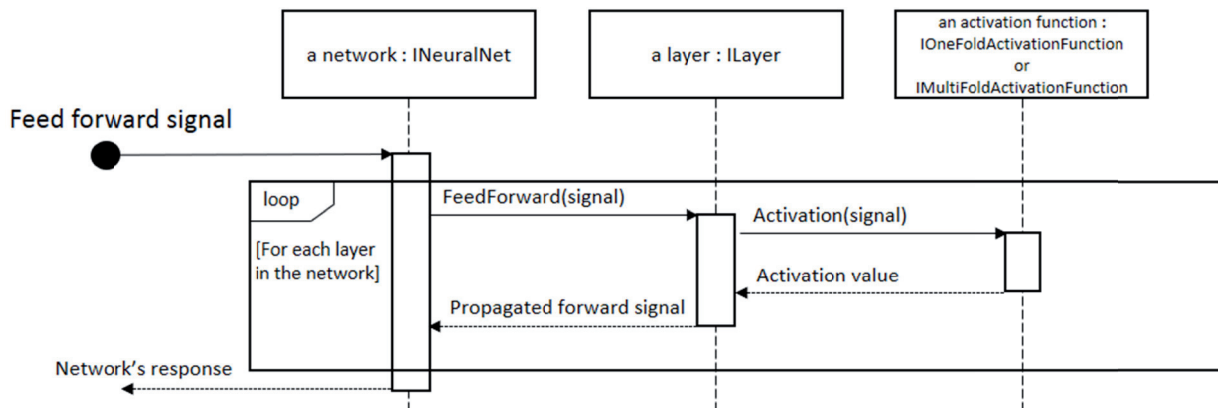


Рисунок 1 - UML-диаграмма последовательности, иллюстрирующая переход управления при прямом прохождении сигнала в ИНС

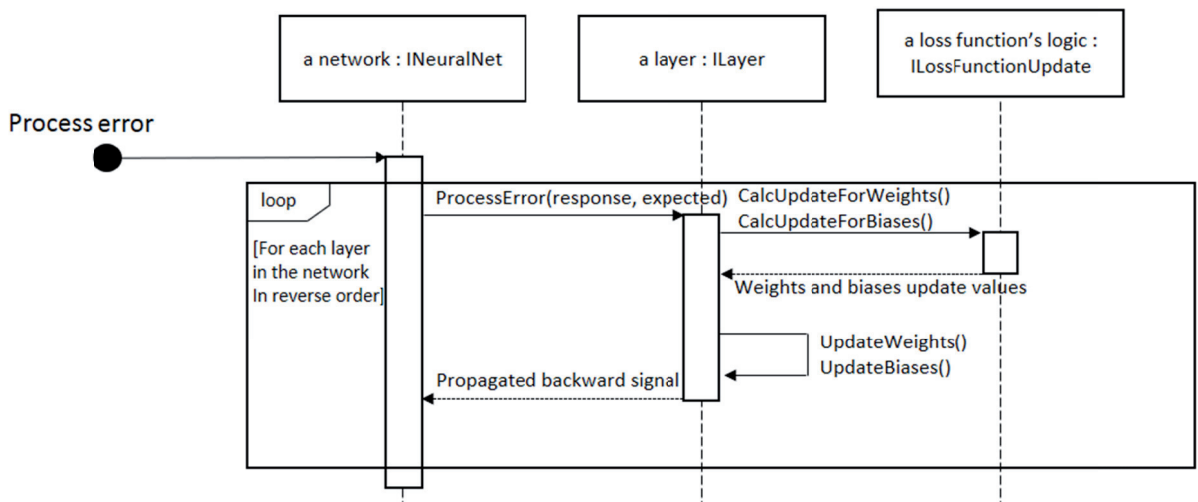


Рисунок 2 - UML-диаграмма последовательности, иллюстрирующая переход управления при обратном прохождении сигнала в ИНС

На приведённых диаграммах легко видеть взаимосвязь между выделенными в ходе функциональной декомпозиции совокупностями функциональностей. При прямом прохождении сигнала (см. рисунок 1) ИНС последовательно делегирует задачу прямой обработки сигнала каждому из входящих в неё слоёв нейронов. Каждый из слоёв нейронов, в свою очередь, делегирует объекту класса функции активации задачу вычисления её значения. При обратном прохождении сигнала (см. рисунок 2) ИНС последовательно делегирует задачу обратной обработки сигнала каждому из входящих в неё слоёв нейронов. Переход управления от слоя к слою в этом случае выполняется, начиная с последнего и заканчивая первым.

В некоторых технических решениях требования таковы, что необходимо использовать уже обученную нейронную сеть, без последующей её корректировки и дообучения. Для обеспечения эксплуатации нейронной сети в таких условиях важно выполнить группировку функционала, составляющего ИНС, в двух вариантах: функционал, необходимый для эксплуатации и функционал, необходимый для обучения.

Рассматривая выполненную выше структурную декомпозицию, легко заметить, что для проектирования и эксплуатации ИНС достаточно иметь в распоряжении функцию активации, один или несколько слоёв нейронной сети и сущность, соответствующую самой нейронной сети. Сущность, соответствующая функции ошибки, требуется только в процессе обучения ИНС.

Функциональная декомпозиция показывает, что для работы ИНС достаточно иметь логику, обеспечивающую прямое прохождение сигнала, в совокупности с некоторой долей вспомогательной логики (например, логики, выполняющей матричные операции). Логика обратного прохождения сигнала и логика корректировки весов синаптических связей требуются только в процессе обучения ИНС.

Используя описанный подход при объектно-ориентированном проектировании, будем разделять логику ИНС на две группы классов:

- классы, инкапсулирующие логику обучаемых ИНС и их компонентов;
- классы, инкапсулирующие логику необучаемых ИНС и их компонентов.

Важно заметить, что обучаемые ИНС в себе содержат логику необучаемых, поэтому эти классы связаны наследованием.

В приводимой декомпозиции выделяется логика нейронов, находящихся на отдельно взятом слое ИНС. Дальнейшая декомпозиция на уровне логики отдельно взятых нейронов не выполняется в силу следующих соображений. На практике в структуре нейронных сетей на отдельно взятом слое используются нейроны схожей конфигурации. Учитывая это, арифметические операции, выполняемые на отдельно взятом слое ИНС, можно выразить и реализовать в коде в виде матричных операций. Некоторые матричные операции, такие как умножение, достаточно ресурсоёмки и обладают большой асимптотической сложностью [15], однако при этом они обладают алгоритмическим параллелизмом. Кроме того, такие вычисления можно выполнять на графических процессорах GPU (*Graphics Processing Unit*) в рамках концепции GPGPU (*General-purpose computing for GPU* - неспециализированные вычисления на графических процессорах) [12, 16].

## 2 Технология проектирования программного кода логики ИНС

При проектировании описываемой реализации ИНС использовались следующие основные принципы:

- функциональность отдельно взятой выделенной сущности должна представлять собой *логическую целостность* для удобства чтения и понимания кода;
- совокупность выделенных сущностей должна обладать достаточной степенью *слабой связанности* для обеспечения возможности независимо развивать, тестировать и использовать различные сущности;
- каждый отдельно взятый класс, инкапсулирующий какую-либо логику, должен соответствовать принципу *единственной ответственности*, то есть быть ответственным за решение какой-либо отдельно взятой задачи и только её.

Следование этим принципам нацелено на проектирование кода так, чтобы заложенные концепции позволяли удобно компоновать различные ИНС и расширять их функционал (добавление новых функций активации, функций оценки ошибки, поддержка различных конфигураций ИНС и т.п.) [5].

На основании проведённых выше анализа и декомпозиции программной логики ИНС была выполнена её имплементация с использованием языка C#. На рисунке 3 приведена UML-диаграмма, которая даёт хорошее представление о перечне и связях, используемых в какой-либо программной библиотеке классов.

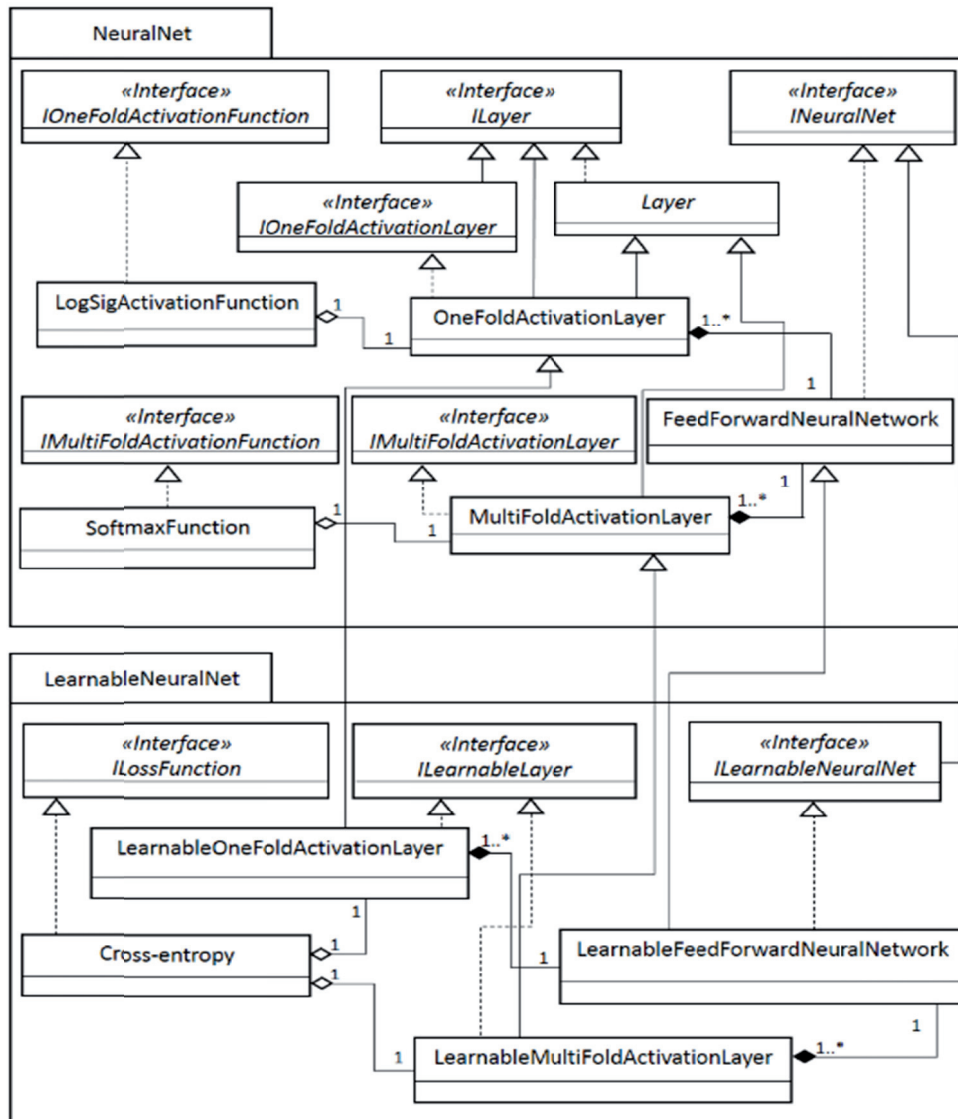


Рисунок 3 - UML-диаграмма классов, иллюстрирующая объектно-ориентированную декомпозицию программной логики ИНС

На диаграмме легко выделить два пакета. В пакете, расположенном в верхней части изображения, содержатся интерфейсы и классы, обеспечивающие прямое прохождение сигнала в ИНС. В нем расположены интерфейсы, объявляющие специфичные члены для многосвёрточных и односвёрточных функций активации, `IMultiFoldActivationFunction` и `IOneFoldActivationFunction`, соответственно. Для наглядности представлена одна реализация каждого из этих интерфейсов (классы `LogSigActivationFunction` и `SoftmaxFunction`). В действительности их может быть больше в зависимости от числа поддерживаемых функций активации. Интерфейс `ILayer` объявляет перечень членов, специфичных для отдельно взятого слоя нейронов. Его непосредственной реализацией является абстрактный класс `Layer`, который задаёт базовую реализацию, общую как для слоёв содержащий многосвёрточную функцию активации, так и односвёрточную. Наследники класса `Layer`, классы `MultiFoldActivationLayer` и `OneFoldActivationLayer`, содержат реализацию логики многосвёрточных и односвёрточных слоёв. Интерфейс `INeuralNet` объявляет члены, присущие непосредственно ИНС. Его реализацией является класс `FeedForwardNeuralNetwork`, название которого говорит само

за себя – он реализует функциональность сети, выполняющей прямое прохождение сигнала. Разумеется, можно добавить реализацию и других нейронных сетей посредством реализации интерфейса *INeuralNet*. Каждый класс ИНС, и в частности, *FeedForwardNeuralNetwork*, образует композицию с одним или более слоями нейронов.

В пакете, расположенном в нижней части диаграммы, находятся интерфейсы и классы, обеспечивающие обучение ИНС. Интерфейс *ILossFunction* объявляет члены, присущие функциям потерь. Приведена одна из реализаций данного интерфейса – класс *Cross-Entropy*. Интерфейсы *ILearnableLayer* и *ILearnableNeuralNet* объявляют перечень членов, необходимых для обучения отдельно взятого слоя и сети в целом, соответственно. Классы *LearnableMultiFoldActivationLayer* и *LearnableOneFoldActivationLayer* – потомки *MultiFoldActivationLayer* и *OneFoldActivationLayer*, соответственно. Они добавляют своим предкам реализацию *ILearnableLayer*, а класс *LearnableFeedForwardNeuralNetwork* добавляет своему предку *FeedForwardNeuralNetwork* реализацию *ILearnableNeuralNet*.

Приведённая диаграмма более детально отражает описанную выше структурную декомпозицию логики ИНС. Используются два интерфейса, описывающих классы, содержащие логику вычисления значений функции активации, *IOneFoldActivationFunction* и *IMultiFoldActivationFunction* (см. рисунок 4).

Совокупность одного или нескольких слоёв нейронов образует собой ИНС. Интерфейс, соответствующей ей, представлен на рисунке 5.

```
public interface IOneFoldActivationFunction
{
    double Activation(double parameter);
    double Derivative(double parameter);
}
public interface IMultiFoldActivationFunction
{
    double[] Activation(double[] parameters);
    double[,] Derivative(double[] parameters);
}
```

Рисунок 4 – Интерфейсы *IOneFoldActivationFunction* и *IMultiFoldActivationFunction*

```
public interface INeuralNet
{
    double[] ProcessInput(double[] input);
    IEnumerable<ILayer> Layers { get; }
}
```

Рисунок 5 - Интерфейс *INeuralNet*

Поскольку слой нейронов ИНС обладает значительной частью логики, которая является общей для многих разновидностей слоёв нейронных сетей, она вынесена в базовый интерфейс *ILayer*. Такая практика выделения базовых сущностей на основании анализа признаков общности и различности необходима, чтобы не дублировать код [6, 18]. Наследниками данного базового интерфейса являются классы *OneFoldActivationLayer* и *MultiFoldActivationLayer*.

Как видно из названий этих классов, они декларируют только те члены, которые необходимы для реализации логики слоя нейронной сети в той части, где они различны в случае применения односвёрточной и многосвёрточной функции активации. Интерфейсы *ILayer*, *IOneFoldActivationLayer* и *IMultiFoldActivationLayer* приведены на рисунке 6.

```

public interface ILayer
{
    double[] Biases { get; }
    double[,] Weights { get; }
    int NumberOfInputs { get; }
    int NumberOfNeurons { get; }
    double[] FeedForward(double[] inputVector);
    //некоторые члены интерфейса опущены для краткости
}
public interface IOneFoldActivationLayer : ILayer
{
    IOneFoldActivationFunction OneFoldActivationFunction { get; }
}
public interface IMultiFoldActivationLayer : ILayer
{
    IMultiFoldActivationFunction MultiFoldActivationFunction { get; }
}

```

Рисунок 6 – Интерфейсы ILayer, IOneFoldActivationLayer и IMultiFoldActivationLayer

Следование принципу разделения объявления и реализации программных сущностей позволяет в дальнейшем гибко изменять функциональность путём использования разных реализаций заданных интерфейсов. Имплементация сформированного набора интерфейсов выполнена следующим образом.

Одна из реализаций интерфейса *IOneFoldActivationFunction* – это класс *LogSigActivationFunction*, который представлен на рисунке 7.

```

public class LogSigActivationFunction :
IOneFoldActivationFunction
{
    public double Activation(double parameter)
    {
        return 1.0 / (1.0 + Math.Pow(Math.E, (-1.0 *
parameter)));
    }
    public double Derivative(double parameter)
    {
        double e = 1.0 / (1.0 + Math.Pow(Math.E, (-1.0 *
parameter)));
        return e * (1.0 - e);
    }
}

```

Рисунок 7 - Класс *LogSigActivationFunction*

Логика этого класса достаточно проста, поэтому паттерн «Внедрение зависимостей» (*Dependency injection*) [19, 20] в данном случае не применяется. Однако этот класс служит примером, образно говоря, начального элемента цепочки зависимостей между доменными классами рассматриваемой программной библиотеки.

В классах, имплементирующих интерфейсы *IOneFoldActivationFunction* и *IMultiFoldActivationFunction*, применённая декомпозиция позволяет сосредоточиться только на логике вычисления значений функции активации в зависимости от переданных парамет-

ров и вычислении значений производной данной функции. Это полностью соответствует принципу единственной ответственности для отдельного взятого класса.

Следующим логическим блоком в рассматриваемой логике является слой нейронов ИНС. Класс *OneFoldActivationLayer* показан на рисунке 8.

```

public class OneFoldActivationLayer : Layer,
IOneFoldActivationLayer
{
    public OneFoldActivationLayer(double[,] weights,
double[] biases, IOneFoldActivationFunction af)
        : base(weights, biases)
    {
        //имплементация конструкторов не приведена
для краткости
    }
    public IOneFoldActivationFunction
OneFoldActivationFunction
    {
        get
        {
            return oneFoldActivationFunction;
        }
    }
    //другие члены класса не приведены для
краткости
}

```

Рисунок 8 - Класс *OneFoldActivationLayer*

Видно, что один из параметров конструкторов данного класса – это реализация *IOneFoldActivationFunction*. Реализация этого класса выполняет делегирование соответствующих вычислений переданной реализации. Для краткости имплементация класса *Layer* не приводится, но в ней содержится общая базовая функциональность, специфичная для слоя нейронов ИНС.

Аналогичным образом зависимость внедряется и для связки нейронной сети в целом, и составляющих её слоёв нейронов. Конструктор класса принимает в качестве параметра перечень слоёв, входящих в состав нейронной сети, а впоследствии делегирует выполнение прямой передачи сигнала от слоя к слою, обеспечивая прямую передачу сигнала по сети в целом.

Таким образом, вся логика ИНС, декомпозированная описанным способом, объединена в общую функциональность посредством концепции «Внедрение зависимостей» (рисунок 9).

Совокупность функциональностей, соответствующая логике обучения ИНС, реализована следующим образом. Выделен интерфейс *ILossFunction*. Различные имплементации данного интерфейса позволяют привнести в описываемую библиотеку использование различных функций потерь. Посредством объекта, имплементирующего данный интерфейс, определяются значения, на которые корректируются значения весов и сдвигов.

Независимо от того, какая имплементация слоя нейронов используется (одна из описанных выше или добавленная в ходе доработки библиотеки), слой нейронов можно использовать в обучаемой ИНС при выполнении всего одного условия: класс, описывающий этот слой, должен реализовывать интерфейс *ILearnableLayer*. В логике этого класса выполняется делегирование задачи вычисления величины корректировки весов и сдвигов объекту, реализующему *ILossFunction*.

```

public class FeedForwardNeuralNetwork : INeuralNet
{
    protected List<ILayer> layers = new
List<ILayer>();
    public FeedForwardNeuralNetwork(params
ILayer[] layers)
    {
        //имплементация конструктора не приводится
для краткости
    }
    public IEnumerable<ILayer> Layers
    {
        get
        {
            return layers;
        }
    }
    public double[] ProcessInput(double[] input)
    {
        for (int i = 0; i < layers.Count; i++)
            input = layers[i].FeedForward(input);
        return input;
    }
}

```

Рисунок 9 - Класс *FeedForwardNeuralNetwork* как одна из реализаций интерфейса *INeuralNet*

Интерфейс *ILearnableNeuralNet* расширяет интерфейс *INeuralNet* перечнем членов, необходимых для выполнения обучения ИНС.

Таким образом, для получения функциональности обучаемой ИНС достаточно выполнить реализацию интерфейсов *ILossFunction*, *ILearnableLayer* и *ILearnableNeuralNet*. В качестве примера на рисунке 10 показана одна из реализаций обучаемой ИНС в виде класса *LearnableFeedForwardNeuralNetwork*.

Важно отметить, что такая реализация является наследником класса *FeedForwardNeuralNetwork* для предотвращения дублирования кода, описывающего логику прямого прохождения сигнала. Как было подробно описано выше, *FeedForwardNeuralNetwork* инкапсулирует логику и все необходимые ссылки на объекты для обеспечения работы логики прямого прохождения сигнала.

### 3 Модульное тестирование

Предлагаемая методика реализации ИНС обеспечивает приемлемую производительность и точность нейронных сетей, обученных на массивах данных [21], а также даёт возможность выполнять эксперименты, связанные с апробацией различных конфигураций ИНС [22]. Повышение качества существующего кода достигается посредством рефакторинга [5, 23], технология которого предполагает применение модульного тестирования с использованием юнит-тестов (*unittesting*). Рассмотрим предлагаемую декомпозицию логики ИНС с точки зрения удобства реализации юнит-тестов.

Для примера проведём тестирование логики класса *OneFoldActivationLayer*. Один из юнит-тестов, выполняющий проверку корректности метода *FeedForward* данного класса, приведён на рисунке 11.

```

public class LearnableFeedForwardNeuralNetwork :
FeedForwardNeuralNetwork, ILearnableNeuralNet
{
    private List<ILearnableLayer> learnableLayers = new
List<ILearnableLayer>();
    public LearnableFeedForwardNeuralNetwork(params
ILearnableLayer[] layers) : base(layers)
    {
        learnableLayers.AddRange(layers);
    }
    public IEnumerable<ILearnableLayer> LearnableLayers
    {
        get
        {
            return learnableLayers;
        }
    }
    public void ProcessError(double[] target, double alpha,
double momentum)
    {
        double[] error =
learnableLayers.Last().ErrorVectorFrom(target);
        for (int i = learnableLayers.Count - 1; i >= 0; i--)
            error = learnableLayers[i].ProcessError(target,
error, alpha, momentum);
        foreach (var layer in learnableLayers)
        {
            layer.UpdateWeights();
            layer.UpdateBiases();
        }
    }
    //некоторые другие члены не приведены для
краткости
}

```

Рисунок 10 - Класс *LearnableFeedForwardNeuralNetwork*

Легко видеть, что подменой логики вычисления функции активации (impleментация интерфейса *IOneFoldActivationFunction*) в ходе тестирования можно проверить логику работы метода *FeedForward* класса *OneFoldActivationLayer* независимо от других классов. Это очень удобно, т.к. позволяет легко локализовать ошибки в коде и вносить корректировки.

Таким же образом можно проверить корректность использования хранимых данным классом весов, логику вычисления поля индуцирования входных значений для функции активации. При этом достаточно использовать в качестве, как их часто называют на практике, тестового дублёра достаточно простую условную реализацию интерфейса *IOneFoldActivationFunction*.

Аналогичным образом упрощается проверка реализации интерфейса *INeuralNet* класса *FeedForwardNeuralNetwork*. Путём внедрения зависимости в виде тестовых дублёров, представленных в виде значительно упрощённых имплементаций интерфейса *ILayer*, легко проверяется логика работы класса *FeedForwardNeuralNetwork*. Подобное разделение позволило декомпозировать достаточно сложную логику работы ИНС в несколько относительно независимых блоков. При этом каждый из них имеет свою семантическую целостность в сочетании с умеренной сложностью.



```

[TestMethod]
public void OneFoldActivationLayerFeedForward()
{
    //Arrange
    double[,] weightMatrix = new double[,] { { 1, 1, 1, 1
}, { 1, 1, 1, 1 } };
    double[] biasVector = new double[] { 1, 1 };
    OneFoldActivationLayer layer = new
OneFoldActivationLayer(weightMatrix, biasVector, new
DummyOneFoldActivationFunction());
    //Act
    double[] output = layer.FeedForward(new double[] {
1, 2, 3, 4 });
    //Assert
    Assert.AreEqual(11, output[0]);
    Assert.AreEqual(11, output[1]);
}

```

Рисунок 11 - Юнит-тест, проверяющий корректность метода *FeedForward* класса *OneFoldActivationLayer*

## Заключение

Выполненная реализация декомпозиции программной логики ИНС опирается на использование объектно-ориентированной парадигмы. Предложенное разделение программной логики между классами и обеспечение слабой связанности между ними упрощает процесс поиска ошибок в коде, делает код более управляемым, повышает продуктивность разработчика при проведении рефакторинга. Приведённые примеры демонстрируют, как рациональная декомпозиция логики на семантически независимые между собой блоки в сочетании с применением концепции «Внедрение зависимостей» способствует большей структурированности кода.

Предлагаемый подход к декомпозиции программной логики ИНС можно применить к широкому спектру различных нейронных сетей: конволюционные и деконволюционные сети; сети, содержащие одновременно конволюционные и полносвязанные слои нейронов; и т.п. При этом единообразие реализации различного рода сетей обуславливает упрощение их понимания, снижает трудозатраты на сопровождение и развитие реализующего их кода.

## Список источников

- [1] *Хайкин, С.* Нейронные сети: полный курс, 2-е изд. – М.: Издательский дом «Вильямс», 2006. – 1104 с.
- [2] *Russell, S.* Artificial Intelligence: A Modern Approach, 3<sup>rd</sup> Edition / S. Russell, P. Norvig. - Pearson Education, 2010. - 1095 p.
- [3] *Мак-Каллок, У.С.* Логическое исчисление идей, относящихся к нервной активности / У.С. Мак-Каллок, У. Питтс // В сб. «Автоматы» под ред. К.Э. Шеннона и Дж. Маккарти. — М.: Изд-во иностр. лит., 1956. — С. 363—384.
- [4] *Fowler, M.* Patterns of Enterprise Application Architecture / M. Fowler. - Addison-Wesley, 2002. – 560 p.
- [5] *Jacobsen, I.* Object Oriented Software Engineering: A Use Case Driven Approach / I. Jacobsen, M. Christerson, P. Jonsson, G. Overgaard. - Addison-Wesley / ACM Press, 1992. – 56 p.
- [6] *Gamma, E.* Design Patterns: Elements of Reusable Object-Oriented Software / E. Gamma, R. Helm, R. Johanson, J. Vlissides. – Addison-Wesley Professional, 1994. - 395 p.
- [7] *Буч, Г.* Объектно-ориентированный анализ и проектирование с примерами приложений. 3-е изд. / Г. Буч. - М.: ООО «И.Д. Вильямс», 2008. - 720 с.
- [8] *Richter, J.* CLR via C#, 4th Edition / J. Richter. - Microsoft Press, 2012. – 896 p.
- [9] OpenNN, Сайт проекта. – <http://www.opennn.net/>
- [10] *Yangqing, J.* Caffe, Сайт проекта. – <http://caffe.berkeleyvision.org/>

- [11] *Nielsen, M.* Neural Networks and Deep Learning, free online book / M. Nielsen. – <http://neuralnetworksanddeeplearning.com/>.
  - [12] *Vasilev, I.* Java deep learning library, Репозиторий проекта. – <https://github.com/ivan-vasilev/neuralnetworks/>
  - [13] *Rumelhart, D.E.* Learning Internal Representations by Error Propagation / D.E. Rumelhart, G.E. Hinton, R.J. Williams // Parallel Distributed Processing: explorations in the microstructure of cognition. Cambridge, MA: MIT Press.– 1986. - Vol. 1. – P. 318-362.
  - [14] *Goodfellow, I.* Deep learning / I. Goodfellow, Y. Bengio, Y. Courville. – <http://www.deeplearningbook.org/>
  - [15] *Sedgewick, R.* Algorithms. 4th Edition / R. Sedgewick, K. Wayne. - Addison-Wesley Professional, 2011. – 992 p.
  - [16] *Mittal, S.* A Survey of CPU-GPU Heterogeneous Computing Techniques / S. Mittal, J. Vetter // ACM Computing Surveys (CSUR). - 2015. – Vol.47(4). - P. 1–35.
  - [17] *Буч, Г.* Язык UML. Руководство пользователя / Г. Буч, Дж. Рамбо, И. Якобсон. – М.: ДМК Пресс, 2006. – 496 с.
  - [18] *Troelsen, A.* Pro C# 7: With .NET and .NET Core. 8th Edition / A. Troelsen, P. Japikse. - Apress, 2017. – 1410 p.
  - [19] *Seemann, M.* Dependence Injection in .Net. - Manning, 2011. - 584 p.
  - [20] *Fowler, M.* Inversion of Control Containers and the Dependency Injection pattern / M. Fowler - <https://www.martinfowler.com/articles/injection.html/>.
  - [21] *Grother, P.* NIST Special Database 19 Handprinted Forms and Characters, 2nd Edition / P. Grother, K. Hanaoka. - National Institute of Standards and Technology, 2016. – 30 p.
  - [22] *Гурин, Р.В.* Двухстадийная нормализация выходных сигналов искусственных нейронных сетей / Р.В. Гурин, С.П. Орлов // Вестник СамГТУ. Серия «Технические науки». – 2017. - №4(56). - С. 7-16.
  - [23] *Fowler, M.* Refactoring: Improving the Design of Existing Code, 1st Edition / M. Fowler. - Addison-Wesley Professional, 1999. – 431 p.
- 

## OBJECT-ORIENTED DECOMPOSITION OF ARTIFICIAL NEURAL NETWORK'S PROGRAM LOGIC

R.V.Girin<sup>1</sup>, S.P. Orlov<sup>2</sup>

<sup>1</sup>ООО "Intellect Soft", Samara, Russia  
[romangirin@gmail.com](mailto:romangirin@gmail.com)

<sup>2</sup>Samara State Technical University, Samara, Russia  
[orlovsp1946@gmail.com](mailto:orlovsp1946@gmail.com)

### Abstract

The article presents a decomposition of neural networks logic for its implementation as a set of loosely coupled domain classes. The structural and functional decomposition of logic are considered. The technique of program logic decomposition is illustrated by UML diagrams. The implementation an artificial neural network's logic in the C # language based on object-oriented approach is briefly described. A similar implementation was used in the design of neural networks configurations for subsequent training, operation, and experimentation. The proposed separation of program logic between classes, as well as providing weak connectivity between them simplifies the process of errors localization in the program code, makes the code more controllable, besides increased developer productivity. The examples demonstrated in the article show how the rational decomposition for neural network program logic with semantically mutually independent units in combination with "Dependency injection" concept, contributes to a more structured code. The new result is the application of object-oriented decomposition for neural network logic, which allows to significantly simplify the process of designing code. An example of testing the neural network decomposed logic is presented. The proposed approach to decomposition of artificial neural networks program logic can be applied to a wide range of different neural networks, such as convolutional, deconvolutional networks, the networks containing fully connective layers of neurons. In this case, the uniformity of implementation of neural networks to simplify their understanding of the causes reduces labor costs for maintenance and development of implementing their code.

**Key words:** *artificial neural network, system analysis, decomposition, object-oriented analysis.*

**Citation:** *Girin RV, Orlov SP.* Object-oriented decomposition of artificial neural network's program logic [In Russian]. *Ontology of designing.* 2018; 8(1): 110-123. DOI: 10.18287/2223-9537-2018-8-1-110-123.

## References

- [1] **Hykin S.** Neural Networks. A Comprehensive Foundation. 2nd Edition. Prentice Hall; 1999.
- [2] **Russell S, Norvig P.** Artificial Intelligence: A Modern Approach, 3rd Edition. Pearson Education; 2010.
- [3] **McCulloch WS, Pitts W.** A logical calculus of ideas immanent in nervous activity // Bull. Math. Biophys., - 1943. – v.5. – p.115–133.
- [4] **Fowler M.** Patterns of Enterprise Application Architecture. Addison-Wesley; 2002.
- [5] **Jacobsen I, Christerson M, Jonsson P, Overgaard G.** Object Oriented Software Engineering: A Use Case Driven Approach. Addison-Wesley/ACM Press; 1992.
- [6] **Gamma E, Helm R, Johanson R, Vlissides J.** Design Patterns: Elements of Reusable Object-Oriented Software. Addison Wesley Professional; 1994.
- [7] **Booch G.** Object-oriented analysis and design with applications. 3rd Edition [InRussian]. - Moscow: I.D. Williams; 2008.
- [8] **Richter J.** CLR via C#, 4th edition. Microsoft Press; 2012.
- [9] OpenNN, Project Site. - <http://www.opennn.net>
- [10] **Yangqing J.** Caffe, Project Site. -<http://caffe.berkeleyvision.org>
- [11] **Nielsen M.** Neural Networks and Deep Learning, free online book. - <http://neuralnetworksanddeeplearning.com>
- [12] **Vasilev I.** Java deep learning library, project repository. - <https://github.com/ivan-vasilev/neuralnetworks>
- [13] **Rumelhart DE, Hinton GE, Williams RJ.** Learning Internal Representations by Error Propagation. Parallel Distributed Processing:explorations in the microstructure of cognition.Cambridge, MA: MIT Press 1986; 1: 318-362.
- [14] **Goodfellow I, Bengio Y, Courville Y.** Deep learning. - <http://www.deeplearningbook.org>
- [15] **Sedgewick R, Wayne K.** Algorithms. 4th Edition. Addison-Wesley Professional; 2011.
- [16] **Mittal S, Vetter J.** A Survey of CPU-GPU Heterogeneous Computing Techniques. ACM Computing Surveys (CSUR) 2015; 47 (4): 1–35.
- [17] **Booch G, Rambo J, Jacobson I.** The Unified Modeling Language. Users Guide. Addison-Wesley; 2005.
- [18] **Troelsen A, Japikse P.** Pro C# 7: With .NET and .NET Core. 8th Edition. Apress; 2017.
- [19] **Seemann M.** Dependence Injection in .Net. Manning; 2011.
- [20] **Fowler M.** Inversion of Control Containers and the Dependency Injection pattern. - <https://www.martinfowler.com/articles/injection.html>
- [21] **Grother P, Hanaoka K.** NIST Special Database 19 Handprinted Forms and Characters, 2nd Edition. National Institute of Standards and Technology; 2016.
- [22] **Girin RV, Orlov SP.** Two-stage normalization of output signals of artificial neural networks [In Russian]. Bulletin of Samara State Technical University. “Technical Sciences” 2017; 4(56): 7-16.
- [23] **Fowler M.** Refactoring: Improving the Design of Existing Code, 1st Edition. Addison-Wesley Professional; 1999.

## Сведения об авторах



**Гирин Роман Викторович**, 1979 г. рождения. Окончил Самарский государственный технический университет в 2003 г. Разработчик проектов в компании ООО «Интеллект Софт», Самара. Соискатель ученой степени кандидата технических наук при Самарском государственном техническом университете. Область научных интересов – нейронные сети и их приложения.

**Roman Victorovich Girin** (b. 1979) graduated Samara Technical State University in 2003. He is project developer of ООО “Intellect Soft”, Samara. He is grandaunt to the degree of Cand.Tech.Sci. at the Samara State Technical University. The field of scientific interests is neural networks and their applications.

**Орлов Сергей Павлович**, 1946 г. рождения. Окончил Куйбышевский политехнический институт в 1969 г., д.т.н.(1991), профессор.

Заведующий кафедрой «Вычислительная техника» Самарского государственного технического университета. Член Международной академии информатизации, член международного общества IEEE. В списке научных трудов более 180 работ в области компьютерных технологий, моделирования сложных систем, создания интеллектуальных систем поддержки принятия решений в технических и социальных системах.

**Sergey Pavlovich Orlov** (b. 1946) graduated from the Kuibyshev Polytechnic Institute in 1969, D. Sc. Eng. (1991), professor. He is Head of Computer Technology Department of Samara State Technical University. He is member of International Informatization Academy and IEEE member. He is co-author of more than 180 publications in the field of computer technology, complex systems simulation, knowledge based decision support systems in technical and social systems.



УДК 72.01

## ОНТОЛОГИЯ ТЕРМИНА «ПРИНЦИП СОВМЕСТИМОСТИ» В РЕСТАВРАЦИОННОМ ПРОЕКТИРОВАНИИ

О.А. Егорова

Санкт-Петербургский государственный архитектурно-строительный университет,  
Санкт-Петербург, Россия  
olga\_2802@list.ru

### Аннотация

В статье обсуждается вопрос актуальности терминологической базы современного реставрационного проектирования. В качестве примера рассматриваются термины «совместимость» и «принцип совместимости», обозначаются их понятийные пространства, раскрывается сущность применения данных терминов. На основе анализа основных терминологических дефиниций, полученных из словарей, учебников и научной литературы, обозначены онтологический и гносеологический аспекты понятия «совместимость», как фундаментального принципа реставрационного проектирования. На основе анализа лингвистических различий понятия «совместимость» в русском, итальянском и английском языках выявлены сущностные различия в употреблении данного термина, а так же в реализации «принципа совместимости» в реставрации памятников архитектуры. Рассматривается проблема понимания термина «совместимость», в зависимости от страны и языка, на котором данный термин трактуется. В статье предложено новое определение понятия «совместимость», рекомендуемое к включению в словарь реставратора-проектанта.

**Ключевые слова:** реставрационное проектирование, совместимость, «принцип совместимости», понятийный аппарат, терминология.

**Цитирование:** Егорова, О.А. Онтология термина «принцип совместимости» в реставрационном проектировании / О.А. Егорова // Онтология проектирования. – 2018. – Т. 8, №1(27). – С.124-133. – DOI: 10.18287/2223-9537-2018-8-1-124-133.

«..**принцип совместимости**. В нём отражён объективный закон существования любого целостного явления, и особенно таких высокоцелостных систем, как живые, социальные, технические системы. Суть данного закона ...: необходимым условием существования целостного объекта является совместимость его компонентов друг с другом, т.е. чтобы они сопрягались, не отторгали друг друга»<sup>1</sup>

Антонюк Г.А., доктор философских наук,  
профессор (Беларусь, Минск)

### Введение

Понятие «совместимость» встречается в разных сферах жизни. Психологическая и культурная совместимость, совместимость материалов и конструкций в строительстве, межличностная совместимость, медицинская совместимость, совместимость техники, технологий, программного обеспечения и пр. Важность совместимости возрастает пропорционально росту темпов технического развития, обмена информацией и процессу глобализации. При решении вопросов, связанных с совместимостью, часто возникают споры и недопонимания,

---

<sup>1</sup> Антонюк Г.А. Принцип совместимости в социальном проектировании // Всесоюзная научно-теоретическая конференция по фундаментальной междисциплинарной проблеме «Организация и управление». – Тез. докл.- Минск, 13-15 ноября 1989 г. – С. 154-159. Дополнено редакцией.

поскольку трактовка данного термина неоднозначна. Особенно это касается вопроса совместимости теоретических понятий и концепций, на который влияет разнообразие культур, характеров, обычаев, традиций, профессионального подхода, человеческого воспитания, а также политика государства и настроения социума. Ситуация усугубляется тем, что терминология и её взаимопереводы на языки разных стран не являются абсолютными эквивалентами друг другу.

В хартии «Ла Валлетта» 2011 года [1], посвящённой сохранению и управлению историческими городами, лейтмотивом проходит тема совместимости. Она объявляется обязательным условием взаимодействия современного человека и общества с исторической средой. Необходимость совместимости обозначена.

Цель данной статьи - проследить соотношение и взаимосвязь дефиниций термина «совместимость» на разных языках (русском, итальянском и английском), для того, чтобы как можно точнее выявить разницу в языке и практическом применении указанного термина.

## 1 Проблема терминологии в науке и реставрационном проектировании

Термин - это смысловое ядро специального языка, передающее основную содержательную информацию. В результате роста научно-технического знания появляются новые слова, более 90% из них - *специальные слова*, т.е. термины и их определения [2]. От слов общепринятой лексики они отличаются специализированностью значения и точностью. Становясь терминами, слова обретают необходимую однозначность, обозначают то или иное действие или понятие, уточняют его, отделяют от смежных. Кроме однозначности, к терминам предъявляются следующие требования: доступность, краткость и точность. Нарушение данных норм приводит к тому, что терминология становится непонятной [3].

В реставрационном проектировании развитие понятийного аппарата напрямую связано с обоснованием проектных предложений, с выбором технологий и методик проведения работ на памятнике. Устаревшая терминология усложняет процесс научной коммуникации, лишает его участников возможности вести конструктивный диалог. Основополагающий термин «реставрация» - до сих пор не унифицирован, трактуется по-разному, в зависимости от источника, точки зрения автора, сферы применения. Термин формировался в течение долгого времени, соответственно тому, как развивалась реставрационная практика, неоднократно менял свое значение.

В последнее время в русской терминологии появляется много новых понятий, заимствованных из иностранных языков, преимущественно европейских. Некоторые заимствованные или переведённые иностранные термины интерпретированы слишком обобщённо или неверно. Это происходит, если у понятия на иностранном языке есть несколько отличающихся по смыслу и оттенку определений. Если основой для перевода на русский язык становится только одно из них, а не совокупность смыслов каждого, то формулировка русифицированного термина оказывается однобокой и не раскрывает его сути. Сферы его применения оказываются ограничены.

Сложная ситуация может возникнуть и с исконно русскими словами. Встречаются определения, не раскрывающие смысла термина. Если слово переходит из ранга общеупотребительных в специальные, возникает потребность его идентификации в научной сфере, необходимость в уточнении его значения. Русский язык располагает возможностью точно передавать смысловые и стилистические оттенки. Поэтому вопрос неоднозначности некоторых терминов, связан со спецификой национального мышления.

Если термин становится часто употребляемым, особенно на международном уровне, а его определение на русском языке является неточным или неполным – это существенно замедля-

ет процесс развития, совершенствования теоретической базы и обмен опытом между странами. Трудность и неоднозначность перевода могут оказаться препятствием, а могут стать ключом к решению фундаментальных проблем теоретического знания, помочь верно интерпретировать и адаптировать термин и его определение к национальному менталитету.

Ярким примером является слово «совместимый» - *такой, который можно с чем-нибудь совместить* [4]. Наиболее часто встречающееся определение термина «совместимость» имеет столь же лаконичную, малоинформативную формулировку, которая не даёт исчерпывающего представления о его значении. Подобная ситуация происходит, когда термин привнесён из других научных сфер и не был адаптирован к конкретной области, в рассматриваемом случае, для области реставрационного проектирования.

## **2 Понятия «совместимость» и «принцип совместимости» в реставрации памятников архитектуры**

«Принцип совместимости» был обозначен в 1964 году в Венецианской хартии [5]. Данный документ возник отчасти благодаря теории, которую предложил известный итальянский реставратор Чезаре Бранди [6, 7]. Он обозначал четыре главных принципа: принцип отличности, принцип обратимости, принцип минимального вмешательства и принцип совместимости. При этом следует обратить внимание на то, что если с первыми тремя принципами дело обстоит более-менее понятно, то «принцип совместимости» - это объект непрерывных обсуждений, споров и противоречий.

Руководитель кафедры реставрации Университета LaSapienza, профессор Джованни Карбонара говорит о том что, согласно данному принципу все новые включения должны существовать и взаимодействовать с подлинными частями памятника, не рискуя навредить им, как это происходит, например, при неправильном использования бетона в непосредственной близости или при контакте с фресками [8].

Сегодня реставраторы более осторожны в отношении совместимости. В своё время считалось, что суперстойкие и высокотехнологичные материалы идеально подходят для реставрационных работ, но это не всегда так. Считается, что материал должен иметь подобные характеристики или даже чуть более низкие показатели по сравнению с древним (подлинным, исходным) материалом, так чтобы в случае внешнего неблагоприятного воздействия на памятник архитектуры сначала изменениям подвергались новые включения, а только после этого сам памятник.

Мысль о необходимости соблюдения принципа совместимости прослеживается не только в иностранных источниках, посвященных вопросам реставрации. Отечественные реставраторы так же считали, что подбор материалов для проведения реставрации должен производиться с учётом специфики данной сферы. В пособии для архитекторов-реставраторов неоднократно упоминается о необходимости учёта совместимости материалов и конструкций при реставрации памятников архитектуры [9].

Описанные примеры показывают, что совместимости материалов и совместимости конструкций уделяется большое внимание. Это те части реставрационного проекта, результат которых можно оценить визуально. Но, до момента реализации утверждённых проектных предложений в разработке у архитекторов-реставраторов остаётся ещё целый ряд вопросов о совместимости идеологической: исторической, социальной, психологической и культурной. Ни одно суждение, в отсутствие упорядоченной системы категорий совместимости, не может однозначно восприниматься как правильное или ошибочное. Совместимость образа, функций, объекта и среды должны рассматриваться индивидуально на каждом архитектурном памятнике. Главный критерий совместимости - гармония. Одной из основных онтологи-

ческих задач современной реставрационной теории является создание и описание алгоритма выявления признаков, в совокупности приводящих к целостному восприятию архитектурного памятника и поздних дополнений.

### 3 Анализ лексических значений термина «совместимость» в разных языковых группах

На сегодняшний день не существует источника, в котором дано чёткое, однозначное определение понятию «*принцип совместимости*» в архитектуре и реставрации. Даже термина «*совместимость*» в специализированных архитектурных, строительных и реставрационных словарях нет.

Учебник по экономике предприятия содержит следующее определение: принцип совместимости - это «принцип, реализация которого обеспечивает создание единства составляющих производственный процесс элементов» [10]. По другим источникам принцип совместимости - это «международный принцип бухгалтерского учета, являющийся исходным моментом в построении бухгалтерских систем, подчеркивающий необходимость совместимости последних с организационными и человеческими факторами бизнеса» [11].

Чтобы вплотную подобраться к пониманию разницы реставрационных подходов в Италии и России возьмём за основу лексические значения и определения слов. Базовое определение слова «совместимость» или «*compabilità*» - *é possibilità di coesistere di convivere di dialogare* [12], что при дословном переводе обозначает «возможность сосуществовать, жить и вести диалог». Дополнительное определение «*che può essere compatito, perdonato*» [13], то, что может быть совместимым, сопутствующим, допущенным, прощённым». Также слово «совместимость» употребляется в значении «*sostenibile*», что в переводе означает «уважительный по отношению к окружающей среде, устойчивый в своём развитии».

В фармацевтике «совместимость» – понятие, используемое в отношении медикаментов, которые могут вводиться одновременно, так как они не мешают друг другу: не реагируют химически, физически или терапевтически.

В области ботаники совместимыми считаются разновидности фруктовых или плодовых деревьев, которые при перекрёстном опылении приносят плоды [14]. В математике уравнение, совместимое с другими, по итальянским источникам - такое, которое имеет с этими уравнениями общее решение; также это могут быть общие условия, которые могут сосуществовать, не вызывая логических противоречий.

Ключевыми словами, наиболее часто встречающимися в вышеуказанных определениях, являются понятия *жить вместе, вести диалог, жалеть, приносить плоды, допускать общее решение*. Все они звучат лояльно и дружелюбно, так же и по отношению к памятникам архитектуры. Если же говорить об определении слова «совместимость» в русскоязычной литературе, то можно выделить следующие:

*Совместимость* это:

- взаимное соответствие, сочетание, сосуществование в гармонии [15];
- способность устройств или систем, отличающихся конструктивными особенностями, выполнять идентичные функции, возможность совместной работы без создания взаимных помех [16];
- пригодность продукции для совместного использования при условии удовлетворения заданным требованиям без возникновения недопустимых последствий от их совместного применения [17];
- «вид отношений» между понятиями и суждениями (в логике два понятия называются совместимыми, если их объёмы совпадают полностью или частично) [18];

- степень, с которой защитные системы организма переносят вторжение чужеродного биологического материала», или «пригодность к смешиванию без неблагоприятных результатов» [19];
- отсутствие отрицательных результатов взаимного влияния различных материалов узла примыкания в процессе эксплуатации (совместимость материалов) [20].

Если применить определение термина «совместимость» в контексте архитектурной реставрации, то в русском языке может возникнуть представление, что действия по реставрации памятников архитектуры априори являются недружественными чужеродными внешними воздействиями. Хотя основной задачей реставрационного проектирования является «продление жизни памятника». При этом следует понимать разницу между продлением «жизни» и продлением «существования». Ясно, что основным принципом подобных формулировок является «не навреди».

Обращаясь к английскому языку «sustainable - designating, of, or characterized by a practice that sustains a given condition, as economic growth or a human population, without destroying or depleting natural resources, polluting the environment, etc.» [21], совместимость - ёмкое понятие обозначающее возможность поддержания текущего состояния чего-либо без истощения ресурсов. Совместимость является основой устойчивого развития. Более употребимое определение слова «совместимый» на английском - «compatible - able to exist or occur together without problems or conflict» [22]. Как и в русском языке, приведённая формулировка демонстрирует направленность на уход от конфликта. В основном, понятием «совместимость» в английском языке оперируют в отношении компьютерных технологий и программного обеспечения. Акцент - на процессе, получении взаимной выгоды, усовершенствовании одного посредством другого. Но существуют определения, которые близки по своему значению к итальянским аналогам. Встречаются слова: «гармония», «отношения».

Сравнительный анализ ключевых слов, встречающихся в определениях понятий «совместимость» и «совместимый» на различных языках приведён в таблице 1.

Таблица 1 – Сравнительный анализ ключевых слов, наиболее часто встречающихся в определениях понятий «совместимость» и «совместимый» на итальянском, русском и английских языках

Язык	Итальянский	Русский	Английский
Ключевые слова	Coesistere – сосуществовать, convivere – жить вместе, dialogare – вести диалог, si può conciliare con – примириться, che si può compatire – жалеть, perdonare – прощать, dare frutti приносить плоды, ammettere una soluzione comune - допускать общее решение, non essere in contrasto - не противоречить	Соответствие, сочетаемость, сосуществование в гармонии, идентичные функции, пригодность, смешивание без неблагоприятных результатов, отсутствие отрицательных результатов, совместная работа, чужеродное вторжение	Exist together – сосуществовать, harmonious relationship – гармоничные отношения, able to be used with – возможно использовать вместе с... cross-fertilizing – перекрестное опыление designed to work with another device – разработан для использования с другим приспособлением
Вывод	Ключевые слова отвечают на вопрос: « <b>Что делать?</b> » Акцент на взаимодействии, динамичность процесса	Ключевые слова отвечают на вопрос « <b>Что? Какие?</b> » Акцент на нейтральном или неотрицательном результате, статичность процесса	Ключевые слова отвечают на вопрос « <b>Что делать? Какие? Для чего?</b> » Акцент на процессе, получении взаимной выгоды, усовершенствовании одного посредством другого



#### 4 Понимание и реализация «принципа совместимости» в реставрационном проектировании

Языковая интерпретация понятия «совместимость» раскрывает суть неоднозначного решения вопросов реставрации в Российской Федерации. Если сравнивать реставрацию Италии и России, разница прослеживается на основе различия ключевых понятий (см. таблицу 1).

С момента издания трудов [6, 7] прошло уже более 50 лет. Наука остро нуждается в обновлении и актуализации понятийной базы. Реставрационное проектирование в России опирается на Свод реставрационных правил [23] и Реставрационные нормы и правила [24]. Сейчас появились новые ГОСТы [25, 26], в которых термин «совместимость» не встречается и упоминаний о необходимости проводить реставрационные работы с учётом «принципа совместимости» нет.

Если оставить ситуацию без изменений, то есть вероятность, что в дальнейшем свершившийся акт реставрации «по-русски» будет рассматриваться только как «статичный результат». В то время как более перспективным направлением может стать ориентация на итальянское – «начало совместной жизни».

#### 5 Перспективы развития и расширения понятия «совместимость» в отношении реставрационного проектирования

Так как архитектура и реставрация включают в себя и активно используют научную базу практически всех направлений от философии до механики, то уровней совместимости архитектурных объектов (памятников) и дополнений предполагается несколько. За основу можно взять градацию совместимостей, приводимую в психологии [27, 28].

В условиях совместимости «человек-объект» или «объект-объект» возникает несколько уровней совместимости, которые представлены на рисунке 1.

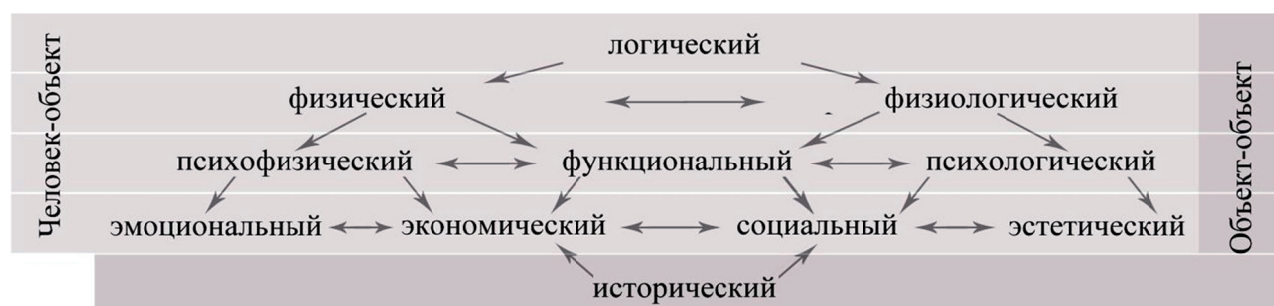


Рисунок 1 – Уровни совместимости

Эти уровни совместимости являются результатом анализа этапов осознания совместимости тех или иных элементов (архитектуры, материи, образа и т.д.) человеком в отношении какого-либо объекта. Первый уровень – логический, определяет саму возможность возникновения совместимости между теми или иными элементами. Решение на данном этапе принимается подсознательно, основываясь на имеющемся эмпирическом опыте и не вызывает противоречий. Следующий уровень – анализ физической и физиологической совместимости. На третьем уровне возникают психофизическая, функциональная и психологическая совместимость человека и окружающей среды – совокупности объектов, а также каждого объекта в отдельности. Если не возникает критических противоречий на предыдущих уровнях, то проводится анализ объекта по определённым критериям: эмоциональному, экономическому, социальному и эстетическому. Последняя ступень на рисунке отражает трансформацию характеристик вышеперечисленных уровней совместимости с течением времени в новый уро-

вень – исторический, представляющий собой их «квинтэссенцию» и неделимое целое. Указанный уровень совместимости возникает уже между объектами. Человеческая жизнь быстротечна, поэтому не может быть мерилom данных отношений. Человек становится наблюдателем. На данном этапе желание и потребность человека вмешиваться в совместимость тех или иных элементов объектов в условиях их исторической совместимости может оказать очень сильное влияние, а в случае пренебрежительного отношения - нанести непоправимый вред.

Предлагается скорректировать существующее определение термина «совместимость». *Совместимость – это совокупное взаимное соответствие различных свойств и характеристик объектов, способствующее их сосуществованию в гармонии, появившейся в момент созидания и перспективной в своем развитии.*

Если говорить о совместимости новых включений в реставрационном проектировании, то *совместимость дополнений - способность новых элементов обеспечивать и поддерживать заданные (необходимые) эстетические, социальные и функциональные свойства существующего объекта.*

### **Заключение**

Рассмотрев некоторые из существующих определения терминов «совместимость» и «принцип совместимости», можно констатировать тот факт, что их применение требует внесения корректировок в понятийный аппарат реставрационного проектирования.

Понятие совместимости требует нового онтологического прочтения. В настоящее время учёт принципа совместимости выходит на первый план при постановке реставрационных задач. Чаще всего встаёт вопрос технической и химической совместимости. Их отсутствие приводит к отрицательным последствиям для реставрируемого объекта. Соблюдение принципа совместимости ещё более важно на этапе теоретического осмысления самой идеи какого-либо преобразования памятника архитектуры. Международные конвенции и хартии указывают, что при взаимодействии с историческими зданиями и городской средой учёт выше-названной совместимости является обязательным. В российских нормативных и регулирующих документах (СНиП, ГОСТ по реставрации) отсутствует термин «совместимость». Следствием этого являются трудности в составлении технических заданий, отвечающих современным требованиям реставрационного проектирования.

Лингвистический сравнительный анализ понятия «совместимость» на русском, английском и итальянском языках позволяет внести бóльшую ясность и составить представление о существующих базовых отличиях восприятия указанного термина в разных культурных группах. Это способствует более широкому и разностороннему осмыслению термина и может стать основой для внесения изменений и дополнений в формулировку его определения на русском языке и включения в словарь реставратора-проектанта. В итоге мы можем получить новый терминологический инструмент, использование которого в реставрационном проектировании в настоящее время актуально и необходимо.

### **Список источников**

- [1] Принципы Валлетты по сохранению и управлению историческими городами и урбанизированными территориями. Приняты 17-ой Генеральной Ассамблеей ИКОМОС 28 ноября 2011 г. – <https://www.icomos.org/en/charters-and-texts>.
- [2] *Граудина, Л.К., Ширяев, Е.Н.* Культура русской речи. Учебник для вузов. — М.: Издательская группа НОРМА-ИНФРА, 1999. — 560 с.

- [3] **Сложеникина, Ю.В.** Основы терминологии: лингвистические аспекты теории термина / Ю.В. Сложеникина. – М.: Книжный дом ЛИБРОКОМ, 2013. – 120 с.
- [4] **Ожегов, С.И.** Толковый словарь русского языка / С.И. Ожегов, Н.Ю. Шведова. – М.: ООО «ИТИ Технологии», 2003. – 944 с.
- [5] Международная хартия по вопросам сохранения и реставрации памятников и достопримечательных мест (Венецианская хартия 1964 г.). – <http://www.icomos.org/venicecharter2004/russian.pdf>.
- [6] **Brandi C.** Teoria del restauro. Lezioni raccolte da Licia Vlad Borrelli, Joselita Raspi Terra, Giovanni Urbani. Roma. Edizioni di storia e letteratura, 1963. – 164.
- [7] **Бранди, Ч.** Теория реставрации и другие работы по теме охраны, консервации и реставрации. М.: Альфа-Дизайн, 2011. – 266 с.
- [8] Архитектурная реставрация, обратимость и совместимость: интервью с Джованни Карбонарой. – <http://www.architetto.info/news/protagonisti/restauro-architettonico-reversibilita-e-compatibilita-intervista-a-giovanni-carbonara/>.
- [9] **Мазаков, Г.П., Ратия, Ш.Е.** Методика реставрации памятников архитектуры. Пособие для архитекторов реставраторов. – М.: Госстройиздат, 1961. – 216 с.
- [10] **Загородников, С.В.** Экономика предприятия. 2009. – [https://bizbook.online/ekonomika\\_teorija/printsiip-sovmestimosti-110331.html](https://bizbook.online/ekonomika_teorija/printsiip-sovmestimosti-110331.html).
- [11] Большой бухгалтерский словарь / Под ред. А.Н.Азрилияна. – М.: Институт новой экономики, 1999. – 574 с.
- [12] Corriere della sera. Итальянский словарь. Понятие «совместимость» – [http://dizionari.corriere.it/dizionario\\_italiano/C/compatibilita.shtml](http://dizionari.corriere.it/dizionario_italiano/C/compatibilita.shtml).
- [13] Corriere della sera. Итальянский словарь. Понятие «совместимый» – [http://dizionari.corriere.it/dizionario\\_italiano/C/compatibile.shtml](http://dizionari.corriere.it/dizionario_italiano/C/compatibile.shtml).
- [14] Treccani. Итальянский словарь. – <http://www.treccani.it/vocabolario/compatibile/>.
- [15] **Ушеров-Маршак, А.В.** Бетонведение: лексикон. М.: РИФ Стройматериалы. – 2009. – 112 с.
- [16] **Рязанцев, В.Д.** Большая политехническая энциклопедия. – М.: Мир и образование. 2011. – 707 с.
- [17] ГОСТ Р ИСО/МЭК 19762-4-2011: Информационные технологии. Технологии автоматической идентификации и сбора данных (АИСД). Гармонизированный словарь. Часть 4. Общие термины в области радиосвязи. – <http://docs.cntd.ru/document/1200091436>.
- [18] Словарь логики. – <http://slovarslov.ru/slovar/log/s/sovmestimost.html>.
- [19] Психологическая энциклопедия. – <http://www.insai.ru/slovar/sovmestimost-4>.
- [20] СТО 75298253-009-2008: Швы монтажные оконные. Технические условия. ООО "Гарантия-Строй" - М., 2008. – <http://docs.cntd.ru/document/471813569>.
- [21] Universalium. Словарь английского языка. Понятие «sustainable» - <https://universalium.academic.ru/45095/sustainable>.
- [22] Оксфордский словарь. Понятие «compatible» - <https://en.oxforddictionaries.com/definition/compatible>.
- [23] СРП-2007. Рекомендации по проведению научно-исследовательских, изыскательских, проектных и производственных работ, направленных на сохранение объектов культурного наследия (памятников истории и культуры) народов Российской Федерации. – [www.northarch.ru/downloads/SRP-2007.doc](http://www.northarch.ru/downloads/SRP-2007.doc)
- [24] РНИП 1.02.01-94. Инструкция о составе, порядке разработки, согласовании и утверждении научно-проектной документации для реставрации недвижимых памятников истории и культуры. – [http://www.snip-info.ru/Rnip\\_1\\_02\\_01-94.htm](http://www.snip-info.ru/Rnip_1_02_01-94.htm).
- [25] ГОСТ Р 55528-2013. Состав и содержание научно-проектной документации по сохранению объектов культурного наследия. Памятники истории и культуры. Общие требования. – <http://docs.cntd.ru/document/1200104243>.
- [26] ГОСТ 21.501-2011. Система проектной документации для строительства. Правила выполнения рабочей документации архитектурных и конструктивных решений. – <http://www.normacs.info/discussions/687>.
- [27] **Крысько, В.** Словарь-справочник по социальной психологии. СПб.: Питер. 2003. – 416 с.
- [28] **Платонов, К.К.** Структура и развитие личности. – М.: Наука, 1986. – 256 с.

## ONTOLOGY OF THE TERM "PRINCIPLE OF COMPATIBILITY" IN RESTORATION DESIGNING

O.A. Egorova

Saint Petersburg State University of Architecture and Civil Engineering, Saint Petersburg, Russia  
olga\_2802@list.ru

### Abstract

The article raises the question of the relevance of the terminological base of modern restoration design. As an example, the terms "compatibility" and "compatibility principle" are considered, their conceptual spaces are designated, the essence of the application of these terms is disclosed. The ontological and gnoseological aspects of the concept of "compatibility" are designated as the fundamental principle of restoration design on the basis of an analysis of the basic terminological definitions, derived from dictionaries, textbooks and scientific literature. Based on the analysis of linguistic differences of the concept of "compatibility" in Russian, English and Italian, the essential differences in the use of this term are revealed, as well as in the implementation of the "compatibility principle" in the restoration of architectural monuments. The problem of understanding the term "compatibility", depending on the country and language, on which the term is interpreted, is considered. The article suggests a new definition of the concept of "compatibility", recommended for inclusion in the vocabulary of the restorer-designer.

**Key words:** restoration designing, compatibility, "principle of compatibility", conceptual apparatus, terminology.

**Citation:** Egorova OA. Ontology of the term "Principle of Compatibility" in restoration designing [In Russian]. Ontology of designing. 2018; 8(1): 124-133. DOI: 10.18287/2223-9537-2018-8-1-124-133.

### References

- [1] The Valletta Principles for the Safeguarding and Management of Historic Cities, Towns and Urban Areas – 2011. - <https://www.icomos.org/en/charters-and-texts/>.
- [2] **Graudina LK, Shiryayev EN.** Culture of Russian speech. Textbook for high schools. [In Russian]. – Moscow: NORMA-INFRA publ.; 1999; 560.
- [3] **Slozhenenikina YuV.** Fundamentals of terminology: linguistic aspects of the theory of the term - Moscow: Book house LIBROKOM, 2013. - 120 p.
- [4] **Ozhegov SI.** Dictionary of the Russian language [In Russian]. Moscow, ITI Technologies LLC, 2003.
- [5] International Charter for the Conservation and Restoration of Monuments and Sites [In Russian]. – 1964. - <http://www.icomos.org/venicecharter2004/russian.pdf>.
- [6] **Brandi C.** Restoration theory. Lezioni raccolte da Licia Vlad Borrelli, Joselita Raspi Terra, Giovanni Urbani. Edizioni di storia e letteratura; Roma; 1963; 164.
- [7] **Brandi C.** Restoration theory and other works on the preservation, conservation and restoration [In Russian]. – Moscow: Alpha-Design publ.; 2011; 266.
- [8] Architectural restoration, reversibility and compatibility: an interview with Giovanni Carbonara. [In Italian]. - <http://www.architetto.info/news/protagonisti/restauro-architettonico-reversibilita-e-compatibilita-intervista-a-giovanni-carbonara/>.
- [9] **Magakov GP, Ratia SE.** Method of restoration of architectural monuments [In Russian]. Handbook for architects-restorers. – Moscow; Gosstroyizdat publ.; 1961; 216.
- [10] **Zagorodnikov SV.** Economics of the enterprise [In Russian]. - [https://bizbook.online/ekonomika\\_teoriya/printsip-sovmestimosti-110331.html](https://bizbook.online/ekonomika_teoriya/printsip-sovmestimosti-110331.html).
- [11] A large accountant dictionary [In Russian]. Ed. Azriliyan AN. – Moscow: Institute for New Economics; 1999; 574.
- [12] Evening courier. Italian dictionary. Word «compatibilità». [In Italian]. - [http://dizionari.corriere.it/dizionario\\_italiano/C/compatibilita.shtml](http://dizionari.corriere.it/dizionario_italiano/C/compatibilita.shtml).
- [13] Evening courier. Italian dictionary. Word «compatibile». [In Italian]. - [http://dizionari.corriere.it/dizionario\\_italiano/C/compatibile.shtml](http://dizionari.corriere.it/dizionario_italiano/C/compatibile.shtml).
- [14] Treccani. Italian dictionary. - <http://www.treccani.it/vocabolario/compatibile/>.
- [15] **Usherov-Marshak AV.** Concrete studies: lexicon [In Russian]. – Moscow: RIF Stroimateriali publ.; 2009; 112.
- [16] **Ryazantsev VD.** Great polytechnical encyclopedia [In Russian]. – Moscow: Mir I Obrazovanie publ.; 2011; 707.

- [17] GOST 19762-4-2011: Information Technology. Automatic identification and data collection technologies. Harmonized dictionary. Part 4. General terms in the field of radio communication [In Russian]. - <http://docs.cntd.ru/document/1200091436>.
- [18] Dictionary of Logic [In Russian]. - <http://slovarslov.ru/slovar/log/s/sovместimost.html>.
- [19] The Psychological Encyclopedia [In Russian]. - <http://www.insai.ru/slovar/sovместimost-3>.
- [20] STO 75298253-009-2008: Window's assembly seams. Technical specifications [In Russian]. - <http://docs.cntd.ru/document/471813569>.
- [21] Universalium. English dictionary. Word «sustainable». [In English] - <https://universalium.academic.ru/45095/sustainable>.
- [22] Oxford dictionary. Word «compatible» - <https://en.oxforddictionaries.com/definition/compatible>.
- [23] Code of Restoration Rules – 2007. Recommendations for the conduct of research, survey, design and production work aimed at preserving the objects of cultural heritage (monuments of history and culture) of the peoples of the Russian Federation [In Russian]. – [www.northarch.ru/downloads/SRP-2007.doc](http://www.northarch.ru/downloads/SRP-2007.doc)
- [24] RNiP 1.02.01-94. Instruction on the content, procedure for the development, approval and maintenance of scientific and project documentation for the restoration of immovable historical and cultural monuments [In Russian]. - [http://www.snip-info.ru/Rnip\\_1\\_02\\_01-94.htm](http://www.snip-info.ru/Rnip_1_02_01-94.htm).
- [25] GOST R 55528-2013. Content of scientific and project documentation for the preservation of cultural heritage sites. Monuments of history and culture. General requirements [In Russian]. - <http://docs.cntd.ru/document/1200104243>.
- [26] GOST 21.501-2011. System of documents for construction. Rules for the implementation of working documentation of architectural and constructive solutions [In Russian]. - <http://www.normacs.info/discussions/687>.
- [27] *Krysko V.* Dictionary-reference book on social psychology [In Russian]. – St. Petersburg: Peter publ.; 2003; 416.
- [28] *Platonov K.K.* Structure and development of personality [In Russian]. – Moscow: Nauka publ.; 1986; 256.

## Сведения об авторе



*Егорова Ольга Александровна*, 1991 г. рождения. Окончила Санкт-Петербургский государственный архитектурно-строительный университет (СПбГАСУ), архитектурный факультет в 2015 году. Аспирантка кафедры архитектурно-строительных конструкций СПбГАСУ. Научные интересы: теория, история и методика реставрации, теория архитектуры, вопросы взаимодействия современного человека и исторической архитектурной среды.

*Egorova Olga Alexandrovna* (b. 1991). Graduated from the St. Petersburg State University of Architecture and Civil Engineering (SPSUACE), Faculty of Architecture in 2015. Post-graduate student of the department of architectural and building constructions of SPSUACE. Scientific interests: theory, history and methods of restoration, theory of architecture, interaction of modern man and the historical architecture environment.

УДК 004.82:004.912

## ПОДХОД К МОДЕЛИРОВАНИЮ ПРОЦЕССА ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ ТЕКСТА НА ОСНОВЕ ОНТОЛОГИИ

**Е.А. Сидорова**

*Институт систем информатики им. А.П. Ершова СО РАН, Новосибирск, Россия  
lsidorova@iis.nsk.su*

### **Аннотация**

В статье рассматриваются модели и методы представления знаний, ориентированные на задачи автоматической обработки текста и извлечения информации. В рамках предлагаемого подхода извлечение информации рассматривается как процесс пополнения онтологии информацией, представленной в виде объектов – экземпляров понятий предметной области. Для описания данного процесса предложены три базовые модели. Модель представления текста задаёт общую схему обработки текста и обеспечивает отображение полученной информации на текст. Модель представления знаний включает описание предметной лексики, жанровые модели текста и модели фактов, которые позволяют смоделировать процессы извлечения информации в терминах семантических классов предметной лексики и онтологии предметной области. Используемая атрибутивная модель представления данных обеспечивает сохранение информационных потоков данных, возникающих в процессе извлечения информации, и позволяет применять онтологические методы для решения задач снятия неоднозначности интерпретации текста и разрешение кореференции. Таким образом, предложена оригинальная методика, позволяющая пользователям проектировать систему анализа текста и моделировать процессы извлечения информации на основе онтологии предметной области.

**Ключевые слова:** *извлечение информации, модель текста, словарь предметной лексики, модель факта, пополнение онтологии.*

**Цитирование:** *Сидорова, Е.А. Подход к моделированию процесса извлечения информации из текста на основе онтологии / Е.А. Сидорова // Онтология проектирования. – 2018. – Т. 8, №1(27). – С.134-151. – DOI: 10.18287/2223-9537-2018-8-1-134-151.*

### **Введение**

Автоматическая обработка и анализ разнородной информации, представленной на естественном языке, является одним из самых востребованных на сегодняшний день направлений исследований. Отдельное положение в ряде решаемых задач в рамках данного направления занимает задача извлечения информации из предметно-специфического контента, представленного текстами, поскольку данная задача плохо разрешима статистическими методами и, как правило, требует привлечения знаний специалистов как о предметной области (ПрО), так и об особенностях языка.

Данная задача тесно связана с задачей пополнения онтологии [1]. Под пополнением онтологии понимается автоматический анализ различных источников и наполнение найденной информацией контента информационной системы, база данных которой опирается на онтологию ПрО. Извлекаемая информация в таких системах представляется экземплярами понятий и отношений заданной онтологии. Для решения данных задач, как правило, используют разнообразные знания в формализованном виде, такие как тезаурусы (WordNet, RusNet), толково-комбинаторные словари [2], аннотированные корпуса текстов (например, Национальный корпус русского языка - [www.ruscorgpora.ru](http://www.ruscorgpora.ru)) и т.п. Работа со знаниями в свою очередь требует создания технологий, которые автоматизируют процессы проектирования и разра-

ботки программных систем посредством предоставления пользователям (в том числе не программистам) средств моделирования, которые позволяют абстрагироваться от непосредственной разработки программных компонент и сконцентрироваться на вопросах обеспечения системы всеми необходимыми знаниями и моделирования непосредственно процессов извлечения информации в предметных терминах.

Вопросам моделирования процессов извлечения информации уделяется мало внимания, как правило, рассматриваются решения конкретных задач, что не позволяет выделить технологические и методологические аспекты решения подобных задач в целом. Целью данной работы является попытка восполнить этот пробел с помощью ряда модельных описаний процессов на различных уровнях представления.

Наиболее изученными являются онтологии и тезаурусы [3]. Онтология, как инструмент моделирования ПрО, является основой для формализации интересов пользователя, формирования словаря и представления конечного результата работы создаваемой системы, а также содержит необходимые знания для этапа семантического анализа текста [4]. Тезаурус, как инструмент описания предметной лексики, позволяет характеризовать термин и его связи с точки зрения особенностей употребления в данной ПрО [5, 6]. Для того, чтобы зафиксировать формально-лингвистические свойства, обусловленные языковой практикой описания объектов и ситуаций данной ПрО, необходимо использовать модели, которые бы с одной стороны описывали различные варианты представления в тексте одной и той же информации, с другой – моделировали бы процесс извлечения данной информации. К моделям данного класса можно отнести синтаксические или семантико-синтаксические модели управления [7, 8], лексико-синтаксические правила и шаблоны [9, 10], правила на основе онтологии [11] и т.п.

Указанные выше модели позволяют формировать в первую очередь базу знаний разрабатываемой системы, т.е. моделировать «заранее» заданные параметры системы или входные данные. Для моделирования целостного процесса извлечения информации их оказывается недостаточно, необходимо указать структуры данных для представления информации на всех промежуточных этапах обработки текста, а также обеспечить решения «специализированных» для данной проблемной области задач, связанных с неоднозначностью, присущей естественному языку.

В данной работе предложен подход к моделированию процессов извлечения информации из текста, который включает компоненты моделирования базы знаний системы, модели представления текста в процессе его обработки, а также способы описания информационных потоков данных, возникающих в процессе извлечения информации, которые позволяют применять онтологические методы для решения задач снятия неоднозначности и разрешение ко-референции.

## 1 Модель представления знаний

Особенностью развиваемого подхода к извлечению информации из текста является применение знаний о ПрО, преимущественное использование лексико-семантической информации и жанровых особенностей документов.

Рассматриваемая лингвистическая модель знаний включает три компонента. Словарь ПрО задаёт лексическую модель подязыка ПрО, жанровая модель текста формирует жанровую структуру рассматриваемого текстового источника, сужая область поиска определённой информации, и модели фактов, связывающие семантико-синтаксические модели, описывающие структуру выражений, принятых в данной области для описания информации, с формальным представлением этой информации, определяемым онтологией ПрО.

## 1.1 Словарь предметной лексики

В рамках предлагаемого подхода в качестве лексической модели языка рассматриваются информационно-поисковые словари ПрО. Такого рода словари ориентированы на автоматическую обработку текста и содержат дополнительную информацию, позволяющую распознавать термины в текстах.

Формально предметный словарь определяется системой вида  $\langle V, M, T, S \rangle$ , где

$V = W \cup P \cup L$  – множество предметных терминов, включающих:

- $W$  – множество лексем (каждой лексеме сопоставлена информация обо всей совокупности её форм);
- $P$  – множество словокомплексов или многословных терминов, характеризующихся высокой частотностью в анализируемом подязыке (словокомплекс описывается парой  $\langle L$ -грамма, тип структуры $\rangle$ , где  $L$ -грамма задаёт последовательность лексем, а тип структуры определяет вершину и правила согласования элементов  $L$ -граммы);
- $L$  – множество лексических конструкций, каждая из которых описывается с помощью шаблонов, используемых для распознавания регулярных текстовых фрагментов (лексические конструкции предназначены для распознавания таких структур как сокращения, аббревиатуры, численные или буквенно-численные обозначения объектов ПрО или значения их атрибутов).

$M$  – морфологическая модель языка, включающая описание морфологических классов и атрибутов (атрибуты в рамках каждого класса делятся на словообразующие, присущие всем формам лексемы данного класса, и словоизменяемые, различающие формы одной лексемы).

$T$  – множество тематических признаков, организованных в иерархию (каждому термину может быть сопоставлен набор признаков с указанием веса связи, где вес – это значение из интервала  $[0, 1]$ , отражает степень принадлежности термина признаку).

$S$  – лексико-семантическая модель ПрО.

Рассмотрим подробнее последний компонент, который является особенно важным при моделировании процесса извлечения информации.

Модель предметной лексики должна включать описание структуры семантики терминов и позволять, в конечном итоге, сопоставлять текстовым единицам их смысловые эквиваленты. Предложенная модель включает грамматическую, тезаурусную и семантическую информацию о термине, а также необходимые данные для описания валентностной структуры предикатных слов.

Для кодирования семантической информации о слове предусмотрены следующие возможности (см. рисунок 1).

*Семантический класс.* Термин может быть отнесён к определённому семантическому классу. Иерархия классов позволяет отнести термин к определённому уровню иерархии, более общему или конкретному с наследованием свойств общего класса.

*Семантический атрибут.* Для представления лексического значения термина используются семантические атрибуты. Совокупность значений атрибутов, приписанных слову, в определённой мере моделирует компонентную семантическую структуру слова. Основные компоненты семантической структуры термина могут рассматриваться как тезаурусные дескрипторы.

*Группировка семантических классов и атрибутов* для описания многозначного слова. Если термин имеет более одного формально различимого контекстом значения, формируется соответствующее число семантических статей слова, объединяющих в себе с помощью механизма группировки семантический класс и совокупность семантических атрибутов с их значениями.



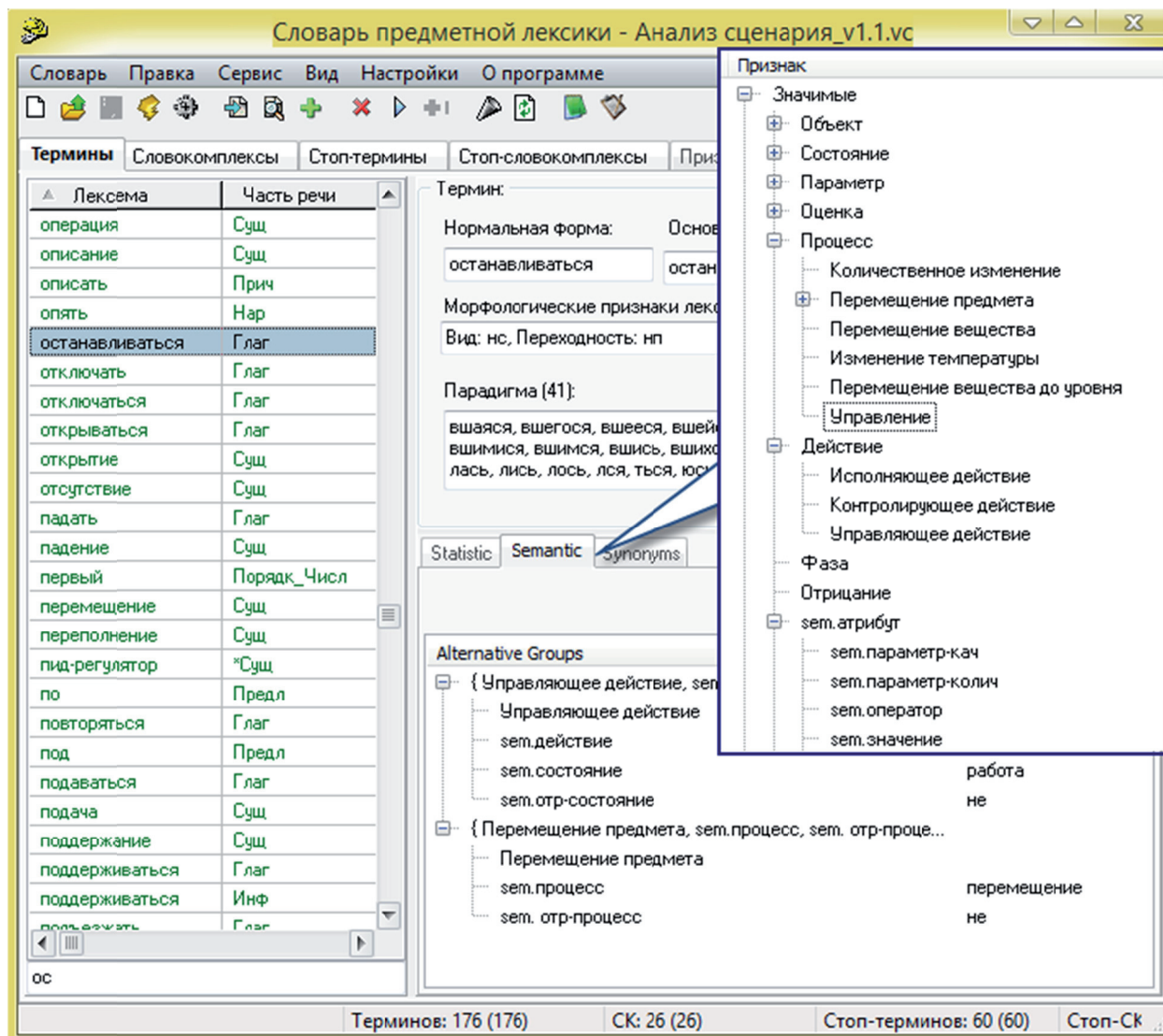


Рисунок 1 – Пример описания семантики термина

Семантика найденного в тексте термина обеспечивает формирование лексического объекта. Структуру лексического объекта можно представить следующим образом:

```

LexObject
Name: string;           // название термина
Descriptor: string;    // имя дескриптора
Semantic: set of LexClass; // множество лексико-семантических классов
Neg: bool;             // наличие отрицания
Attributes: set of Semantic_attribute; // множество атрибутов
Value: string;        // значение или форма, в которой встретился термин в тексте
Position: int;        // позиция термина в тексте
Grammatics: set of gram_parameters; // грамматические характеристики термина

```

Параметр *Name* задаёт нормализованную форму термина, полученную либо в соответствии с правилами морфологической нормализации (в случае однословного термина), либо в соответствии с правилами согласования слов в словосочетании (для многословных терминов), либо как имя лексического шаблона (для символьно-числовых конструкций). Параметр

*Descriptor* определяет имя понятия, к которому относится термин. В частности, дескриптор служит для формирования синсетов или групп синонимов, когда все термины с одинаковым дескриптором являются синонимами или квазисинонимами в данной ПрО. Поле *Semantic* определяет лексико-семантический класс(ы) термина. Наличие нескольких значений в поле *Semantic* лексического объекта соответствует составному значению термина (например, сложносоставное слово *пятиметровый* имеет лексическое значение *Числа* и *Единицы измерения*) в отличие от неоднозначности, когда формируется несколько лексических объектов *LexObject* (например, по термину *останавливаться* будут созданы объекты с семантикой *Действия* и *Процесса*). Отметим, что морфологическая омонимия не порождает отдельных лексических объектов, а разрешается, если необходимо, в рамках одного объекта. Поле *Neg* задаёт наличие отрицания при значении термина, которое в тексте может выражаться, например, приставкой *не*. Поле *Attributes* задаёт множество семантических атрибутов для представления структуры семантики термина. Поля *Grammatics*, *Position* и *Value* определяются словарными и текстовыми характеристиками термина.

## 1.2 Жанровая модель текста

Предлагаемые решения по анализу и извлечению информации из текстов опираются на понятие жанра как совокупности содержательных и формально-лингвистических (логико-композиционных и лексико-грамматических) аспектов.

Жанровая модель текста представляется системой вида  $GMT = \langle g, F_S, LC, V_G, M_G, P_G \rangle$ , где:

$g$  – жанровый тип документа;

$F_S$  – множество формальных жанровых сегментов (каждый сегмент, найденный в тексте в соответствии с жанровой моделью, определяется типом  $t_f \in F_S$  и начальной и конечной текстовой позицией);

$LC$  – логико-композиционная структура текста, определяемая множеством взаимосвязей между текстовыми фрагментами;

$V_G$  – словарь жанровой лексики;

$M_G$  – множество жанровых маркеров, которые задаются с помощью терминов из  $V_G$ ;

$P_G: M_G \rightarrow F_S$  – множество жанровых шаблонов, которые связывают маркеры и структурные блоки текста.

Таким образом, жанровые особенности текста передаются его разбиением на содержательные блоки, которые:

- включают определённую жанровую лексику,
- имеют определённую структурную организацию,
- реализуются в рамках определённых формальных сегментов.

Логико-композиционная структура текста выявляется с помощью лексикона жанровых маркеров и шаблонов, выделяющих содержательные блоки. Маркеры извлекаются из текстов, и прежде всего – из заголовков подразделов, вводных предложений и списков, включенных в состав документа. Простые маркеры сопоставляются терминам предметного словаря или группе терминов (синонимы). Более сложные формируются на основе простых: подерживаются альтернативы, совместная встречаемость, а также вложенное использование маркеров.

На рисунке 2 в качестве примера приведён фрагмент жанровой модели текста стандартного протокола клинических испытаний. Данные тексты включают жанровые сегменты, маркированные жанровыми тегами, на основе которых могут извлекаться фрагменты текста для поиска той или иной информации.

Выявленные на основе жанровых моделей значимые сегменты могут в дальнейшем использоваться в качестве условия поиска определённой информации.

<b>TextGenre</b> <b>Block</b> genre_segment <b>Block</b> genre_segment ... <b>Block</b> genre_segment <b>Block</b> genre_segment ...	<pre> &lt;study_design_info&gt;   &lt;allocation&gt;Randomized&lt;/allocation&gt;   &lt;intervention_model&gt;Parallel Assignment&lt;/intervention_model&gt;   &lt;primary_purpose&gt;Prevention&lt;/primary_purpose&gt;   &lt;masking&gt;Double Blind (Participant, Care Provider, Investigator,   Outcomes Assessor)&lt;/masking&gt; &lt;/study_design_info&gt; </pre>
--	--

Рисунок 2 – Фрагмент жанровой модели текста, описывающего дизайн клинического исследования

### 1.3 Модель факта

Модель факта является основой для моделирования процесса извлечения информации. Каждая модель задаёт схему извлечения единицы информации или единичного факта, представленного связанным языковым выражением. При извлечении факта из текста необходимо учитывать множество языковых способов репрезентации данного факта носителями подъязыка и обеспечивать их трансформацию в формальную структуру. Совокупность моделей должна описывать процесс целиком, начиная с инициализации объектов, заполнения атрибутов и установления онтологических отношений между объектами.

Факт, представляя собой зафиксированное в высказывании (языковом выражении) эмпирическое знание об объектах, их свойствах и ситуациях, может быть формализован в виде когнитивной схемы, соотносящей его с понятиями и отношениями онтологии. Таким образом, модель фактов формирует знания о согласовании имеющихся лингвистических знаний с предметными знаниями и фиксирует формально-лингвистические свойства, обусловленные языковой практикой описания объектов и ситуаций данной ПрО.

Предлагаемый способ моделирования обеспечивает преимущественное использование лексико-семантической информации, что не исключает применения частичного синтаксического анализа и синтаксических ограничений, накладываемых на семантический каркас моделей фактов. Известные системы отличаются полнотой и ролью синтаксического анализа в процессе извлечения фактической информации из текстов. Так, технология [12] предполагает построение полного семантико-синтаксического дерева предложения, к которому применяются шаблоны (своего рода фильтры), описывающие искомые факты. В предлагаемом подходе, как и в [13], синтаксический анализатор применяется локально (при обнаружении ключевых единиц и их конфигураций), в частности, предусмотрено определение актантных позиций предикатных слов [14].

Основой лингвистического описания факта является семантико-синтаксическая модель, которая ограничивает синтаксическую сочетаемость и согласованность грамматических и семантических признаков терминов (вершин синтаксических групп) в соответствии с правилами согласования и управления. Такие модели описываются в виде актантной структуры, связанной с одной или несколькими обобщёнными лексемами [9, 15]. Под обобщённой лексемой понимается либо термин словаря (или его форма), либо группа лексем, описанных в терминах грамматических и семантических категорий без указания нормальной формы. Актантная структура описывает набор актантов, характеризующих соответствующую валентность, в терминах семантических и грамматических характеристик, которые являются ограничениями для зависимых слов.

Семантико-синтаксическая модель характеризуется парой вида  $\langle lg, M_A \rangle$ , где:

$lg = \langle L_{lg}, S_{lg}, M_{lg} \rangle$  – обобщённая лексема, характеризующая группу терминов  $L_{lg} \subseteq V$ , обладающую набором семантических признаков  $S_{lg} \subseteq S$  и заданными значениями морфологических атрибутов  $M_{lg} \subseteq Mv$ ;

$M_A = \langle A_1, \dots, A_n \rangle$  – последовательность актантов, описывающих модель, где каждый актант  $A_i \subseteq S \times Mv \times V_{pre}$  задаёт множество допустимых сочетаний семантических  $S$  и грамматических  $Mv$  характеристик терминов, а также условия на наличие предлога  $V_{pre}$ .

Предложенная структура семантико-синтаксических моделей предоставляет широкие возможности моделирования языковых связей в тексте. Так, модель может не содержать синтаксических ограничений и представлять собой онтологические отношения или описываться без семантических характеристик и соответствовать чисто синтаксическим моделям управления. Обобщение лексем в моделях позволяет компактно определить многие языковые конструкции, варианты взаимосвязи слов в выражениях и словарные группы.

Модель фактов задаётся структурой, аналогичной актантной структуре, которая описывается либо в терминах классов онтологии, либо в терминах семантических признаков словаря и связывается с фрагментом онтологии. Дополнительно накладываются ограничения на онтологические признаки элементов структуры и их взаиморасположение в тексте.

Формально модель фактов – это система вида  $F = \langle M_{AF}, M_{SS}, Cs, Cp, O_{Res} \rangle$ , где:

$M_{AF}$  – последовательность аргументов факта;

$M_{SS}$  – множество семантико-синтаксических моделей, которым должны удовлетворять аргументы на уровне лексического состава;

$Cs$  – множество семантических ограничений;

$Cp$  – множество структурных ограничений;

$O_{Res}$  – фрагмент онтологии, в соответствии с которым формируется результат применения модели фактов – новые объекты и/или изменения атрибутов уже существующих объектов, которые являются экземплярами понятий онтологии (относящихся к заданному фрагменту).

Выделяются два основных типа моделей фактов: модели, служащие для начальной инициализации объектов, и модели для выявления связей. Модели первой группы необходимы для начального формирования онтологических сущностей на основании словарных признаков. Модели второй группы моделируют процессы «обнаружения» фрагментов онтологии.

Рассмотрим эти два типа на примерах.

Система семантических признаков словаря формируется на основе онтологических сущностей, что позволяет инициализировать начальное формирование объектов непосредственно на основании словарных признаков. Рассмотрим пример описания модели для инициализации объектов класса *Препарат*, которые могут быть представлены в тексте аппозитивной именной группой. В этой группе опорным словом является родовое слово или словокомплекс (тип), а имя примыкает к нему в постпозиции.

**Scheme** Препарат3 : segment *Клауза*

arg1: Term::Препарат(SemClass: тип)

arg2: Term:: Препарат(SemClass: имя)

**Condition** PrePos(arg1,arg2), Contact(arg1,arg2)

⇒ Object :: Препарат(Тип: arg1.Class & arg2.Class, Наименование: arg2.Norm)

В данной модели термины должны иметь семантический класс *Препарат*, с учётом иерархии наследования признаков в словаре, а также первый термин должен обладать семантическим признаком *тип*, а второй – *имя*. При применении данной модели, например, к фразе:

*вакцина* <Препарат, SemClass: тип> "АСАМ2000" <Препарат, SemClass: имя> <sup>1</sup>

создаётся объект – экземпляр понятия онтологии *Препарат*, тип препарата (например, фармакологическая группа) может уточниться в соответствии с семантическим признаком первого или второго термина, атрибут *Наименование* у объекта заполняется наименованием второго термина *Norm*.

В качестве другого примера инициализирующей модели рассмотрим случай, когда на основе лексического шаблона (LexTerm) выделяется фрагмент текста в кавычках и формируется гипотеза о том, что это имя объекта, но уточнение его класса возможно только при наличии термина-классификатора:

**Scheme** Новый\_объект : segment *Клауза*  
 arg1: Term:: (SemClass: тип)  
 arg2: LexTerm::Именованный\_объект()  
**Condition** PrePos(arg1,arg2), Contact(arg1,arg2)  
 ⇨ Object (Тип: arg1.Class, Наименование: arg2.Name)

Формирование объекта с помощью данной модели осуществляется аналогично предыдущей.

При поиске и выявлении характеристик объектов и их связей, как правило, требуется проверить сочетаемость семантических и/или грамматических признаков объектов.

Рассмотрим примеры модели, используемой для извлечения атрибутов объектов.

**Scheme** ТипКогорты: genre\_segment <*arm\_group*>  
 arg1: Object::Группа(), genre\_segment <*arm\_group\_label*>  
 arg2: Term::Параметр(), genre\_segment <*arm\_group\_type*>  
 ⇨ arg1: Группа (тип: arg2.Name)

Данная модель позволяет уточнить тип группы на основе информации о жанровой структуре документа, которая в соответствии с принятым стандартом содержится строго в определённых жанровых фрагментах.

Рассмотрим пример модели для построения отношения в соответствии с рассматриваемой ситуацией.

**Scheme** УсловияПримененияПрепарата: genre\_segment <*group*>  
 arg1: Object::Группа()  
 arg2: Object::Препарат()  
 arg3: Term::Параметр(SemClass: кратность)  
 arg4: LexTerm::Доза()  
 arg5\*: Term::Параметр(SemClass: время)  
**Condition** genre\_segment (arg2, arg3, arg4, arg5) <*description*>, Contact(arg2, arg3), Contact\_weak(arg2, arg4), Contact\_weak(arg2, arg5)  
 ⇨ имеетНазначение (группа: arg1, препарат: arg2, тип: «препарат», размер дозы: arg4.value, кратность: arg3.value, время: arg5.value)

Условия назначения препарата для конкретной группы людей описываются такими характеристиками, как наименование препарата, его дозировка, кратность применения и время приёма. Данная информация в соответствии с принятым стандартом содержится строго в определённых жанровых фрагментах, однако в рамках фрагмента *description* описание пара-

<sup>1</sup> В примерах в скобках указываются признаки терминов, заданные в словаре.

метров разворачивается в виде текста, что требует применения более сложного лингвистического анализа. В результате применения модели создается описание ситуации терапевтического вмешательства.

Приведённый набор моделей фактов демонстрирует подход к извлечению информации о проводимом клиническом исследовании на основе структуры протокола.

## 2 Модель представления текста

Важным компонентом моделирования процесса извлечения информации является модель представления текста, которая последовательно изменяется, обогащаясь на каждом этапе анализа новыми знаниями. Для описания происходящих изменений предложена модель, близкая по смыслу к схемам, используемым при создании размеченных корпусов текстов [16]. Отличия заключаются в:

- единообразной поддержке всех этапов обработки текста, включая семантический;
- использовании «внешнего» аннотирования (а не систему тэгов), синхронизированного с текстом [17];
- ориентации на объектно-ориентированное представление данных.

Модель представлена набором *покрытий текста*, когда промежуточные результаты обработки представляются однотипными объектами с заданной проекцией на текст (текстовыми интервалами), что позволяет наглядно интерпретировать полученные результаты и выделять контекстно связанные с каждым элементом знания.

Модель текста определяется пятеркой  $\langle C_A, C_L, C_G, C_{Th}, C_{IO} \rangle$ , где:

$C_A = \{atom_1, \dots, atom_n\}$  – графематическое покрытие, содержащее множество атомов  $atom_i = \langle t, id, pos \rangle$ , где атом – это объект, сопоставляемый неразрывному фрагменту текста, состоящему из символов одного типа  $t$ ; атомы упорядочены по встречаемости в тексте и определяются порядковым номером  $id$  и текстовыми позициями  $pos = \langle pos\_begin, pos\_end \rangle$ ;

$C_L = \{lex_1, \dots, lex_n\}$  – терминологическое покрытие, содержащее множество лексических объектов вида  $lex_i = \langle v, m_v, s_v, pos \rangle$ , где  $v \in V$  – термин словаря;  $m_v$  – множество грамматических характеристик термина  $v$ ;  $s_v$  – множество семантических признаков  $v$ ;  $pos$  – текстовая позиция;

$C_G = \{s_1, \dots, s_n\}$  – сегментное (или жанровое) покрытие, отражающее логико-композиционную структуру текста и включающее множество сегментов вида  $s_i = \langle t_f, pos, R_G \rangle$ , где  $t_f$  – тип или формальный сегмент жанровой модели текста,  $pos$  – текстовая позиция и  $R_G$  – связи с другими сегментами, определяющими их взаиморасположение в тексте;

$C_{Th} = \{st_1, \dots, st_n\}$  – тематическое покрытие, которое определяется множеством тематических фрагментов вида  $st_i = \langle th, pos, R_{Th} \rangle$ , где  $th$  – тематика из заранее определённого тематического классификатора,  $pos$  – текстовая позиция и  $R_{Th} \subseteq C_G$  – подмножество структурных сегментов, покрываемых данным фрагментом;

$C_{IO} = \{IO_1, \dots, IO_n\}$  – информационное покрытие, содержащее множество найденных в тексте информационных объектов вида  $IO_i = \langle I, Pos, R_I \rangle$ , где  $I$  – онтологический объект или экземпляр понятия, определённого онтологией ПрО,  $Pos$  – текстовая позиция (в общем случае разрывная, т.е. определяемая множеством неразрывных позиций  $pos$ ),  $R_I$  – множество информационных зависимостей объекта, полученных в процессе обработки текста при использовании информации из одного объекта для генерации или обновления другого.

В зависимости от решаемой задачи могут быть выделены и другие типы покрытий. Приведённая модель ориентирована в первую очередь на задачи семантического анализа и извлечения информации. На рисунке 3 представлена общая схема преобразования входящих текстовых данных и промежуточные результаты в виде покрытий.

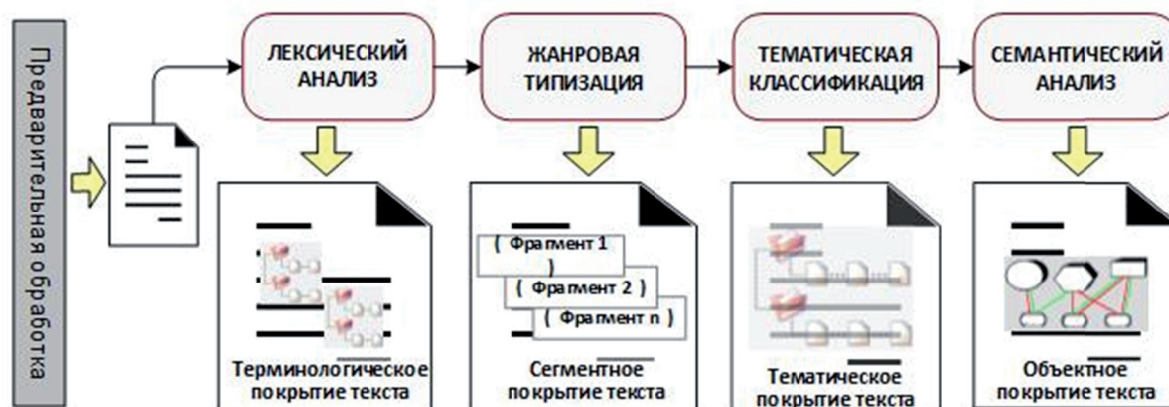


Рисунок 3 – Этапы формирования модели представления текста

*Графематическое покрытие* текста является результатом его графематического анализа, в процессе которого входящий линейный текст разбивается на элементарные атомы. Основная задача данного этапа – сгруппировать символы одного типа в последовательности и дать им необходимую интерпретацию: слово определённого алфавита, число, знак препинания и т.п. Для решателей, работающих с разметкой (например, html-тексты), можно дополнительно задать типизацию тэгов или помет. Важным свойством данного представления является то, что элементы покрытия задают все возможные границы элементов для всех последующих представлений, т.е. при дальнейшей обработке ни один атом не может быть «разделён».

*Терминологическое покрытие* состоит из словарных терминов, найденных в данном тексте, с учетом возможной омонимии и пересечений многословных терминов. Терминологическое покрытие текста – это лексическая модель текста, которая строится на основе лексической модели подязыка ПрО, и включает найденные в тексте термины с привязкой к позиции в тексте. После того, как термин найден в тексте, формируется лексический объект, который снабжается набором семантических атрибутов, заданных в предметном словаре для найденного термина (см. раздел 1.1).

*Сегментное покрытие* отражает структурное деление текста на логические (абзац, предложение, заголовок и т.п.) и жанровые фрагменты. Жанровое покрытие является результатом сегментации текста и одним из способов отражения его формальной структуры. Сегментация рассматривается на макроуровне, т.е. на уровне всего текста (в отличие от локального анализа предложения и выделения совокупности взаимосвязанных фрагментов (клауз), рассматриваемых в рамках синтаксического анализа предложения) и опирается как на формально-текстовые, так и на жанровые особенности документа (см. раздел 1.2), которые передаются разбиением текста на концептуальные части. При анализе текста разбиение на жанровые фрагменты помогает сузить область поиска информации определённого вида и, тем самым, повысить качество анализа. С помощью жанровой структуры текста могут решаться задачи определения жанровой релевантности документов, полученных из неизвестных источников, например, при поиске в Интернет [18].

*Тематическое покрытие* определяет текстовые границы тематически-связанных областей текста для каждой рассматриваемой тематики. Формирование таких областей осуществляется на основе словаря, в котором задано соответствие между терминами и тематическими признаками. Тематическое покрытие строится над терминологическим и жанровым покры-

тием. Здесь элемент тематического покрытия определяется как фрагмент текста, включающего кластер терминов, относящихся к одной теме, в границах формального сегмента (или последовательности сегментов) жанрового покрытия. Аналогично сегментам, тематические сегменты могут сужать область поиска информации определённого вида.

*Информационное (объектное) покрытие* описывает найденную информацию в виде семантической сети объектов ПрО. Информационное покрытие текста является самым информационно-насыщенным и представляет результаты семантической обработки документа. Чтобы построить информационное покрытие необходимо представить содержание документа в понятной компьютеру форме. Условием этого является наличие формата данных, задающего строгую структуру представления и хранения полученной информации. Данная структура должна быть «заранее» осмыслена, т.е. иметь заранее заданную семантическую интерпретацию. В современных подходах для этого используют онтологии ПрО [19].

В упрощённом виде онтологию можно представить как  $O = \langle C, T, A \rangle$ , где:

$C$  – множество классов, описывающих понятия ПрО;

$T = \cup T_i$  – множество типов данных и  $V_i$  – множество значений типа  $T_i$ ;

$A$  – множество атрибутов,  $F_A: C \rightarrow 2^{A \times (T \cup C)}$  – функция, которая определяет имена и типы атрибутов классов  $C$ ; выделяются атрибуты простых типов (owl:DataProperty) и связи (owl:ObjectProperty), также определяется подмножество ключевых атрибутов  $A_K \subseteq A$ , которые служат для однозначной идентификации объектов.

Контент информационной системы, построенной на основе онтологии  $O$ , представляется множеством экземпляров классов онтологии и описывается как  $I = \{I_1, \dots, I_n\}$ , где  $I_i$  класса  $C_i$  представляется набором атрибутов со значениями  $Av_i = \{(a_1, v_1), \dots, (a_k, v_k)\}$ .

Объектное покрытие описывает информационный контент текста в терминах информационных объектов, которые формируются на основе онтологических понятий и должны быть сопоставлены определённому экземпляру онтологии, который либо уже присутствует в контенте информационной системы, либо будет добавлен в результате анализа текста. Таким образом, информационные объекты являются своего рода «вхождениями» или упоминаниями онтологических объектов, обнаруженными в тексте.

Информационные объекты формируются на основе моделей фактов (см. раздел 1.3.), при этом порождаются информационные зависимости между объектами, выступающими в качестве аргументов модели, и её результатом. Для точного описания данных зависимостей используется атрибутивная модель извлечения информации.

Преимущество предложенной модели текста заключается: во-первых, в наглядном представлении результатов работы анализатора; во-вторых, предложенное описание может являться основой для формального описания алгоритмов и доказательства их свойств, а также служить в качестве абстракции верхнего уровня для программной реализации; в-третьих, использование данного представления в рамках информационных систем обеспечит достоверность результата, подтверждаемого непосредственно текстовым источником, что позволит проводить широкий спектр корпусных исследований.

### 3 Атрибутивная модель извлечения информации

При описании последовательных процессов извлечения информации с помощью моделей фактов интенсивно используются атрибуты объектов и связи между ними. Поэтому для детального представления происходящих процессов и отражения возникающих информационных зависимостей не только на уровне объектов, но и между их атрибутами, решено воспользоваться подходом к моделированию процессов на основе атрибутной конвейерной модели (АКМ).



Атрибутная модель – это логическая модель данных, предназначенная для отображения связей между свойствами или атрибутами объектов, участвующих в одном процессе. Для корректного моделирования процессов извлечения информации модели должны обладать свойством конвейерности, т.е. позволять описывать процессы, внутри которых не возникают циклы. Такого рода модели широко используются в различных ПрО, например, при описании технологических процессов с помощью диаграмм сборки Сервис-Компонентной Архитектуры SCA [20]. Использование АКМ позволяет выстраивать общую схему процесса с зависимостями по атрибутам.

В общем виде модель АКМ можно описать системой вида  $\langle O_M, A, L \rangle$ , где:

$O_M$  – множество объектов моделирования;

$A$  – множество атрибутов объектов (подразумеваются атрибуты простых типов данных);

$L$  – множество направленных связей между атрибутами объектов.

При моделировании процесса извлечения информации связь между атрибутами интерпретируется как передача значения от одного атрибута к другому. Для обеспечения целостности модели на связи накладываются следующие ограничения:

- атрибут может иметь только одну входящую связь и множество исходящих связей;
- связь можно установить только между атрибутами одного типа.

Для поддержки моделирования процессов извлечения информации на основе онтологии и моделей фактов было необходимо расширить АКМ таким образом, чтобы она поддерживала следующие возможности:

- в качестве атрибута объекта может выступать другой объект, и, соответственно, модель должна предоставлять возможность устанавливать связь между объектом и атрибутом другого объекта (соответствующего типа);
- возможность устанавливать значение атрибута напрямую без использования связей, т.е. без передачи значения от другого объекта;
- в случае порождения нового объекта (в соответствии с моделями фактов) возможность устанавливать связи между аргументами факта и порождаемым объектом.

Для реализации первой возможности была расширена типизация свойств АКМ – добавлена возможность использовать класс онтологии в качестве типа данных атрибута (аналогично онтологическому объекту), также в модель введены соответствующие связи между объектом и атрибутом, которые обладают теми же возможностями и ограничениями, что и исходящие связи. Для поддержки второй возможности были добавлены *независимые атрибуты* – атрибуты, значения которых можно изменять напрямую, не используя соединения (при этом возникают неявные зависимости между аргументами факта и независимым атрибутом, которые пока не моделируются). И, наконец, в случае если других связей не установлено, между объектом-аргументом и результирующим объектом устанавливается *объектная связь*, которая реализуется на уровне имён классов, т.е. имя класса объекта в данной модели рассматривается как его атрибут.

Предложенная модель имеет два практических применения. В первом случае АКМ является основой (метамоделью) для моделирования процессов с помощью моделей фактов: как для описания единичных моделей (см. рисунок 4), так и для их совокупности. В другом случае в процессе обработки текста она используется для сохранения информационных зависимостей между объектами информационного покрытия текста (данные информационные зависимости согласуются с зависимостями, заданными в моделях фактов) и, тем самым, представляет «историю» создания объектов. Эта информация позволяет, в частности, оценить степень связности объекта с контекстом и осуществить корректное удаление объекта из системы.

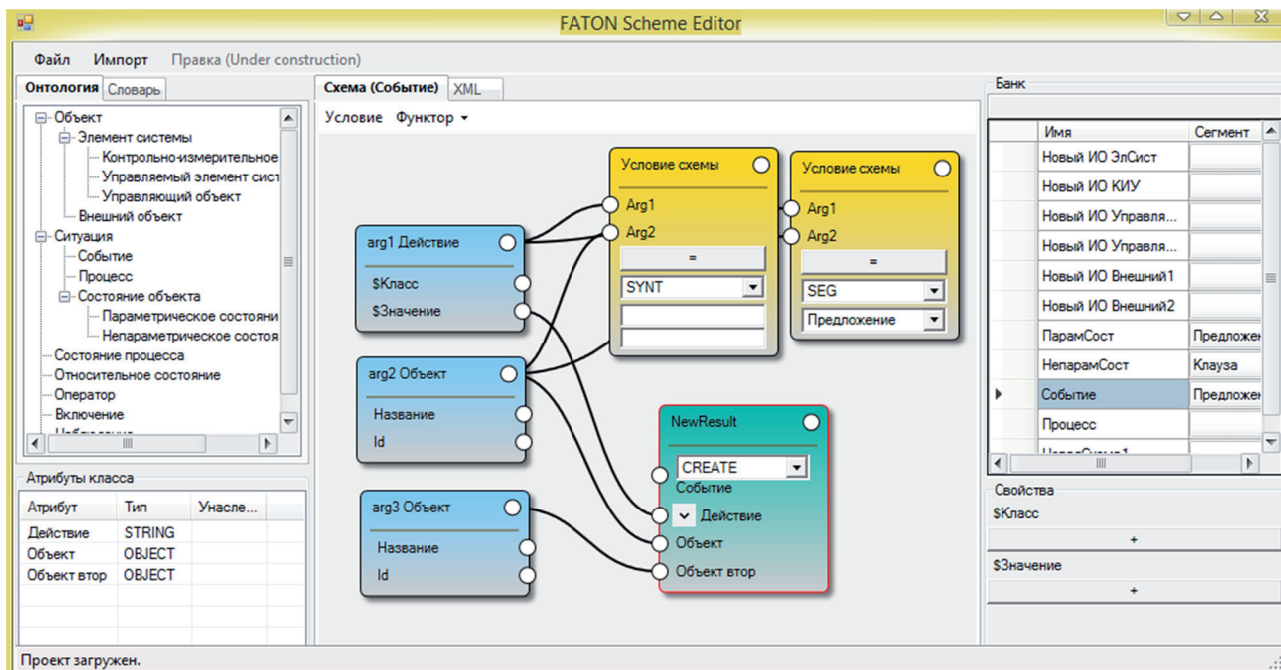


Рисунок 4 – Пример использования АКМ для моделирования процесса извлечения информации о единичном факте

#### 4 Извлечение информации

Задача извлечения информации в соответствии с предложенными моделями представления текста сформулирована в общем виде следующим образом.

Для заданной четверки  $\langle O, I, LM, T \rangle$ , где:  $O$  – онтология ПрО,  $I$  – информационный контент системы,  $LM$  – модель подязыка ПрО,  $T$  – текстовый источник, построить информационное покрытие текста  $C_{IO}$ , обладающее свойством однозначности и сопоставимости с  $I$ . Под однозначностью понимается наличие не более одной интерпретации любого фрагмента заданного текста  $T$ . Это означает, что при наличии нескольких вариантов разбора текста, например, в случае омонимии, должен быть выбран один вариант. Свойство сопоставимости с контентом системы означает, что каждому информационному объекту, входящему в результирующее информационное покрытие  $C_{IO}$ , должен быть однозначно сопоставлен экземпляр онтологии  $O$ , что обеспечивается наличием заданных ключевых атрибутов. Сопоставляемый экземпляр, по сути, является онтологическим референтом найденного в тексте объекта.

На рисунке 5 представлена общая схема реализации семантического этапа анализа и извлечения информации из текста. На вход системы основного анализа поступают результаты предварительного этапа обработки текста в виде терминологического, сегментного и тематического покрытия текста. База знаний системы включает знания об онтологии ПрО и моделях фактов, заданных для данной онтологии. Результатом работы системы является объектное покрытие текста, на основе которого формируется результирующее множество онтологических объектов, описывающих контент документа в терминах онтологии ПрО.

В процессе основного анализа можно выделить следующие три основных задачи:

- 1) Непосредственно извлечение фактов на основе моделей фактов.
- 2) Разрешение кореференции и поиск референтов объектов в онтологическом контенте системы.
- 3) Разрешение неоднозначности.

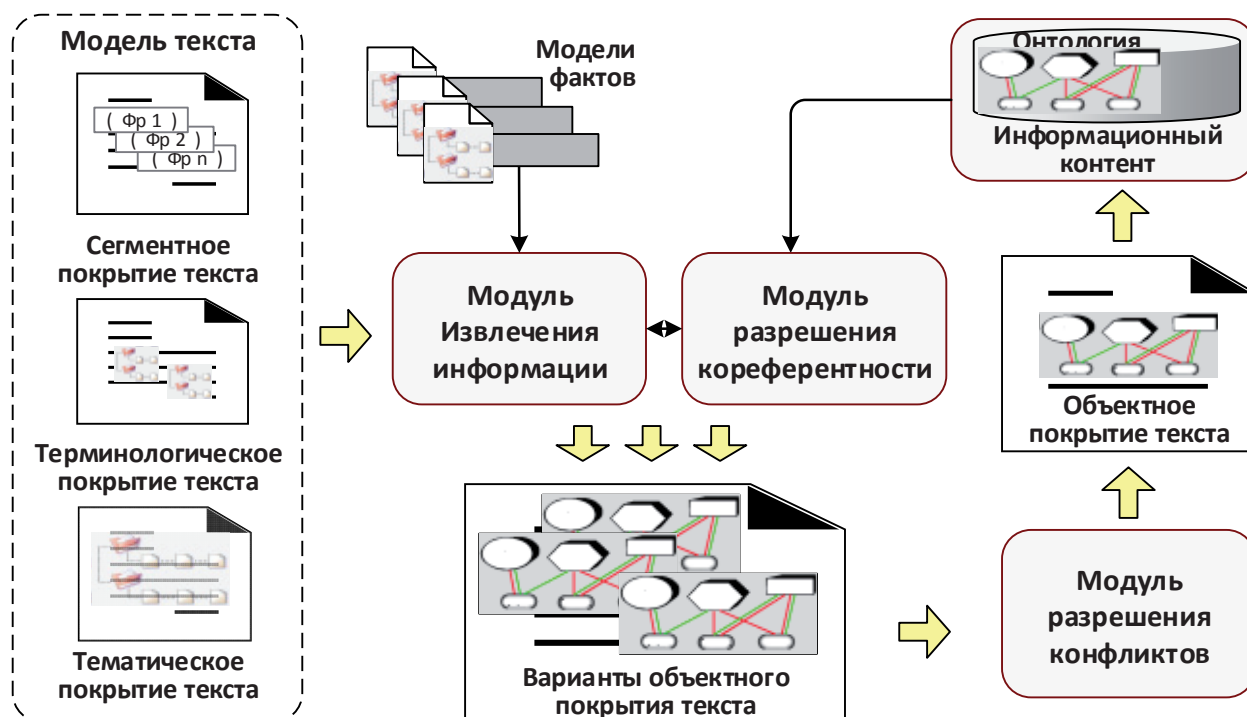


Рисунок 5 – Схема извлечения информации из текста

Решение первой задачи моделируется пользователем при разработке моделей фактов. Технологическая поддержка данного процесса осуществляется системой, которая по моделям фактов формирует правила и применяет их к входным данным (одна из возможных реализаций данной системы представлена в [21]). Качество извлечения информации в большой степени зависит от качества модели процесса, созданной пользователем. В результате формируется АКМ процесса извлечения информации для данного текста.

От решения второй задачи во многом зависит полнота анализа. В работе [22] предложен метод разрешения кореферентности на основе онтологии, в рамках которого осуществляется семантическое сравнение найденных в текстах объектов с учётом их онтологических свойств. На данном этапе ищутся все упоминания каждого онтологического объекта, формируются кореферентные группы информационных объектов и разрешаются возможные конфликты. На качество решения данной задачи влияют степень проработанности заданной пользователем онтологии и наполненность информационного контента системы.

Решение третьей задачи обеспечивает формирование конечного варианта объектного покрытия текста для размещения в информационном контенте системы. Неоднозначности при анализе текста возникают вследствие особенностей естественного языка. Языковая неоднозначность — это способность слова или выражения иметь различные интерпретации, в результате которой порождаются различные варианты разбора текста, конфликтующие между собой. В разрабатываемом подходе конфликты рассматриваются на уровне проекции результатов анализа текста на ПрО, т.е. в контексте заданной онтологии. Полученная при извлечении фактов АКМ характеризует информационные зависимости, возникшие в процессе формирования данной модели. Предлагаемый метод основан на идее вычисления степени информационной связанности информационных объектов, извлечённых из заданного текста [23]. Модуль разрешения конфликтов должен разрешить все неоднозначности таким обра-

зом, чтобы система была свободной от конфликтов и при этом сохранила максимально возможное количество объектов и связей.

Результат работы системы анализа помещается в информационное хранилище, предварительно обеспечив однозначную идентификацию найденных объектов методом, предложенным в работе [24].

### **Заключение**

Представленный в работе подход к моделированию процессов извлечения информации существенным образом опирается на знания о ПрО, явно формализованные в виде онтологии, что позволяет применять методы локального семантического и синтаксического анализа, не требуя наличия полного корректного синтаксического разбора и грамматически правильно построенного текста. Сужение области значения предметных терминов значительно уменьшает неоднозначность текста. Использование информационного контента онтологии при идентификации и сравнении объектов, найденных в тексте, позволяет использовать неявные знания, т.е. информацию, не содержащуюся в тексте.

Рассмотренные модели и их технологическое обеспечение предоставляют конечным пользователям – экспертам в ПрО, лингвистам и инженерам знаний – инструменты для моделирования процессов извлечения информации и их отладки, а разработчикам информационных систем – инструменты для проектирования систем автоматической обработки текста и концептуальные схемы представления данных.

### **Благодарности**

Работа выполнена при финансовой поддержке Президиума СО РАН (Блок 36.1. Комплексной программы ФНИ СО РАН II.1) и РФФИ (грант № 17-07-01600).

### **Список источников**

- [1] *Petasis, G.* Ontology Population and Enrichment: State of the Art / G. Petasis, V. Karkaletsis, G. Paliouras, A. Krithara, E. Zavitsanos // In Knowledge-driven multimedia information extraction and ontology evolution. - LNAI 6050. - Springer-Verlag Berlin, 2011. - P.134–166.
- [2] *Мельчук, И.А.* Опыт теории лингвистических моделей: «Смысл-Текст». Семантика, синтаксис / И.А. Мельчук. – М.: Школа «Языки русской культуры», 1999. – 992 с.
- [3] *Нариньяни, А.С.* ТЕОН-2: от Тезауруса к Онтологии и обратно / А.С. Нариньяни // Труды международного семинара Диалог'2002 по компьютерной лингвистике и ее приложениям. –М.: Наука, 2002. – Т.1. – С.307–313.
- [4] *Загоруйко, Ю.А.* Семантическая технология разработки интеллектуальных систем, ориентированная на экспертов предметной области / Ю.А. Загоруйко // Онтология проектирования. - 2015. - Т.5. - №1 (15). – С.30-46.
- [5] *Добров, Б.В.* Онтологии и тезаурусы: модели, инструменты, приложения: учебное пособие / Б. В. Добров, В. В. Иванов, Н.В. Лукашевич, В.Д. Соловьев. - М.: Интернет-университет информационных технологий; БИНОМ. Лаборатория знаний, 2009. - 173 с.
- [6] *Рубашкин, В.Ш.* Семантический компонент в системах понимания текста / В.Ш. Рубашкин // Труды Десятой национальной конференции по искусственному интеллекту с международным участием КИИ-2006. – М.: Физматлит, 2006. – Т.2. – С.455-463.
- [7] *Розенталь, Д.Э.* Управление в русском языке / Д.Э. Розенталь. – М.: Книга, 1986. – 173 с.
- [8] *Журавлев, А.О.* Создание базы данных моделей управления слов русского языка / А.О. Журавлев // Искусственный интеллект. 2013. № 1. – С.91-97.
- [9] *Большакова Е.И.* Язык лексико-синтаксических шаблонов LSPL: опыт использования и пути развития / Е.И. Большакова // Программные системы и инструменты: Тематический сборник. - №15. – М.: МАКС Пресс, 2014. – [http://www.lspl.ru/articles/Paper\\_19\\_LSPL.pdf](http://www.lspl.ru/articles/Paper_19_LSPL.pdf).

- 
- [10] **Ковалев, А.И.** Инструмент разработки предметных словарей на основе лексических шаблонов DigLex / А.И. Ковалев, Е.А. Сидорова // Материалы Всероссийской конференции с международным участием «Знания – Онтологии – Теории» (ЗОНТ–2015), 6 - 8 октября 2015 г. – Новосибирск: Институт математики им. С.Л. Соболева СО РАН, 2015. – Т1. – С.123-130.
- [11] **Хорошевский, В.Ф.** OntosMiner: Семейство систем извлечения информации из мультязычных коллекций документов / В.Ф. Хорошевский // Девятая Национальная конференция по искусственному интеллекту с международным участием КИИ-2004: Труды конференции. В 3-х т. М.: Физматлит, 2004, т. 2. - С.573-581.
- [12] **Ермаков, А.Е.** Автоматическое извлечение фактов из текстов досье: опыт установления анафорических связей / А.Е. Ермаков // Труды международной конференции Диалог' 2007 «Компьютерная лингвистика и интеллектуальные технологии». - М.: Наука, 2007. – С.131–135.
- [13] **Гершензон, Л.М.** Система извлечения и поиска структурированной информации из больших текстовых массивов СМИ. Архитектурные и лингвистические особенности / Л.М. Гершензон, И.М. Ножов, Д.В. Панкратов // Труды международной конференции Диалог'2005 «Компьютерная лингвистика и интеллектуальные технологии». М.: Наука, 2005. - С.97-101.
- [14] **Азарова, И.В.** Использование маркеров актантных позиций при анализе деловых текстов для расширения логической схемы ПрО / И.В. Азарова, А.С. Гребеньков, Т.М. Ландо // Труды международной конференции Диалог' 2008 «Компьютерная лингвистика и интеллектуальные технологии». - М.: РГГУ, 2008. - Вып. 7 (14). - С.11-16.
- [15] **Яковчук, Е.И.** Обобщенные семантико-синтаксические модели в задачах обработки текста / Е.И. Яковчук, Е.А. Сидорова // Труды рабочего семинара «Наукоемкое программное обеспечение НПО-2011». Ершовская конференция по информатике. – Новосибирск: ИСИ СО РАН, 2011. – С.287-292.
- [16] **Апресян, Ю.Д.** Синтаксически и семантически аннотированный корпус русского языка: современное состояние и перспективы / Ю.Д. Апресян, И.М. Богуславский, Б.Л. Иомдин, Л.Л. Иомдин, А.В. Санников, В.З. Санников, В.Г. Сизов, Л.Л. Цинман // Национальный корпус русского языка: 2003-2005. – М.: Индрик, 2005. – С.193-214.
- [17] **Blanco, X.** Using NooJ for Multipurpose Analysis of Romance Languages Corpora / X. Blanco // Труды международной конференции «Корпусная лингвистика–2008». – СПб., 2008. – С.40–44.
- [18] **Конonenko, И.С.** Жанровые аспекты классификации веб-сайтов / И.С. Кононенко, Е.А. Сидорова // Программная инженерия № 8. – 2015. – С.32–40.
- [19] **Загоруйко, Ю.А.** Автоматизация сбора онтологической информации об интернет-ресурсах для портала научных знаний / Ю.А. Загоруйко // Известия Томского политехнического университета, 2008. - Т.312. - №5. - С. 114-119.
- [20] **Marino, J.** Understanding SCA (Service Component Architecture) / J. Marino, M. Rowley. – Addison-Wesley Professional. – 2009. – 360 p.
- [21] **Garanina, N.** A multi-agent text analysis based on ontology of subject domain / N. Garanina, E. Bodin, E. Sidorova // Proc. of 9th Ershov Informatics Conference (PSI'2014). – Novosibirsk: A.P. Ershov Institute of Informatics systems, 2014. – P.50-56.
- [22] **Garanina, N.** A Distributed Approach to Coreference Resolution in Multiagent Text Analysis for Ontology Population / N. Garanina, E. Sidorova, I. Kononenko // Springer International Publishing AG 2018 A. K. Petrenko and A. Voronkov (Eds.): PSI 2017, LNCS 10742, 2018. - P.1–16.
- [23] **Garanina, N.O.** Conflict resolution in multi-agent systems with typed relations for ontology population / N.O. Garanina, E.A. Sidorova, I.S. Anureev // Programming and Computer Software. – 2016. – Volume 42. – Issue 4. – P. 206–215.
- [24] **Серый, А.С.** Идентификация объектов в задаче автоматической обработки документов / А.С. Серый, Е.А. Сидорова // Труды международной конференции Диалог'2011 «Компьютерная лингвистика и интеллектуальные технологии». – Вып. 10(17). – М.: РГГУ, 2011. – С.580-590.
-

## ONTOLOGY-BASED APPROACH TO MODELING THE PROCESS OF EXTRACTING INFORMATION FROM TEXT

**E.A. Sidorova**

*A.P. Ershov Institute of Informatics Systems SB RAS, Novosibirsk, Russia  
lsidorova@iis.nsk.su*

### Abstract

The article deals with models and methods of knowledge representation, focused on tasks of automatic text processing and information extraction. In the framework of our approach, information extraction is considered as a process of ontology population with information represented as instances of domain concepts. To describe this process three basic models are proposed. The model of the text representation defines the general scheme of text processing and provides the mapping of the received information on the text. The knowledge representation model includes a description of the subject vocabulary, genre models of the text and the models of facts, which allow modeling the processes of information extraction in terms of semantic classes of subject vocabulary and ontology of the subject domain. The attributive model of data representation ensures the preservation of information streams of data that arise in the process of extracting information, and allows the use of ontological methods for solving ambiguity problems and resolving the coreference. Thus, an original technique that allows users to design a text analysis system and simulate the information extraction based on the domain ontology is proposed.

**Key words:** *information extraction, model of text, domain vocabulary, model of fact, ontology population.*

**Citation:** *Sidorova EA. Ontology-based approach to modeling the process of extracting information from text [In Russian]. *Ontology of designing*. 2018; 8(1): 134-151. - DOI: 10.18287/2223-9537-2018-8-1-134-151.*

### Acknowledgment

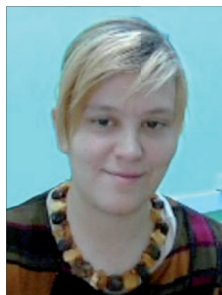
The work was financially supported by the Presidium of the SB RAS (Block 36.1 of the Comprehensive Program of the FNI SB RAS II.1) and the RFBR (grant No. 17-07-01600).

### References

- [1] *Petasis G, Karkaletsis V, Paliouras G, Krithara A, Zavitsanos E.* Ontology Population and Enrichment: State of the Art. In Knowledge-driven multimedia information extraction and ontology evolution. LNAI 6050. Springer-Verlag Berlin; 2011: 134–166.
- [2] *Mel'chuk IA.* Opyt teorii lingvisticheskikh modelei «Smysl ↔ Tekst». Semantika. Sintaksis [Experience of Theory of Linguistic Models “Meaning ↔ Text”. Semantics. Syntax] [In Russian]. – Moscow: Iazyki russkoi kul'tury Publ.; 1999.
- [3] *Narin'jani AS.* TEON-2: from Thesaurus to Ontology and backwards [In Russian]. DIALOG'2002: Proc.of the Int. Workshop. – Moscow: Nauka Publ. 2002; Vol.1: 307–313.
- [4] *Zagorulko YuA.* Semantic technology for development of intelligent systems oriented on experts in subject domain [In Russian]. *Ontology of designing*. 2015; 1(15): 30-46.
- [5] *Dobrov BV, Ivanov VV, Lukashevich NV, Solov'ev VD.* Ontologii i tezaury: modeli, instrumenty, prilozheniia [In Russian]. – Moscow: Binom. Laboratoriia znaniia, Internet-universitet informatcionnykh tekhnologii; 2009.
- [6] *Rubashkin VSh.* Semanticheskii komponent v sistemakh ponimaniia teksta [In Russian]. *Artificial Intelligence: Proc. of the 10th Nat. Conf. with Int. Part CAI-2006.* – Moscow: Fizmatlit; Vol.2: 455-463.
- [7] *Rozental' DE.* Upravlenie v russkom iazyke [In Russian]. – Moscow: Kniga publ.; 1986.
- [8] *Zhuravlyov AO.* Creating of Database of Governance Models for Russian Words [In Russian]. *Artificial intelligence*; 2013; 1: 91-97.
- [9] *Bol'shakova EI.* Iazyk leksiko-sintaksicheskikh shablonov LSPL: opyt ispol'zovaniia i puti [In Russian]. *Programmnye sistemy i instrumenty: Tematicheskii sbornik №15.* – Moscow: MAX Press; 2014. - [http://www.lspl.ru/articles/Paper\\_19\\_LSPL.pdf](http://www.lspl.ru/articles/Paper_19_LSPL.pdf).

- [10] **Kovalev AI, Sidorova AE.** Tool for developing subject dictionaries based on lexical templates DigLex [In Russian]. Knowledge-Ontology-Theory (KONT-15): Proc. of Russian Conf. – Novosibirsk. Mathematics Institute of SB RAS; Vol.1; 2015: 123-130.
- [11] **Horoshevskii VF.** OntosMiner: Semei'stvo sistem izvlecheniia informacii iz mul'tiiazy'chny'kh kollekcii' dokumentov [In Russian]. Artificial Intelligence: Proc. of the 9th Nat. Conf. with Int. Part CAI-2004. – Moscow: Fizmatlit; Vol.2: 573-581.
- [12] **Ermakov AE.** Automatical extraction of facts from texts of personal files: experience in anaphora resolution [In Russian]. Computational Linguistics and Intellectual Technologies: Proc. of the Int. Conf. "Dialogue". - Moscow: Nauka publ.; 2007: 131–135.
- [13] **Gershenson LM, Nozhov IM, Pankratov DV.** Sistema izvlecheniia i poiska strukturirovannoi' informacii iz bol'shikh tekstovy'kh massivov SMI. Arhitekturny'e i lingvisticheskie osobennosti [In Russian]. Computational Linguistics and Intellectual Technologies: Proc. of the Int. Conf. "Dialogue". - Moscow: Nauka Publ.; 2005: 97-101.
- [14] **Azarova IV, Grebenkov AS, Lando TM.** The context schema of predicate arguments for automatic expansion of a domain ontology [In Russian]. Computational Linguistics and Intellectual Technologies: Proc. of the Int. Conf. "Dialogue". - Moscow: RSUH; 2008; 7 (14): 11-16.
- [15] **Yakovchuk EI, Sidorova EA.** Generalized semantic-syntactic models in text processing tasks [In Russian]. A.P. Ershov Informatics Conference (the PSI Conference Series, 11th edition). Science-Intensive Software SIS-PSI 2011 Satellite Workshop. – Novosibirsk: IIS SB RAS; 2011: 287-292.
- [16] **Apresian IuD, Boguslavskii IM, Iomdin BL, Iomdin LL, Sannikov AV, Sannikov VZ, Sizov VG, Tcinman LL.** Sintaksicheski i semanticheski annotirovannyi korpus russkogo iazyka: sovremennoe sostoianie i perspektivy [In Russian]. Russian National Corpus: 2003-2005. – Moscow: Indrik; 2005: 193-214.
- [17] **Blanco X.** Using NooJ for Multipurpose Analysis of Romance Languages Corpora. Corpus Linguistics-2008: Proc. of the Int. Conf. – St. Petersburg; 2008: 40–44.
- [18] **Kononenko IS, Sidorova EA.** Genre Aspects of Websites Classification [In Russian]. Software Engineering; 2015; N8: 32–40.
- [19] **Zagorulko YuA.** Automation of ontologic information gathering about internet-resources for the scientific knowledge portal [In Russian]. Bulletin of the Tomsk Polytechnic University; 2008; 312(5): 114-119.
- [20] **Marino J, Rowley M.** Understanding SCA (Service Component Architecture). – Addison-Wesley Professional; 2009.
- [21] **Garanina N, Bodin E, Sidorova E.** A multi-agent text analysis based on ontology of subject domain. Proc. of 9th Ershov Informatics Conference (PSI'2014). – Novosibirsk: A.P. Ershov Institute of Informatics systems; 2014: 50-56.
- [22] **Garanina N, Sidorova E, Kononenko I.** A Distributed Approach to Coreference Resolution in Multiagent Text Analysis for Ontology Population. Springer International Publishing AG 2018 A. K. Petrenko and A. Voronkov (Eds.): PSI 2017. LNCS 10742; 2018: 1–16.
- [23] **Garanina NO, Sidorova EA, Anureev IS.** Conflict resolution in multi-agent systems with typed relations for ontology population // Programming and Computer Software, July 2016. 2016; 42(4): 206–215.
- [24] **Seriy AS, Sidorova EA.** Object identification in problem of automatic document processing [In Russian]. Computational Linguistics and Intellectual Technologies: Proc. of the Int. Conf. "Dialogue". - Moscow: RSUH; 2011; 10(17): 580-590.

## Сведения об авторе



**Сидорова Елена Анатольевна**, 1977 г. рождения. Окончила Новосибирский государственный университет в 2000 г., к.ф.-м.н. (2006). Старший научный сотрудник лаборатории искусственного интеллекта Института систем информатики им. А.П. Ершова (Новосибирск) с 2014 г., член Российской ассоциации искусственного интеллекта. В списке научных трудов более 120 работ в области компьютерной лингвистики, мультиагентных систем, представления знаний и онтологического инжиниринга.

**Elena Anatolievna Sidorova** (b. 1977) has been senior researcher of the Laboratory of Artificial Intelligence at the A.P. Ershov Institute of Informatics Systems (Novosibirsk, Russia) since 2014, and was a leader of several projects in computational linguistics. She holds MSc degree from the Novosibirsk State University and PhD degree in Computer Science (2006). Dr. Sidorova has about 120 peer-reviewed publications in the field of NLP Systems, Multi-agent Systems, Knowledge Representation, and Ontology Engineering.

УДК 004.02

## ЭВЕРГЕТИКА КАК МЕТОДОЛОГИЧЕСКАЯ ОСНОВА УПРАВЛЕНИЯ ВЫЯВЛЕНИЕМ ДЕФЕКТОВ НА ПРЕДПРОЕКТНОЙ СТАДИИ ЖИЗНЕННОГО ЦИКЛА СИСТЕМ ОБРАБОТКИ ДАННЫХ

В.Е. Гвоздев<sup>1</sup>, Л.Р. Черняховская<sup>2</sup>, Д.В. Блинова<sup>3</sup>

Уфимский государственный авиационный технический университет, Уфа, Россия

<sup>1</sup>wega55@mail.ru, <sup>2</sup>lrchern@yandex.ru, <sup>3</sup>blinova.darya@gmail.com

### Аннотация

Системы обработки данных рассмотрены не только как компоненты информационной структуры, но и как неотъемлемая компонента современного общества, что позволяет согласовывать подходы к управлению дефектами систем обработки данных с подходами, применяемыми при управлении социальными системами. Основное внимание уделяется дефектам систем обработки данных, возникающим на предпроектной стадии жизненного цикла таких систем. Подобные дефекты вызваны отсутствием единого понимания проблемной ситуации – ситуации идентификации потребности в разработке систем обработки данных - акторами, вовлечёнными в «урегулирование» этой ситуации. На основе положений эвергетики и теории интересубъективного управления выделен класс дефектов, возникающих вследствие нарушения коэволюции между состояниями внешней по отношению к системе обработки данных среды и свойствами системы. Выполнена структуризация процесса, соответствующего предпроектной стадии создания систем, результатом которого является спецификация требований пользователей систем обработки данных. Выделены ключевые факторы возникновения дефектов. Новизна подходов, изложенных в работе, заключается в том, что с позиций эвергетики рассмотрена роль спецификации требований пользователей как «моста» между «жизненным миром» и «миром систем».

**Ключевые слова:** эвергетика, дефект, система обработки данных, актор, проблемная ситуация, спецификация требований пользователей.

**Цитирование:** Гвоздев В.Е. Эвергетика как методологическая основа управления выявлением дефектов на предпроектной стадии жизненного цикла систем обработки данных / В.Е. Гвоздев, Л.Р. Черняховская, Д.В. Блинова // Онтология проектирования. – 2018. – Т. 8, №1(27). - С.152-166. – DOI: 10.18287/2223-9537-2018-8-1-152-166.

### Введение

В настоящее время инфосфера является системообразующим фактором жизни общества, необходимым атрибутом человеческой культуры и экономики современного мира. Инфосфера представляет собой совокупность информации, информационной инфраструктуры, субъектов, осуществляющих сбор, формирование, распространение и использование информации, а также системы регулирования возникающих при этом общественных отношений. Системы обработки данных (СОД) наряду с инфокоммуникационными системами относятся к базовым компонентам информационной инфраструктуры. Рассмотрение СОД как неотъемлемой компоненты современного общества обуславливает необходимость положить в основу управления потребительскими свойствами СОД подходы, согласующиеся с подходами, используемыми при управлении социальными системами.

К числу базовых требований, предъявляемых к современным СОД, относятся надежность и функциональная безопасность [1, 2]. Надежность и функциональная безопасность СОД в том числе определяются дефектами, возникающими на разных стадиях жизненного



цикла систем. Проблемам разработки подходов, методов, инструментов для выявления дефектов на разных стадиях жизненного цикла СОД (матрицы проекций; сценарное, исследовательское, *ad hoc* тестирование; и многие другие) посвящена обширная литература. Традиционная точка зрения состоит в том, что основной задачей управления выявлением дефектов является минимизация внеплановых затрат разных ресурсов, обусловленная необходимостью устранения различия потребительских свойств СОД от тех, которые определены в спецификации требований пользователей. Такой подход к определению управления выявлением дефектов является основанием для использования «классических» методов проектирования СОД, ориентированных на использование формальных моделей. Это является предпосылкой сопоставимости результатов решения выбранными методами различных проектных задач разными проектировщиками [3]. Иными словами, проектировщик становится «элементом», «винтиком» в системе проектирования. При этом в качестве эталонов предполагаемых результатов на разных стадиях жизненного цикла СОД выступают спецификации требований, соотносимые с этими стадиями. Естественно, что дефекты спецификаций трансформируются в дефекты системы. Учитывая критическую значимость спецификации требований пользователей (СТП) для результатов проекта (этот документ является входом проекта [4]), можно утверждать, что дефекты этого документа относятся к числу критических факторов провала проекта. Однако методология, методики и инструменты управления выявлением дефектов на стадии разработки СТП СОД далеки от совершенства [5]. Подтверждением этому служит перечень критических факторов успеха и провала программных проектов, приводимый в отчётах Standish Group (<http://www.standishgroup.com/>), в котором качество спецификации входит первую тройку наиболее значимых факторов [6]. Концептуальную основу настоящей работы составляет следующее положение: поскольку СОД являются неотъемлемой частью социальных систем, подходы к управлению выявлением дефектами должны согласовываться с подходами, применяемыми при управлении социальными системами. Это обусловлено тем, что основу формирования спецификации требований пользователей составляют цели, желания заказчиков, их представления о ценностях.

## 1 Анализ «формулы успеха IT-проекта» с позиции эвергетики

Эвергетика – ценностно-ориентированная наука о процессах интересубъективного управления в сложных системах [7, 8]. Центральным понятием эвергетики является «неоднородный актор» – субъект, вовлечённый в урегулирование проблемной ситуации (ПС). Вовлечённость означает заинтересованность субъекта в изменении ситуации и обладание полезными, с точки зрения урегулирования ситуации, ресурсами. Термином «неоднородный» подчёркивается то обстоятельство, что одна и та же ситуация по-разному воспринимается разными акторами [9].

Классическая наука об управлении, методологической основой которой является системный подход, ориентирована на достижение поставленной цели наилучшим, в смысле заранее определённых критериев и ограничений, образом. С нею ассоциируется «мир систем». С миром систем сопрягается «инструментальная рациональность», неоднородный актор рассматривается как обезличенный элемент системы [10]. Эвергетика же исходит из того, что помимо «мира систем» существует «жизненный мир», в котором существует неоднородный актор, стремящийся понять смысл ПС в рамках своих индивидуальных когнитивных ограничений. Осознание смысла ПС служит основанием для разработки персональных онтологических моделей разных акторов по урегулированию ПС [11], причём у каждого из акторов имеется своя персональная онтологическая модель. Вместе с тем заинтересованность всех акторов в урегулировании ПС служит основанием для формирования признаваемой всеми заин-

тересованными сторонами объединённой онтологической модели [12]. Это, в свою очередь, создаёт основу для разработки согласованной стратегии урегулирования ПС. Персональные онтологические модели неоднородных акторов являются основой достижения согласованного понимания ПС за счёт коммуникативных действий (дискурса) [9]. Таким образом, с жизненным миром сопрягается коммуникативная рациональность [7]. Сформированный признаваемый всеми неоднородными акторами способ урегулирования ПС является входом «мира систем» и составляет основу деятельности узких профессионалов, реализующих взаимосвязанную совокупность задач управления в рамках системного подхода. В работах [8, 10] особо подчёркивается, что эвергетика не противопоставляется системному подходу, а дополняет его. Проанализируем с позиции эвергетики известные «формулы успеха IT-проектов».

В 1994 г. Институте управления проектами (Project Management Institute, PMI) была предложена модель, именуемая «*project triangle*»:

(1) *on Time AND on Budget AND with Required Features & Functions*

В рамках этой модели проект считается успешным, если одновременно выполняются условия:

- проект уложился в согласованные заказчиком и исполнителем сроки (*Time*);
- не перерасходован изначально установленный бюджет (*Budget*);
- реализованы согласованные функциональные и нефункциональные требования (*Required Features & Functions*).

Не останавливаясь на анализе ограничений этой модели (см., например, [13]), отметим, что в рамках этой формулы успех определяется тремя взаимосвязанными факторами [14]:

- поддержка проекта высшим руководством;
- вовлечение пользователей в проект;
- ясные цели бизнеса.

Заметим, что два последних условия являются основой формирования требований, обладающих признаками, определёнными в [15]. При этом центральная идея, заложенная в модель (1) – соответствие потребительских свойств СОД потребностям системы управления сложным объектом – в той или иной форме обсуждается в различных литературных источниках, например в [4, 16, 17].

В отчётах Standish Group, стандартах (например, [18]), руководствах (например, [4]) и других источниках подчёркивается необходимость уделять особое внимание вовлечению пользователей в реализацию проектов, что является гарантией соответствия потребительских свойств СОД целям бизнеса. Вместе с тем, фактическое отсутствие положительных тенденций в показателях, характеризующих успешность проектов, а также отсутствие отрицательных тенденций в показателях, характеризующих неуспешные и проваленные проекты (см. отчёты Standish Group за период с 1994 г. по настоящее время) позволяют сделать заключение об ограниченной применимости предлагаемого подхода. При этом обращает на себя внимание тот факт, что согласно отчётам Standish Group, треть проектов являются успешными, причём эта доля остаётся практически неизменной в течение последних полутора десятков лет. Это обстоятельство и позволяет утверждать, что модель (1) обладает признаками адекватности, однако необходимо более чётко определить область её применимости.

Выделенные в [14] факторы успеха явно принадлежат к «жизненному миру». Можно заключить, что в модели (1) составляющая *Required Features & Functions* является отображением жизненного мира (отображением смысла ПС в сознании высшего руководства и пользователей, представленной в виде спецификации требований пользователей). Составляющие же *on Time* и *on Budget* – отображением мира систем. Эти две компоненты соответствуют тому, что в [7] именуется «инструментальной рациональностью»: реализовать систему, обла-

дающую требуемыми потребительскими свойствами (цель управления) в условиях ограничений на ресурсы (*Time, Budget*).

В 2015 г. Standish Group предложила иную «формулу успеха IT-проектов» [6, 13, 19]:

(2) *on Time, on Budget, on Target, on Goal, Value and Satisfaction*

Предлагаемая формула подчёркивает необходимость достижения проектом установленных целей (*Target*), соответствующих желаниям заказчика (*Goal*). При этом полагается, что заказчик должен получить в результате реализации проекта не только требуемые ценности (*Value*), но и удовлетворённость (*Satisfaction*). (На наш взгляд, компонента *Goal, Value, Satisfaction* по смыслу близка к известной модели Кано).

В [13] высказывается опасение, что переход от (1) к (2) не позволит кардинально изменить ситуацию с эффективностью реализации программных проектов. Такое мнение обосновывается тем, что в рамках (2), как и в (1), не выделяются в качестве самостоятельных проблемы формирования консолидированного мнения субъектами, имеющими различное представление о ценностях, связанных со способами достижения целей проекта. Более того, не обсуждается вопрос о том, как могут быть определены ценности, ассоциированные с группами пользователей, задействованных в решении одних и тех же управленческих задач. А это автоматически влечёт за собой вопрос о неопределённости способов формирования целей, адекватных потребностям, связанным с урегулированием ПС. В упоминаемой работе со ссылкой на первоисточники указывается пять причин, обуславливающих сложность проблемы формирования консолидированного мнения. Перечислим их, не рассматривая детально содержание:

- 1) люди исходно имеют когнитивные ограничения (т.е. *персональные интеллектуальные модели* [20]), которые изначально затрудняют полное взаимопонимание;
- 2) люди в буквальном смысле встречают друг друга «по одежке», и готовность человека к общению сильно зависит от его эмоционального интеллекта;
- 3) для человека нет ничего хуже навязанного общения;
- 4) используемые человеком слова довольно абстрактны. Например, за «хорошо», «плохо» стоят разные представления разных людей о добре и зле. В силу этого построенные фразы не несут никакого смысла, пока точно не определены используемые понятия. Но люди, как правило, опускают в своей речи эти определения, считая, что все и так понятно;
- 5) в любом мероприятии конкретный человек преследует конкретные цели. Конфликты интересов сопровождают любую совместную деятельность.

Выделенные причины, на наш взгляд, с позиций теории intersubъективного управления являются обстоятельствами, препятствующими формированию консолидированного мнения относительно подходов к урегулированию ПС.

Проанализируем (2) с позиции эвргетики. Так же, как и в случае (1), можно утверждать, что модель описывает взаимосвязь «жизненного мира» и «мира систем». Жизненному миру соответствуют компоненты *Target, Goal, Value, Satisfaction*; миру систем – *Time, Budget*.

Таким образом, приходим к следующему заключению: во-первых, как (1), так и (2) могут рассматриваться как частные результаты, полученные в рамках научного направления, именуемого «эвргетикой». Во-вторых, эвргетика может составить методологическую основу разноаспектного управления проектами, связанными с созданием СОД, в том числе управления выявлением дефектов на разных стадиях жизненного цикла систем.

СТП - описание внешнего облика СОД - играет роль «моста» между миром систем и жизненным миром. Учитывая, что СТП является фундаментом, влияющим как на свойства результатов проекта, так и систему управления проектом, представляется целесообразным выделить её в качестве самостоятельной подсистемы в архитектуре системы «жизненный мир – мир систем» (рисунок 1). Влияние качества СТП на исход проекта СОД обсуждается в

многочисленных литературных источниках, например, [4, 5, 21]. Вопросы исследования качества спецификаций обсуждаются, например, в [5, 15, 22, 23].



Рисунок 1 – Роль спецификации требований пользователей в системе «жизненный мир – мир систем»

Если спроецировать вышеизложенные рассуждения в область управления выявлением дефектов, можно сделать следующий вывод: дефекты являются неотъемлемой составляющей всех стадий жизненного цикла СОД. До настоящего времени основное внимание уделялось изучению способов предотвращения возникновения дефектов: управления их локализацией и устранением на стадиях проектирования, реализации и сопровождения СОД, т.е. дефектами, соотносимыми с «миром систем». Вместе с тем, дефекты СОД начинают возникать гораздо раньше, на предпроектной стадии, соотносимой с «жизненным миром». Методологическую основу изучения дефектов предпроектной стадии может составить эвергетика.

## 2 Статические и динамические дефекты

Сложившаяся практика устранения дефектов СОД сфокусирована на дефектах, которые связаны с расхождением характеристик системы с характеристиками эталонов, определёнными в спецификациях разного уровня (требования пользователей, системные требования, спецификации модулей). Иными словами, эти дефекты можно соотнести с определённым временным срезом, которому ставится в соответствие некоторое представление об эталонном изделии. Количество и характер дефектов напрямую зависят от сложности компонент СОД [24]. Отмеченные обстоятельства позволяют классифицировать дефекты, соотносимые со спецификациями, как «статические дефекты, обусловленные сложностью конструкции СОД». Инструментами устранения таких дефектов являются в различной степени формализованные технологии тестирования, локализации и устранения ошибок, которые относятся к «миру систем». Последствиями устранения статических дефектов, обусловленных сложностью конструкции СОД, является повышение степени соответствия фактических характеристик СОД эталонным характеристикам, определённым в техническом задании. При этом не

принимается во внимание то обстоятельство, что имеет место моральное старение системы в силу того, что с течением времени эталонная модель СОД будет всё меньше соответствовать желаниям и представлениям о ценностях пользователей. Иными словами, помимо методов и технологий устранения статических дефектов, обусловленных сложностью конструкции СОД, следует разрабатывать методы и технологии устранения динамических (стратегических) дефектов, существо которых определяется моральным старением СОД.

Причиной морального старения является изменение состояния внешней по отношению к СОД среды в результате сложного взаимодействия существующих в ней усиливающих и противодействующих циклов [20]. Сложность взаимодействия проявляется, в частности, в том, что трудно разграничить причины и следствия событий. Поясним это следующим примером.

Известно, что частое внесение изменений в спецификацию требований пользователей является причиной перепроектирования системы, т.е. возникновения дополнительных статических дефектов. С другой стороны, причиной внесения изменений в спецификации требований может служить несоответствие свойств СОД личным представлениям о ценностях отдельных пользователей.

Ограничение на свободу внесения изменений в требования<sup>1</sup>, с одной стороны, может способствовать сокращению сроков разработки СОД. С другой стороны, может явиться причиной нарушения коэволюции [25] между изменяющимися во времени потребностями пользователей и потребительскими характеристиками СОД. Следует подчеркнуть, что изменения потребностей пользователей обусловлены процессами, происходящими в «жизненном мире».

В рассматриваемом примере решение следует искать не в стабилизации требований, а в сбалансированности интенсивностей поступления заявок на внесение изменений в спецификации требований и интенсивностью решения комплекса задач, обусловленных изменением спецификаций. Это может быть обеспечено, например, внедрением более совершенных технологий проектирования и испытания систем.

### **3 Факторы возникновения ошибок и дефектов на предпроектной стадии формирования спецификации требований пользователей**

Стратегия урегулирования ПС (стратегия бизнес-процессов) есть фундамент стратегии построения системы информационной поддержки управления [16]. Из того, что СТП (как одна из моделей СОД [4]) является производной от стратегии урегулирования ПС, следует, что часть дефектов этой фазы жизненного цикла системы обусловлена дефектами стратегии урегулирования ПС. Структура процесса формирования СТП представлена на рисунке 2. Предпосылкой разработки представленной модели является структура процесса принятия решений по урегулированию ПС на основе теории интересубъективного управления, описанная в [7].

Опишем кратко содержание компонент структуры процесса принятия решений, а также выделим основные факторы возникновения дефектов на предпроектной стадии СОД.

#### **3.1 Регистрация симптомов проблемной ситуации**

Регистрация симптомов ПС заключается в осознании субъектами, вовлечёнными в управление сложной системой, неприемлемого несоответствия состояния внешней среды и

<sup>1</sup> Сложная организационная процедура получения разрешения на внесение изменений в требования рассматривается как один из инструментов стабилизации требований в руководствах (см., например, свод руководств ESA-PSS-05-XX).

состояния объекта управления, что выражается в аномальном отклонении поведения объекта управления от *базовой линии*<sup>2</sup>. Основными факторами возникновения дефектов на этом шаге предпроектной стадии (стадии предподготовки формирования СТП) являются:

- неоднозначность идентификации аномального поведения объекта управления на фоне нестабильности внешнего поведения, обусловленного вариативностью параметров внешней среды и внутренних параметров объекта управления [27];
- неопределенность в выделении коренных причин ПС, неоднозначность содержания причинно-следственных связей между коренными причинами и наблюдаемыми симптомами [28], усиливающих и противодействующих процессов с учётом временных задержек [20].

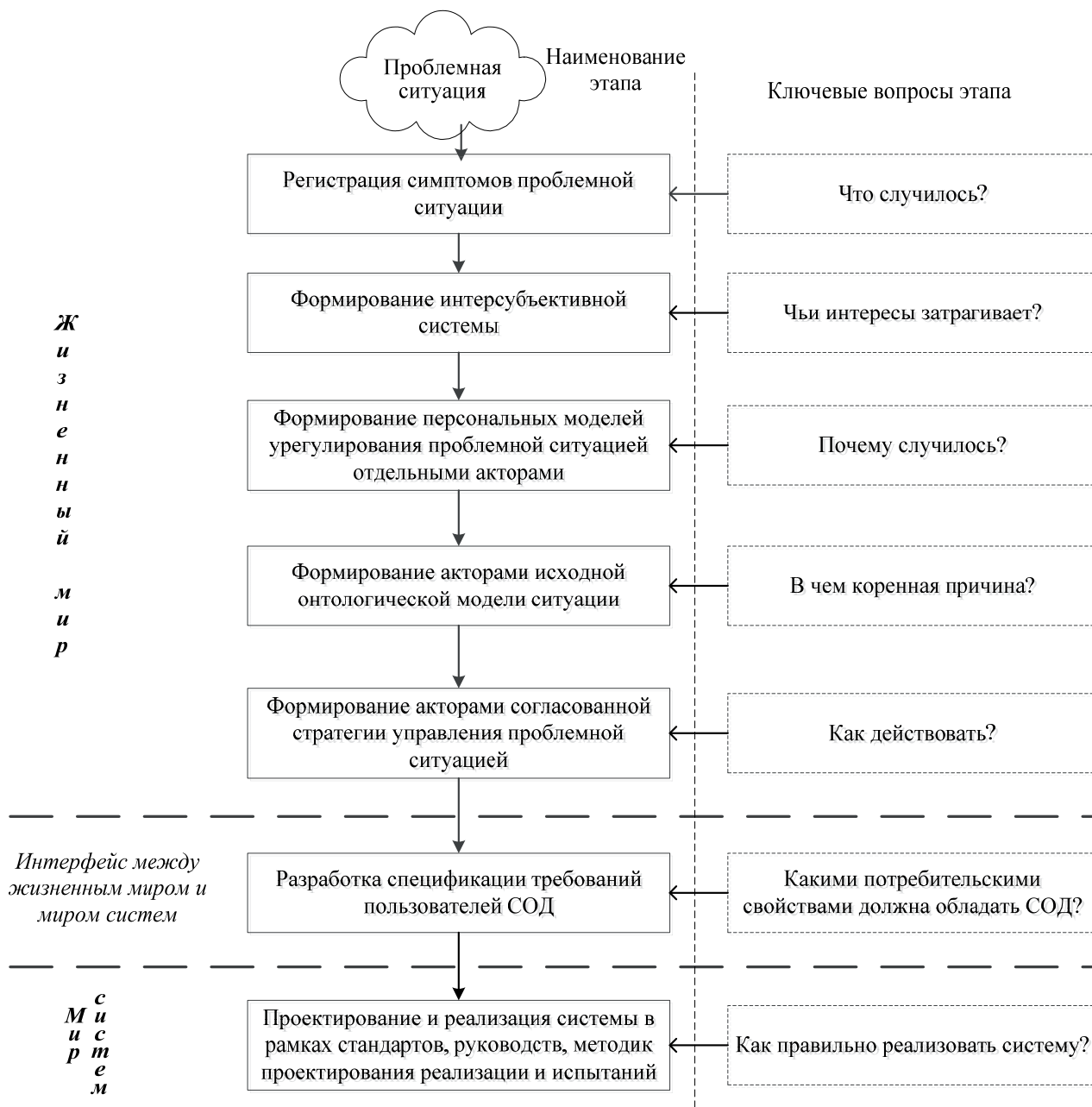


Рисунок 2 – Структура процесса формирования спецификации требований пользователей

<sup>2</sup> Согласно [26] *базовая линия* – спецификация или продукт, которые были официально рассмотрены и согласованы.

### 3.2 Формирование интересубъективной системы

Этот процесс содержит выделение сообщества субъектов, интересы которых затрагиваются ПС; осознающих необходимость урегулирования ПС и способных влиять на ПС (в дальнейшем эти субъекты именуется «неоднородными акторами»<sup>3</sup>). Акторы находятся внутри ПС, то есть являются вовлеченными в её урегулирование. Термином «сообщество» подчёркивается то обстоятельство, что, с одной стороны, акторы имеют общие интересы в урегулировании ПС. С другой стороны, признаётся то, что каждый актор имеет частные интересы, не противоречащие общим [7, 22]. Формирование интересубъективной системы сводится «... к рождению (на основе самоорганизации) сообщества (составного холона), представляющего собою объединение акторов (с имеющимися в их распоряжении ресурсами), которые обладают единым пониманием того, какие вопросы в рамках настоящего объединения регулируются совместно, а какие самостоятельно каждым актором...» [7].

С точки зрения формирования СТП, общие интересы в последующем служат основой формирования обязательных требований к СОД, а частные – опционных [30].

Основными факторами возникновения дефектов этой стадии «предподготовки» формирования СТП являются следующие.

- Привлечение к урегулированию ПС субъектов, не вовлечённых в ПС<sup>4</sup>. Такие субъекты «участвуют» в управлении формально (они согласны с любыми решениями, относящимися к управленческим задачам).
- Привлечение к урегулированию ПС лиц, недостаточно компетентных в решении управленческих задач в условиях неопределённости целей, методов и средств управления. Недостаток компетентности является причиной генерации нереалистичных идей (авторы идей «искренне заблуждаются»). Заметим, что часть дефектов этого типа может быть устранена на стадии формирования онтологической модели ПС.

### 3.3 Формирование отдельными акторами персональных моделей урегулирования проблемной ситуации

Каждый из правообладателей является носителем персональных знаний об окружающем мире. Осмысление ПС, т.е. попытка каждым актором получить ответ на вопрос: «Почему случилось?» и формирование личного видения способов её урегулирования, осуществляется в границах этих персональных знаний с учётом ценностных ориентиров правообладателя. Результатом этой деятельности является построение каждым из правообладателей персональной онтологической модели. Совокупность персональных онтологических моделей правообладателей создаёт коммуникационную<sup>5</sup> основу построения в последующем согласованной онтологической модели ПС, а также подходов к её урегулированию.

Персональные онтологические модели являются ключевой компонентой нематериальных активов системы урегулирования ПС. Они составляют основу проектирования бизнес-процессов, устраняющих коренные причины ПС. Основными факторами возникновения дефектов на этой стадии предподготовки формирования СТП являются:

- ограниченность, «однобокость» предпонимания отдельными правообладателями содержания ПС и связанные с этим дефекты структуризации ПС (выделение объектов, собы-

<sup>3</sup> Подробное толкование термина «неоднородный актор» обсуждается в работах [7, 9]. В области создания и использования СОД достаточно близким по смыслу является понятие «правообладатель». Формальное определение этого понятия приводится в [26]. Проблематика вовлечения правообладателей в управление сложными системами обсуждается в [29].

<sup>4</sup> В [7] со ссылкой на [31] проблемная ситуация определяется как «...когда неудовлетворительное состояние дел уже осознано, но ещё неясно, что нужно делать для его изменения...».

<sup>5</sup> В работе [7] даётся следующее толкование понятия коммуникация: «Коммуникация... не отождествляется с более узким понятием связи и передачи сообщений, а является средством обеспечения прежде всего взаимопонимания акторов...».

тий, процессов, хронологии событий и т. д.) и установление между ними различного рода отношений и связей (причинно-следственных, обобщения, функциональных и т.п.), выделение усиливающих и противодействующих контуров (процессов). В [32] отмечается, что «... для деятельности человека характерно определённое пристрастие к регулярным подходам, представленным упорядоченными, хорошо просматриваемыми и понятными для него действиями, в то время как оригинальные пути решения обычно не относятся к «регулярным»» (заметим, что урегулирование ПС как раз и требует нахождения оригинальных решений);

- различные подходы к структуризации ПС разными акторами [9, 32].

### 3.4 Формирование акторами исходной онтологической модели ситуации

Реализация этой части процесса подготовки СТП предполагает, во-первых, формирование на основе персональных знаний акторов, представленных в форме персональных онтологических моделей, интересубъективного знания<sup>6</sup> (интерсубъективной теории урегулирования проблемной ситуации).

Во-вторых, представление интересубъективного знания в форме, признаваемой всеми акторами онтологической модели ситуации. Основным вопросом этой стадии предпроектной фазы является: «В чём коренные причины ПС?». Фактором возникновения дефектов на этой стадии подготовки СТП является недостаток фактических знаний и ограниченные возможности получения новых знаний. Недостаток знаний заполняется различного рода предположениями и допущениями, часто не имеющими под собою объективной основы.

Онтологическая модель должна обладать свойством полиморфизма. Она должна создавать основу для разработки взаимосвязанных сбалансированных подходов и методов урегулирования ПС как на оперативном, так и на стратегическом уровнях. Это, в частности, означает, что в моделируемой ситуации должны быть выделены совокупности взаимодействующих усиливающих и противодействующих (стабилизирующих) циклов, а также временных задержек. Технологическую основу описания ПС в терминах усиливающих и противодействующих циклов, а также временных задержек составляют архетипы, описанные, например, в [20].

Онтология разрабатывается для создания точных и, по возможности, непротиворечивых определений терминологии в контексте систем обработки данных на основе описательной логики (*Description Logic, DL*). С использованием DL в онтологии создаётся множество аксиом, отображающих знания экспертов в области диагностики и установления причинно-следственных отношений [33]. Наличие аксиом позволяет получить ответы на запросы и организовать поиск решений по урегулированию ПС в соответствии с контекстом предметной области. Онтологическая модель должна давать объяснение наблюдаемым симптомам ПС, а также давать возможность генерации и сопоставления альтернативных вариантов воздействия на коренные причины ПС.

### 3.5 Формирование акторами согласованной стратегии управления проблемной ситуацией

На стадии формирования стратегии управления ПС формулируются ответы на следующие вопросы (строится «дорожная карта» [34]):

- каковы ожидаемые последствия урегулирования ПС для внешней (по отношению к объекту, с которым ассоциируется ПС) среды, т.е. определяется миссия управления ПС;

---

<sup>6</sup> В [7] описано содержание компоненты интересубъективной теории урегулирования ПС: семантической; эмпирической; логической; операционной; нормативной.



- как в результате урегулирования ПС должно измениться внутреннее состояние объекта, с которым ассоциируется ПС, т.е. определяется видение результатов управления ПС;
- определяется архитектура бизнес-процессов, посредством которых осуществляется управление ПС;
- определяется архитектура нематериальных активов (в том числе внешний облик СОД, т.е. СТП, необходимых для реализации бизнес-процессов).

Основным вопросом этой стадии предпроектной фазы СОД является: «Как действовать для урегулирования ПС?». Основными факторами возникновения дефектов при формировании стратегии урегулирования ПС являются ошибки при выделении и анализе взаимодействия усиливающих и противодействующих процессов, а также временных задержек [20].

Последствиями дефектов этой стадии «предподготовки» СТП являются следующие.

- Несоответствие реальных коренных причин ПС причинам, представленным в онтологической модели ПС: «лечат болезнь, а не причину болезни».
- Несоответствие ожидаемого поведения объекта управления (ПС) фактически наблюдаемому (т.е. низкая эффективность управления, вплоть до получения результатов, противоположенных ожидаемому). Заметим, что в этом случае достаточно часто обвиняют систему информационной поддержки управления, а также проектировщиков СОД, которые не могут создавать системы, нужные пользователям (хотя группа пользователей сама может не прийти к единому мнению, что ей нужно).
- Сложность организации коммуникаций между правообладателями. При этом чем выше уровень абстракции компонент персональных моделей, тем больше неопределённость в их описании и смысловой диапазон употребляемых в дискурсе понятий и, следовательно, больше шансов различного толкования разными акторами их содержания. В свою очередь, предлагаемая отдельными акторами высокая степень детализации компонент персональных моделей чревата тем, что непонятные другим акторам термины будут маскировать существо предмета обсуждения.

Основным последствием проявления дефектов является недопустимо частое внесение изменений в функциональные и нефункциональные требования к СОД на стадии формирования спецификации системных требований. Это, в свою очередь, является в последующем причиной низкой эффективности управления программным проектом; неполного соответствия потребительских свойств СОД потребностям и ожиданиям пользователей; нарушения сроков и бюджетов проекта.

### 3.6 Разработка спецификации требований пользователей системы обработки данных

СТП должна давать ответ на вопрос: «Какими потребительскими свойствами должна обладать СОД?» Этот документ играет следующие ключевые, с точки зрения качества программного средства, роли:

- является «мостом» между «жизненным миром» и «миром систем»;
- является фундаментом для построения системы управления проектом;
- является эталоном, с которым пользователи будут сопоставлять результаты реализации проекта;
- является фундаментом реализации всех стадий жизненного цикла СОД [4].

Основными источниками дефектов спецификации требований пользователей являются:

- дефекты стратегии урегулирования ПС (внешний облик СОД является производной от стратегии урегулирования ПС в части нематериальных активов) [16];
- дефекты дорожной карты реализации СОД;

- необоснованные ограничения на способы реализации в формулировках функциональных и нефункциональных требований;
- неопределённость формулировок функциональных и нефункциональных требований, а также описания среды использования СОД, что является причиной различного толкования их содержания разными правообладателями и исполнителями проекта;
- неполнота состава требований, либо излишние требования [5].

Последствиями проявления дефектов в СТП будет получение систем, потребительские свойства которых в неполной мере соответствуют информационным потребностям, необходимым для урегулирования ПС.

### **3.7 Проектирование и реализация программного средства на основе спецификации требований пользователей**

Эта фаза процесса относится к «миру систем» и обеспечивается системой стандартов, руководств, методик, инструментальных средств. Основным вопросом стадии проектирования и реализации программного средства является: «Как правильно реализовать систему, эталонной моделью которой является СТП в рамках общепризнанных подходов проектирования, кодирования, испытаний?»

Основными факторами возникновения дефектов при этом являются следующие.

- Неполное покрытие стандартами, руководствами, методиками системы задач, связанных с реализацией СОД [35]. При этом степень покрытия тем ниже, чем более ранней является стадия реализации программного проекта. Это обусловлено тем, что чем более ранняя стадия программного проекта рассматривается (т.е. чем ближе находимся к началу «конуса неопределённости программного проекта» [17]), тем выше неопределённость содержания задач проекта, и тем более субъект-исполнитель отличается от автомата, преобразующего «вход» в «выход» по определённому алгоритму.
- Неопределённость результатов предыдущей фазы жизненного цикла программного средства. При этом неопределённость уменьшается по мере увеличения формализации постановок задач, т.е. переходу к всё более поздним фазам жизненного цикла программного средства. При этом субъект-исполнитель всё более уподобляется автомату.
- Общий характер рекомендаций стандартов и руководств, низкая дисциплина соблюдения положений стандартов и руководств.
- Недооценка влияния субъективной составляющей на ход проекта и получаемые результаты, необоснованное обезличивание участников проекта, взгляд на них как на обезличенные элементы в составе сложной системы.

Основными последствиями проявления дефектов являются:

- недостаточная совместимость результатов, получаемых разными исполнителями при решении задач, соотносимых с определённой фазой жизненного цикла программного средства;
- компенсация участниками проекта неопределённости содержания спецификации различного рода предположениями, основанными на личных знаниях разработчиков;
- сверхплановые затраты различных ресурсов на обеспечение совместимости результатов, получаемых разными участниками программного проекта.

### **Заключение**

Возрастающая роль инфокоммуникационных систем как фактора управления безопасностью общества, с одной стороны, качественное изменение сложности и масштабов объектов управления с другой, делают необходимым совершенствование методических, теоретиче-

ских, методологических и инструментальных платформ построения, эксплуатации и сопровождения СОД. Причиной несоответствия свойств этих систем реальным информационным потребностям, связанным с урегулированием ПС, являются дефекты разной природы, среди которых помимо проектных и технологических дефектов следует выделить дефекты, соотносимые с «жизненным миром» субъектов, вовлечённых в урегулирование ПС, то есть дефекты предпроектной стадии.

Источником дефектов предпроектной стадии является фундаментальная неопределённость, обусловленная ограниченностью персональных знаний правообладателей об окружающем их мире. Это не позволяет в полной мере устранить коренные причины ПС, оказывать влияние на развивающиеся, уравнивающие процессы, а также временные задержки.

В статье с позиций эвргетики сделана попытка обоснования положения о необходимости выявления дефектов, соотносимых с фазой предподготовки проектирования СОД. Выявление дефектов этого типа делает необходимым дополнение известных подходов выявления дефектов, соотносимых с миром систем, подходами, соотносимыми с жизненным миром акторов.

## Благодарности

Работа поддержана грантом РФФИ 16-08-00442 «Управление функциональной безопасностью аппаратно-программных комплексов в составе сложных технических систем».

## Список источников

- [1] *Липаев, В.В.* Надежность программных средств / В.В. Липаев. – М.: СИНТЕГ, 1998. – 232 с.
- [2] *Липаев, В.В.* Функциональная безопасность программных средств / В.В. Липаев. – М.: СИНТЕГ, 2004. – 348 с.
- [3] *Вагнер, Г.* Основы исследования операций, Т.1 / Г. Вагнер. – М.: Мир, 1972. – 335 с.
- [4] ESA PSS-05-02 Guide to the user requirements definition phase, March 1995.
- [5] *Куликов, С.С.* Тестирование программного обеспечения. Базовый курс / С.С. Куликов. – Минск: Четыре четверти, 2017. – 312 с.
- [6] Standish Group 2015, Chaos Report – Q&A with Jennifer Lynch.
- [7] *Виттих, В.А.* Введение в теорию интересубъективного управления / В.А. Виттих. – Самара: СамНЦ РАН, 2013. – 64 с.
- [8] *Виттих, В.А.* Дополнительность системного и обыденного подходов к урегулированию проблемных ситуаций в повседневности / В.А. Виттих // Проблемы управления и моделирования в сложных системах: Труды XIX Международной конф. (12-15 сентября 2017 г., Самара). – Самара: СамНЦ РАН, 2017. – С. 4-11.
- [9] *Виттих, В.А.* Организация сложных систем / В.А. Виттих. – Самара: СамНЦ РАН, 2010. – 66 с.
- [10] *Виттих, В.А.* Неоднородный актор и повседневность как ключевые понятия эвргетики: препринт / В.А. Виттих. – Самара: ИПУСС РАН, 2014. – 12 с.
- [11] *Смирнов, С.В.* Онтологическое моделирование в ситуационном управлении / С.В. Смирнов // Онтология проектирования. - 2012. - №2(4). - С. 16-24.
- [12] *Смирнов, С.В.* Формальный поход к представлению смысла проблемной ситуации в процессах коллективного принятия решений / С.В. Смирнов // Труды XII Всероссийского совещания по проблемам управления ВСПУ-2014 (16-19 июня 2014 г., Москва, Россия). – М.: ИПУ РАН, 2014. - С. 6261-6270.
- [13] *Тимофеев, А.Н.* Почему падают IT-проекты? / А.Н. Тимофеев // Научно-образовательный журнал «Практика проектирования систем». – 2017. – С. 2-11.
- [14] CHAOS MANIFESTO 2014: Value versus Success & The Orthogonals. – The Standish Group International, Inc.
- [15] IEEE STD 830 Recommended Practice for Software Requirements Specifications, 1998.
- [16] Учебник 4СЮ. Версия 1.0. – М.: 4СЮ, 2011. – 228 с.
- [17] *Макконнелл, С.* Сколько стоит программный проект / С. Макконнелл. – М.: Русская Редакция; СПб.: Питер, 2007. – 304 с.
- [18] ГОСТ Р 51904-2002 Программное обеспечение встроенных систем. Общие требования к разработке и документированию. – М.: Стандартинформ, 2005.

- [19] **Summers, P.** 98.8%! Is project failure acceptable and inevitable? / In proc of Conference: IRNOP 2015, At Barlett School, UCL, London, 2015.
- [20] **Сенге, П.М.** Танец перемен. Новые проблемы самообучающихся организаций: / П. Сенге, Ш. Робертс, Б. Смит, А. Клейнер, Р. Росс, Д. Рот. – М.: ЗАО «Олимп-Бизнес», 2012. – 624 с.
- [21] **Гвоздев, В.Е.** Системное и структурное моделирование требований к программным продуктам и проектам / В.Е. Гвоздев, Д.В. Блинова, Н.И. Ровнейко, О.П. Ямалова // Программная инженерия. – 2013. – №5. – С. 2-10.
- [22] **Гвоздев, В.Е.** Предупреждение дефектов на ранних стадиях проектирования аппаратно-программных комплексов на основе положений теории интересубъективного управления / В.Е. Гвоздев, Д.В. Блинова, Л.Р. Черныховская // Онтология проектирования. – 2016. – Т. 6, №4(22). – С. 452-464.
- [23] **Гвоздев, В.Е.** Анализ функциональных возможностей аппаратно-программных комплексов на ранних стадиях проектирования с учетом мнений правообладателей / В.Е. Гвоздев, Д.В. Блинова // Программная инженерия. – 2016. – Т. 7.– №9. – С. 395-399.
- [24] ESA PSS-05-10 Guide to software verification and validation, March 1995.
- [25] **Rzevski, G.** Self-Management in the Emerging Global Society / G. Rzevski // Проблемы управления и моделирования в сложных системах: Труды XVII Международной конф. (22-25 июня 2015 г., Самара, Россия). – Самара: СамНЦ РАН, 2015. – С. 3-9.
- [26] ГОСТ Р ИСО/МЭК. 15288 – 2005. Информационные технологии. Системная инженерия. Процессы жизненного цикла систем. – М.: Стандартиформ, 2006.
- [27] **Махутов, Н.А.** Оценка уязвимости технических систем и ее место в процедуре анализа риска / Н.А. Махутов, Д.О. Резников // Проблема анализа риска. – 2008. – Т.5. – №3. – С. 72–85.
- [28] **Kataoka, T.** The Analyzing Method of Root Causes for Software Problems / T. Kataoka, K. Furuto, T. Matsumoto // SEI Technical Review. – 2011. – № 73. – P. 81-85.
- [29] **Гельруд, Я.Д.** Управление проектами: методы, модели, системы / Я.Д. Гельруд, О.В. Логиновский // Под ред. А.Л. Шестакова. – Челябинск: Издательский центр ЮУрГУ, 2015. – 330 с.
- [30] **Гвоздев, В.Е.** Онтологический анализ дефектов при проектировании компонентов аппаратно-программных комплексов / В.Е. Гвоздев, Д.В. Блинова // Онтология проектирования. – 2015. – Т. 5, №4(18). – С. 399-410. DOI: 10.18287/2223-9537-2016-6-4-452-464.
- [31] **Новиков, А.М.** Методология / А.М. Новиков, Д.А. Новиков. – М.: Синтег, 2007. – 668 с.
- [32] **Козляков, В.В.** Поиск новых технических решений в машиностроении на ранних этапах проектирования / В.В. Козляков, Д.Л. Раков // Машиностроение и инженерное образование. – 2008. – №1. – С. 2-11.
- [33] **Рассел, С.** Искусственный интеллект: современный подход. 2-е изд. / С. Рассел, П. Норвиг. – М.: Издательский дом «Вильямс», 2007.
- [34] **Каплан, Р.С.** Сбалансированная система показателей: от стратегии к действию / Р.С. Каплан, Д.П. Нортон. – М.: ЗАО «Олимп-Бизнес», 2003. – 214 с.
- [35] **Липаев, В.В.** Процессы и стандарты жизненного цикла сложных программных средств: Справочник / В.В. Липаев // – М.: Синтег, 2006. – 276 с.
- 

## EVERGETICS AS A METHODOLOGICAL BASIS FOR MANAGING OF DEFECTS IDENTIFICATION AT THE PRE-DESIGN STAGE OF DATA PROCESSING SYSTEMS

**V.E. Gvozdev<sup>1</sup>, L.R. Chernyakhovskaya<sup>2</sup>, D.V. Blinova<sup>3</sup>**

*Ufa State Aviation Technical University, Ufa, Russia,*

*<sup>1</sup>wega55@mail.ru, <sup>2</sup>lrchern@yandex.ru, <sup>3</sup>blinova.darya@gmail.com*

### Abstract

Data processing systems are considered not only as components of the information structure, but also as an integral component of modern society. This allows to coordinate approaches for defect management of data processing systems with approaches used in managing social systems. The main attention is paid to the defects of the data processing systems that arise at the pre-project stage, caused by lack of uniform understanding of the problem situation by the actors, involved in resolving the problem situation. On the basis of the provisions of evergetics and the theory of intersubjective management a class of defects is allocated, arising due to violation of co-evolution between the states of external to the data processing systems environment and the properties of the system. The structuration of the process, corresponding to a pre-design stage was carried out, result of which is the specification of user's requirements of data processing

systems. Key factors of defects occurrence are allocated. The novelty of the approaches described in the paper is in considering from the positions of the evergetics the role of specification of user requirements as a "bridge" between the "life world" and the "world of systems".

**Key words:** *evergetics, defect, data processing system, actor, problem situation, user requirements specification.*

**Citation:** *Gvozdev VE, Chernyakhovskaya LR, Blinova DV. Evergetics as a methodological basis for managing of defects identification at the pre-design stage of data processing systems [In Russian]. Ontology of designing. 2018; 8(1): 152-166. - DOI: 10.18287/2223-9537-2018-8-1-152-166.*

## Acknowledgment

The work is supported by RFBR grant 16-08-00442 «Management of functional safety of hardware-software systems as part of complex technical systems».

## References

- [1] **Lipaev VV.** Reliability of software [In Russian]. – Moscow: SINTEG; 1998.
- [2] **Lipaev VV.** Functional security of software [In Russian]. – Moscow: SINTEG; 2005.
- [3] **Wagner HM.** Principles of operations research. – New Jersey: Prentice Hall; 1969.
- [4] ESA PSS-05-02 Guide to the user requirements definition phase, Mar 1995.
- [5] **Kulikov SS.** Testing of the software. A basic course [In Russian]. – Minsk: Chetyri Chetverti; 2017.
- [6] Standish Group 2015, Chaos Report – Q&A with Jennifer Lynch.
- [7] **Vittikh VA.** Introduction to the Theory of Intersubjective Management [In Russian]. – Samara: Samara Scientific Center of RAS; 2013.
- [8] **Vittikh VA.** Complementarity of system and ordinary approaches to settlement of problem situations in daily occurrence. [In Russian]. Complex Systems: Control and Modeling Problems: Proc. of the XIX Int. Conf. (Samara, Russia, 2017, Sept. 12-15). – Samara: Samara Scientific Center of RAS; 2017: 4-11.
- [9] **Vittikh VA.** Organization of complex systems [In Russian]. – Samara: Samara Scientific Center of RAS; 2010.
- [10] **Vittikh VA.** Heterogeneous Actor and Everyday Life as Key Concepts of Evergetics: pre-print [In Russian]. – Samara: Institute for Control of Complex Systems of RAS; 2014.
- [11] **Smirnov SV.** Ontological modeling in situational management [In Russian]. *Ontology of Designing.* 2012; 2: 16-24.
- [12] **Smirnov SV.** Formal approach to representation of the situation sense in collaborative decision-making [In Russian]. Proc. of the 12th All-Russian Meeting on the Management (June 16-19, 2014, Moscow, Russia). - Moscow: Institute of Control Sciences of RAS; 2014: 6261-6270.
- [13] **Timofeev AN.** Why IT projects is fall? [In Russian]. Scientific and educational magazine "Praktika Proyektirovaniya Sistem"; 2017: 2-11.
- [14] CHAOS MANIFESTO 2014: Value versus Success & The Orthogonals. – The Standish Group International, Inc.
- [15] IEEE STD 830 Recommended Practice for Software Requirements Specifications; 1998.
- [16] 4 CIO: Textbook. Version 1.0. – Moscow: 4 CIO; 2011.
- [17] **McConnell S.** Software Estimation: Demystifying the Black Art. Microsoft Press; 2007.
- [18] GOST R 51904-2002 Embedded system software. General requirements for development and documentation [In Russian]. – Moscow: Standartinform; 2005.
- [19] **Summers P.** 98.8%! Is project failure acceptable and inevitable? In proc of Conference: IRNOP 2015, At Barlett School, UCL, London; 2015.
- [20] **Senge PM, Roth G, Roberts C, Ross R, Smith B, Kleiner A.** The Dance of Change: The Challenges of Sustaining Momentum in Learning Organizations. – Nicholas Brealey Publishing Lt; 1999.
- [21] **Gvozdev VE, Blinova DV, Rovneiko NI, Yamalova OP.** Systemic and Structural Modelling of Requirements to Software and Software Projects [In Russian]. *Programmaja Injenerija.* 2013; 5:2-10.
- [22] **Gvozdev VE, Blinova DV, Chernyakhovskaya LR.** Prevention of defects on the early stages of designing hardware-software complexes on the basis of the positions of the theory of intersubjective management [In Russian]. *Ontology of Designing.* 2016; V.6; 4(22): 452-464. DOI: 10.18287/2223-9537-2016-6-4-452-464.
- [23] **Gvozdev VE, Blinova DV.** Analysis of Functional Possibilities of Hardware-Software Complexes in the Early Stages of Designing with Considering the Opinions of Stakeholders [In Russian]. *Programmaja Injenerija.* 2016; V.7; 9:395-399.
- [24] ESA PSS-05-10 Guide to software verification and validation, March 1995.

- [25] **Rzevski G.** Self-Management in the Emerging Global Society. Complex Systems: Control and Modeling Problems: Proc. of the XVII Int. Conf. (Samara, Russia, 2015, June 22-25). - Samara: Scientific Center of RAS; 2015: 3-9.
- [26] ISO/IEC 15288:2002 «System engineering — System life cycle processes».
- [27] **Makhutov NA, Reznikov DO.** Assessment of the vulnerability of technical systems and its place in the risk analysis procedure [In Russian]. Issues of Risk Analysis. 2008; V.5; 3: 72-85.
- [28] **Kataoka T, Furuto K, Matsumoto T.** The Analyzing Method of Root Causes for Software Problems. SEI Technical Review. 2011; 73: 81-85.
- [29] **Gelrud YaD, Loginovsky OV.** Project Management: Methods, Models, Systems [In Russian]. – Chelyabinsk: Publishing center of South Ural State University; 2015.
- [30] **Gvozdev VE, Blinova DV.** Ontological analysis of defects at the designing of hardware-software complexes [In Russian]. Ontology of Designing. 2015; 5(4): 399-410.
- [31] **Novikov AM, Novikov DA.** Methodology [In Russian]. Moscow: Sinteg; 2007.
- [32] **Kozlyakov VV, Rakov DL.** Search of new technical solutions in mechanical engineering at early design stages [In Russian]. Mashinostroeniye i injenernoe obrazovaniye 2008; 1: 2-11.
- [33] **Russell SJ, Norvig P.** Artificial Intelligence: A Modern Approach (2nd Edition). – Prentice Hall, 2002.
- [34] **Kaplan RS, Norton DP.** The Balanced Scorecard: Translating Strategy into Action. – Harvard Business Review Press, 1996.
- [35] **Lipaev VV.** Processes and standards of complicated software lifecycle. Reference book [In Russian]. - Moscow: Sinteg; 2006.

## Сведения об авторах



**Гвоздев Владимир Ефимович**, 1955 г. рождения. Окончил Уфимский авиационный институт в 1978 г., д.т.н. (2000), профессор. Заведующий кафедрой технической кибернетики Уфимского государственного авиационного технического университета. В списке научных трудов более 170 работ, в том числе монография и учебные пособия по проектированию и реализации программных продуктов и проектов. Проводит исследования в области открытых информационных систем, прикладной статистики, теории надёжности, контроля и управления состоянием окружающей среды, управления программными проектами, функциональной безопасности.

**Vladimir Efimovich Gvozdev** (b. 1955) graduated from the Ufa Aviation Institute in 1978, D. Sc. Eng. (2000), prof. He is Head of Technical Cybernetic Department at Ufa State Aviation

Technical University. He is co-author of more than 170 publications in the field of design and production of program products and projects, open information systems, applied statistic, theory of reliability, monitoring and management of environmental condition, function safety.



**Черняховская Лилия Рашитовна**, 1947 г. рождения. Окончила Уфимский авиационный институт в 1970 г., д.т.н. (2004). Профессор кафедры технической кибернетики Уфимского государственного авиационного технического университета. В списке научных трудов около 150 работ в области системного анализа, интеллектуальных информационных систем и систем поддержки принятия решений.

**Liliya Rashitovna Chernyakhovskaya** (b.1947) graduated from the Ufa Aviation Institute in 1970, Dr. of Tech. Sci. (2004). Professor of Ufa State Aviation Technical University (Technical cybernetics department). Co-author of about 150 publications in the field of system analysis, intellectual information systems and decision support systems.



**Блинова Дарья Викторовна**, 1984 г. рождения. Окончила Уфимский государственный авиационный технический университет в 2006 г., к.т.н. (2009). Доцент кафедры технической кибернетики Уфимского государственного авиационного технического университета. Проводит исследования в области прикладной статистики, управления природно-техническими системами, функциональной безопасности, дефектологии компонентов аппаратно-программных комплексов.

**Darya Viktorovna Blinova** (b.1984) graduated from the Ufa State Aviation Technical University in 2006, PhD (2009). She is Assistant Professor at Ufa State Aviation Technical University (Department of Technical Cybernetic). She conducts research in the field of applied statistic, management of natural and technical systems, function safety, defectology of hardware-software complexes components.

## **XVIII International Conference on Data Science and Intelligent Analysis of Information (ICDSIAI'2018)**

**June 4–7, 2018, Kiev, Ukraine**

### **TOPICS**

- *Big data analysis methods and data mining, machine learning*
- *Data analysis using fuzzy mathematics, soft computing, computing with words*
- *Applications of data science to economics, business, and public policy*
- *Applied data science and data mining systems*
- *Economics of blockchain and cryptocurrencies, including governance, monetary policy, contract theory*
- *Knowledge engineering methods and ontology engineering*
- *Intelligent educational systems, intelligent systems for distance learning and knowledge management*
- *Intelligent search and information analysis in local and global networks*

### **CONTACTS**

Applied Mathematics Department Igor Sikorsky Kyiv Polytechnic Institute  
Email: iaikpi@ukr.net. Site of the conference: pma.fpm.kpi.ua/en/icdsiai

---

## **XXIII Байкальская Всероссийская конференция с международным участием**

### **Информационные и математические технологии в науке и управлении**

**1–8 июля 2018 г. Иркутск, Россия**

#### **НАПРАВЛЕНИЯ КОНФЕРЕНЦИИ**

- *Теоретические и методологические аспекты информационных и математических технологий*
- *Математическое моделирование в научных исследованиях, вычислительная математика, оптимизация*
- *Методы, технологии и инструментальные средства создания интеллектуальных энергетических систем*
- *Методы и системы искусственного интеллекта, интеллектуальные вычисления*
- *Информационное и семантическое моделирование, семантические вычисления*
- *Корпоративные информационные, геоинформационные, интеллектуальные системы*
- *Параллельные, распределенные, облачные и агентные вычисления*
- *Кибербезопасность (защита информационных систем критически важных инфраструктур)*

#### **КОНТАКТНЫЕ ДАННЫЕ ОРГАНИЗАТОРОВ КОНФЕРЕНЦИИ**

Ученый секретарь конференции: imt@isem.irk.ru. Сайт журнала: <http://www.imt-journal.ru/>

---

## **XX Международная научная конференция Проблемы управления и моделирования в сложных системах (ПУМСС-2018)**

**3-6 сентября 2018 г. Самара, Россия**

#### **НАПРАВЛЕНИЯ КОНФЕРЕНЦИИ**

- *Методы управления и оптимизации в сложных технических системах*
- *Интеллектуальные технологии в сложных системах*
- *Измерения, контроль и диагностика в экстремальных условиях*
- *Процессы управления в обществе*
- *Агрокибернетика и агроинформатика*
- *Биокибернетика и биоинформатика*

**ИПУСС РАН.** Самара, 443020, ул. Садовая, 61, тел. +7 (846) 333 27 70, e-mail: [cscmp@iccs.ru](mailto:cscmp@iccs.ru)  
Сайт конференции: <http://www.iccs.ru/cscmp/cscmp.html>



**Шестнадцатая национальная конференция  
ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ**  
с международным участием (КИИ-2018)  
24 по 27 сентября 2018 г. Москва, Россия

**НАПРАВЛЕНИЯ КОНФЕРЕНЦИИ**

- Автоматизация рассуждений и интеллектуальный анализ данных
- Мультиагентные системы, интеллектуальные роботы и поведение
- Анализ естественного языка и понимание текстов
- Нечёткие модели и мягкие вычисления
- Когнитивное компьютерное моделирование
- Интеллектуальные системы и приложения методов ИИ

Сайт конференции: <http://2018.mcai.ru>

---

**21st International multi-conference on  
ADVANCED COMPUTER SYSTEMS (ACS 2018)**  
Międzyzdroje, Poland, September 24-26, 2018

**TOPICS**

- Artificial Intelligence
- Software Technologies
- Information Technology Security
- Multimedia Systems
- Design of Information Systems

All papers accepted and presented in ACS 2018 will be invited for further revisions and publication in Springer's "Advances in Intelligent Systems and Computing" (AISC) series. The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink and indexed in Web of Science.

Web site: <http://acs.zut.edu.pl>

---



6-7 September 2018 | London, UK

**Intelligent Systems Conference (IntelliSys) 2018**  
6-7 September 2018, London, United Kingdom

Intelligent Systems Conference (**IntelliSys**) 2018 will focus on areas of intelligent systems and artificial intelligence (AI) and how it applies to the real world. The goal of the conference is to be a premier venue for researchers and industry practitioners to share new ideas, research results and development experiences in various fields.

- Talks by industry experts on the state-of-the-art in computer science
- Lectures by eminent scientists designed to inspire and inform
- Presentations by innovative researchers coming from 50+ countries
- Discussion-oriented sessions and networking breaks to enable collaborations

Web site: <http://saiconference.com/IntelliSys>

---



**Конференция ИИ в бизнес – AI CONFERENCE**  
19 апреля 2018 г., Москва, Россия

Мероприятие для разработчиков, предпринимателей, стартапов, создателей платформ на основе Big Data и компаний, которые используют ИИ в бизнесе. Эксперты обсудят способы применения ИИ в медицине, ритейле, страховании, маркетинге, финансовой отрасли и сфере обслуживания. Спикеры - представители Nvidia, Microsoft, МТС, VisionLabs, Addi, Marketingbot, «Сколтех», «Сбербанк» и других компаний. В выставочной зоне IT-компании представят передовые AI-разработки. В рамках события пройдет конкурс стартапов, предлагающих современные решения в области ИИ. Организатор мероприятия – Smile-Expo.

Сайт конференции: <https://aiconference.ru/ru>

---