

# ОНТОЛОГИЯ ПРОЕКТИРОВАНИЯ



Russian  
Science  
Citation  
Index

40-ой выпуск!

Vol **11**  
N **2**  
2021

Scientific journal

Volume 11

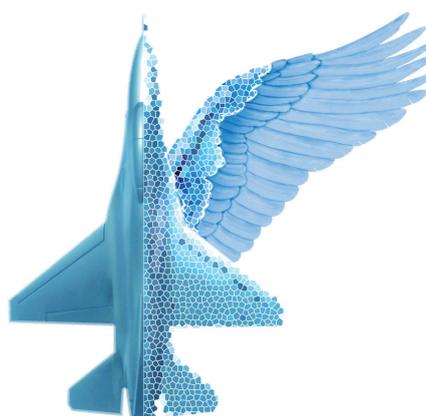
№ 2

**ОНТОЛОГИЯ**  
**ПРОЕКТИРОВАНИЯ**

Научный журнал

Том 11

№ 2



## Editorial Board - Редакционная коллегия

Nikolay M. **Borgest\***, Ph.D., Professor Samara University, Member of IAOA, Samara, Russia  
 Stanislav N. **Vasiliev\***, Doctor of Phys. and Math. Sciences, Professor, Academician of RAS, ICS RAS, Moscow, Russia  
 Tatiana A. **Gavrilova\***, Doctor of Technical Sciences, Professor, GSOM SPbU, St.-Petersburg, Russia  
 Vladimir G. **Gainutdinov\***, Doctor of Technical Sciences, Professor, KNITU-KAI, Kazan, Russia  
 Vladimir V. **Golenkov\***, Doctor of Technical Sciences, Professor, BSUIR, Minsk, Belarus  
 Vladimir I. **Gorodetsky\***, Doctor of Technical Sciences, Professor, InfoVings LLC, St. Petersburg, Russia  
 Valeriya V. **Gribova\***, Doctor of Technical Sciences, Senior Researcher, IAPU of the Far Eastern Branch of RAS, Vladivostok, Russia  
 Yury A. **Zagorulko\***, Ph.D., Senior Researcher, ISI of the Siberian Branch of RAS, Novosibirsk, Russia  
 Anton V. **Ivaschenko\***, Doctor of Technical Sciences, Professor, Samara State Technical University, Samara, Russia  
 Valery A. **Komarov\***, Doctor of Technical Sciences, Professor, Samara University, Samara, Russia  
 Vladik **Kreinovich\***, Ph.D., Professor, University of Texas at El Paso, El Paso, USA  
 Victor M. **Kureichik\***, Doctor of Technical Sciences, Professor, Southern Federal University, Taganrog, Russia  
 Dmitry V. **Lande\***, Doctor of Technical Sciences, Senior Researcher, IPRI NASU, Kiev, Ukraine  
 Paulo **Leitao\***, Professor at Polytechnic Institute of Bragança, Bragança, Portugal  
 Vladimir **Marik\***, Professor, Scientific Director of the CIIRC of the Czech Technical University in Prague, Praha, Czech Republic  
 Lyudmila V. **Massel\***, Doctor of Technical Sciences, Professor, ISEM of the Siberian Branch of RAS, Irkutsk, Russia  
 Aleksandr Yu. **Nesterov\***, Doctor of Philosophy, Professor, Samara University, Samara, Russia  
 Dmitry A. **Novikov\***, Doctor of Technical Sciences, Professor, Corresponding Member of RAS, ICS RAS, Moscow, Russia  
 Alexander V. **Palagin\***, Doctor of Technical Sciences, Professor, Academician of the NASU, Ins. of Cybernetics, Kiev, Ukraine  
 Semyon A. **Piyavsky\***, Doctor of Technical Sciences, Professor, Moscow City Pedagogical University, Samara, Russia  
 Yury M. **Reznik\***, Doctor of Philosophy, Professor, Institute of Philosophy of RAS, Moscow  
 George **Rzevski\***, Professor, Open University, London, UK  
 Peter O. **Skobelev\***, Doctor of Technical Sciences, «Smart solutions» Scientific Production Co., Samara, Russia  
 Sergey V. **Smirnov\***, Doctor of Technical Sciences, ICCS RAS, member of IAOA, Samara, Russia  
 Dzhavdet S. **Suleymanov\***, Doctor of Technical Sciences, Professor, Academician of the Republic of Tatarstan, Kazan, Russia  
 Boris E. **Fedunov\***, Doctor of Technical Sciences, Professor, State Research Institute of Aviation Systems, Moscow, Russia  
 Altynbek **Sharipbay\***, Doctor of Technical Sciences, Professor, Institute of Artificial Intelligence, Nur-Sultan, Kazakhstan  
 Boris Ya. **Shvedin\***, Ph.D., Member of IAOA, Dan Rose LLC, Rostov-on-Don, Russia

**Боргест** Николай Михайлович\*, к.т.н., профессор, Самарский университет, член IAOA, Самара, Россия  
**Васильев** Станислав Николаевич\*, д.ф.-м.н., профессор, академик РАН, ИПУ РАН, Москва, Россия  
**Гаврилова** Татьяна Альбертовна\*, д.т.н., профессор, ВШМ СПбУ, Санкт-Петербург, Россия  
**Гайнутдинов** Владимир Григорьевич, д.т.н., профессор, КНИТУ-КАИ, Казань, Россия  
**Голенков** Владимир Васильевич\*, д.т.н., профессор, БГУИР, Минск, Беларусь  
**Городецкий** Владимир Иванович\*, д.т.н., профессор, ООО «ИнфоВингс», Санкт-Петербург, Россия  
**Грибова** Валерия Викторовна\*, д.т.н., с.н.с., ИАПУ ДВО РАН, Владивосток, Россия  
**Загорюлько** Юрий Алексеевич\*, к.т.н., с.н.с., ИСИ СО РАН, Новосибирск, Россия  
**Иващенко** Антон Владимирович, д.т.н., профессор, СамГТУ, Самара, Россия  
**Комаров** Валерий Андреевич, д.т.н., профессор, Самарский университет, Самара, Россия  
**Крейнвич** Владик, профессор, Техасский университет Эль Пасо, Эль Пасо, США  
**Курейчик** Виктор Михайлович\*, д.т.н., профессор, Южный федеральный университет, Таганрог, Россия  
**Ландэ** Дмитрий Владимирович\*, д.т.н., с.н.с., ИПРИ НАН Украины, Киев, Украина  
**Лейтао** Пауло, профессор, Политехнический институт, Браганса, Португалия  
**Марик** Владимир, профессор, научный директор ЧИИРК Чешского технического университета, Прага, Республика Чехия  
**Массель** Людмила Васильевна\*, д.т.н., профессор, ИСЭМ СО РАН, Иркутск, Россия  
**Нестеров** Александр Юрьевич, д.филос.н., профессор, Самарский университет, Самара, Россия  
**Новиков** Дмитрий Александрович, д.т.н., профессор, член-корреспондент РАН, ИПУ РАН, Москва, Россия  
**Палагин** Александр Васильевич, д.т.н., профессор, академик НАН Украины, Ин-т кибернетики, Киев, Украина  
**Пиявский** Семён Авраамович, д.т.н., профессор, Московский город.педагог.университет, Самара, Россия  
**Резник** Юрий Михайлович, д.филос.н., профессор, Институт философии РАН, Москва, Россия  
**Ржевский** Георгий, профессор, Открытый университет, Лондон, Великобритания  
**Скобелев** Петр Олегович, д.т.н., НПК «Разумные решения», Самара, Россия  
**Смирнов** Сергей Викторович\*, д.т.н., ИПУСС РАН – СамНЦ РАН, член IAOA, Самара, Россия  
**Сулейманов** Джавдет Шевкетович\*, д.т.н., профессор, академик АН РТ, Казань, Россия  
**Федунов** Борис Евгеньевич\*, д.т.н., профессор, ГосНИИ авиационных систем, Москва, Россия  
**Шарипбай** Алтынбек\*, д.т.н., профессор, Ин-т искусственного интеллекта, Нур-Султан, Казахстан  
**Шведин** Борис Яковлевич, к.психол.н., член IAOA, ООО «Дан Роуз», Ростов-на-Дону, Россия

\* - members of the Russian Association of Artificial Intelligence - члены Российской ассоциации искусственного интеллекта - [http://www.raai.org/about/about.shtml?raai\\_list](http://www.raai.org/about/about.shtml?raai_list)

## Executive Editorial Board - Исполнительная редакция

Chief Editor	<b>P.O. Skobelev</b>	Samara, Russia	Главный редактор	Скобелев П.О.	директор НПК «Разумные решения»
Deputy Chief Editor	<b>S.V. Smirnov</b>	Samara, Russia	Зам. главного редактора	Смирнов С.В.	зам. директора ИПУСС РАН – СамНЦ РАН
Executive Editor	<b>N.M. Borgest</b>	Samara, Russia	Выпускающий редактор	Боргест Н.М.	директор издательства «Новая техника»
Editor	<b>D.M. Kozlov</b>	Samara, Russia	Редактор	Козлов Д.М.	доцент Самарского университета
Technical Editor	<b>D.N. Borgest</b>	Samara, Russia	Технический редактор	Боргест Д.Н.	специалист Самарского университета
Executive Secretary	<b>S.A. Vlasov</b>	Samara, Russia	Ответственный секретарь	Власов С.А.	аспирант Самарского университета

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. The journal has been successfully evaluated in the evaluation procedure for the ICI Journals Master List 2014-2019 and journal received the ICV (Index Copernicus Value).

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

Журнал включён в перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание учёных степеней кандидата и доктора наук (Письмо Департамента аттестации научных и научно-педагогических работников Минобрнауки РФ от 01.12.2015 № 13-6518) по научным специальностям 05.13.01, 05.13.17, 05.13.18 и 05.07.02, 05.07.05.

Журнал включен в список журналов, входящих в базу данных Russian Science Citation Index (RSCI) на платформе Web of Science. Пятилетний импакт-фактор РИНЦ **1,00** (2013), **0,93** (2014), **1,34** (2015), **1,07** (2016), **1,00** (2017), **1,17** (2018), **0,86** (2019).

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.)

## СОДЕРЖАНИЕ

### ОТ РЕДАКЦИИ

Оппозиция знания и незнания 141-143

### ОБЩИЕ ВОПРОСЫ ФОРМАЛИЗАЦИИ ПРОЕКТИРОВАНИЯ: ОНТОЛОГИЧЕСКИЕ АСПЕКТЫ

**А.Д. Редозубов** 144-153  
Формализация смысла. Часть 1

### ПРИКЛАДНЫЕ ОНТОЛОГИИ ПРОЕКТИРОВАНИЯ

**Г.Ю. Гуськов, А.М. Наместников, А.А. Романов, А.А. Филиппов** 154-169  
Формирование базы знаний для поддержки процесса архитектурного проектирования программных средств

**С.В. Кузнецов, Д.В. Кознов** 170-184  
Управление мастер-данными в рамках итеративного подхода

### ИНЖИНИРИНГ ОНТОЛОГИЙ

**Н.В. Максимов, А.А. Лебедев** 185-211  
Онтологическая система «знания-деятельность»

**А.В. Видня, Н.О. Дородных, А.Ю. Юрин** 212-226  
Подход к созданию онтологий на основе электронных таблиц с произвольной структурой

### МЕТОДЫ И ТЕХНОЛОГИИ ПРИНЯТИЯ РЕШЕНИЙ

**В.П. Офицеров, С.В. Смирнов** 227-238  
Оптимизация при ограничении числа проектных переменных

### Год науки и технологии в России - 2021

Премия «За верность науке». Заявка участника 239-245

Термины, определения и принципы классификации в новом стандарте 246-247

Рекомендуемые издания 2021 248

### Национальные стратегии в области искусственного интеллекта: европейская перспектива 2021

249-250

### Онтологический Саммит 2021

251-252

---

Журнал ориентирован на учёных и специалистов, работающих по научным направлениям: онтологические аспекты общих вопросов формализации проектирования, прикладные онтологии проектирования, инжиниринг онтологий, методы и технологии принятия решений.

Правила подготовки рукописей статей размещены на сайте журнала «Онтология проектирования»: [http://agora.guru.ru/scientific\\_journal/](http://agora.guru.ru/scientific_journal/).

#### Контакты учредителей

**ФИЦ Самарский научный центр РАН:** 443020, Самара, ул. Садовая, 61, тел./факс.: +7 (846) 333 27 70, Смирнов С.В., smimov@iccs.ru.

**Самарский университет:** 443086, Самара, Московское шоссе 34, корп. 10, тел.: +7 (846) 267 46 47, Боргест Н.М., borgest@yandex.ru.

**ООО «Новая техника» (издательство):** 443010, Самара, ул. Фрунзе, 145, тел.: +7 (846) 332 67 84, факс: +7 (846) 332 67 81.

---

Отпечатано в ООО «Новая техника», г. Самара, пр. К. Маркса, 24-76. Дата выхода 30.06.2021. Тираж 300 экз. Свободная цена. (6+).

---

## CONTENTS

### EDITORIAL

- Opposition of knowledge and ignorance 141-143

### GENERAL ISSUES OF FORMALIZATION IN THE DESIGNING: ONTOLOGICAL ASPECTS

- A.D. Redozubov** 144-153  
Formalization of the meaning. Part 1

### APPLIED ONTOLOGIES OF DESIGNING

- G.Y. Guskov, A.M. Namestnikov, A.A. Romanov, A.A. Filippov** 154-169  
Formation of a knowledge base to support the process of architectural design of software systems
- S.V. Kuznetsov, D.V. Koznov** 170-184  
Master data management in an iterative approach

### ONTOLOGY ENGINEERING

- N.V. Maksimov, A.A. Lebedev** 185-211  
Ontological system «knowledge-activity»
- A.V. Vidia, N.O. Dorodnykh, A.Yu. Yurin** 212-226  
An approach to creating freeform spreadsheet ontology

### METHODS AND TECHNOLOGIES OF DECISION MAKING

- V.P. Ofitserov, S.V. Smirnov** 227-238  
Optimization while limiting the number of design variables

### Year of Science and Technology in Russia - 2021

- Loyalty to Science Award. Participant application 239-245
- Terms, definitions and classification principles in the new standard 246-247
- Recommended edition 2021 248

- National strategies on Artificial Intelligence: A European perspective, 2021 edition** 249-250

- Ontology Summit 2021** 251-252

---

The journal is aimed at scientists and specialists working in the following research areas: ontological aspects of general issues of design formalization, applied design ontologies, ontology engineering, methods and technologies of decision making.

The current version of the Rules for the preparation of manuscripts of articles for the journal «Ontology of Designing» is on the journal website: [http://agora.guru.ru/scientific\\_journal/](http://agora.guru.ru/scientific_journal/).

### Contacts of the Founders

**FIC Samara Scientific Center of the RAS:** 61, Sadovaya st., Samara, 443020, Russia. Tel.: +7 (846) 333 27 70, S.V. Smirnov, smirnov@iccs.ru  
**Samara University:** 34, Moskovskoye shosse, bldg. 10, Samara, 443086, Russia. Tel.: +7 (846) 267 46 47, N.M. Borgest, borgest@yandex.ru  
**New Engineering LLC (publishing house):** 145, Frunze st., Samara, 443010, Russia. Tel.: +7 (846) 332 67 84, fax: +7 (846) 332 67 81



## ОТ РЕДАКЦИИ

## Оппозиция знания и незнания Opposition of knowledge and ignorance

«... Мы на пороге массового вымирания, а всё, о чём вы можете говорить, это деньги и сказки о вечном экономическом росте. Как вы смеете!»

*Грета Тунберг, Климатический саммит ООН в 2019 г.*

«Кто владеет информацией, тот владеет миром»

*Натан Ротшильд, 1815 г.*

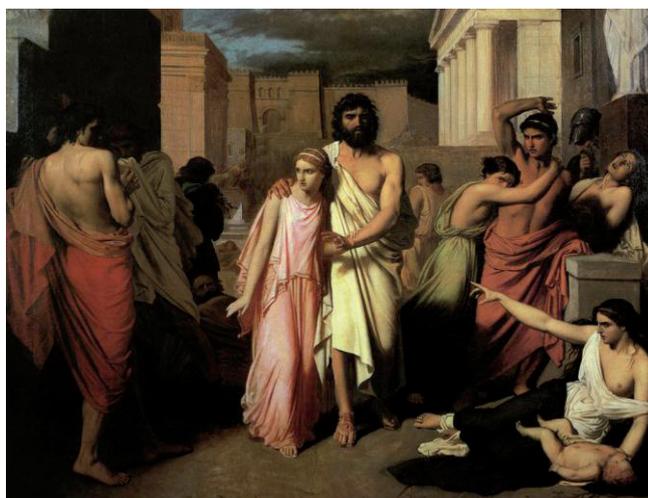
«О знание, знание! Тяжкая обуза, когда во вред ты знающим дано!»

*Тиресий, Трагедия Софокла «Царь Эдип», 467 г. до н.э.*

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

Колесо исторического процесса набирает обороты. Цивилизационные изменения по энтропийному закону наполняются сложностью и непредсказуемостью. Информационный «перебор» зашкаливает: информация и данные, позиция и оппозиция, знания и незнания смешались в одном информационном котле. При этом используемые и внедряемые технологии (за редким исключением) не улучшают состояние «нашего дома», той природной среды, в которой сформировалась жизнь. На пике взлетевших, в буквальном смысле, массовых коммуникаций жителей нашей маленькой планеты, когда миллионы людей, сменяя друг друга, находятся в воздушном пространстве, единицы бороздят космические просторы, а десятки, а то и сотни миллионов перемещаются по планете ежеминутно на созданных ими артефактах – все это, включая производство необходимой энергии и бесчисленных предприятий, перерабатывающих природное сырьё и производящих в конечном итоге мусор, привело к ситуации, способной изменить нашу жизнь на планете... О трагедии изменения климата<sup>1</sup>, предвестником которой может служить в том числе и небывало долгая жара в мае-июне этого года в России, всё активнее говорят и пишут не только в научных журналах.

«О знание, знание! Тяжкая обуза, когда во вред ты знающим дано!», – сетует слепой пророк Тиресий в драме Софокла «Царь Эдип»<sup>2</sup>. Эдип позвал его, чтобы узнать причину эпидемии и экологической катастрофы, обрушившихся на Фивы. Но Тиресий знал, что царь отвергнет правду. Сегодня эта ситуация хорошо знакома многим. Подобно Тиресию, современные учёные знают, куда движется наша планета и почему, но не с помощью пророчеств, как в



*Антигона выводит слепого Эдипа из Фив,  
Жалабер, XIX в.*

<sup>1</sup> Bryan Doerries. The Tragedy of Climate Change. May 18, 2021. - <https://www.project-syndicate.org/commentary/climate-change-pandemic-greek-tragedy-scientific-communication-by-bryan-doerries-2021-05>.

<sup>2</sup> Софокл. Драмы. В пер. Ф.Ф. Зелинского, под ред. М.Л. Гаспарова и В.Н. Ярхо. Издание подготовили М.Л. Гаспаров и В.Н. Ярхо. Серия "Литературные памятники". М., Наука, 1990. (Царь Эдип - [http://lib.ru/POEEAST/SOFOKL/sofokl3\\_1.txt](http://lib.ru/POEEAST/SOFOKL/sofokl3_1.txt)).

греческой трагедии, а благодаря многочисленным экспериментам, испытаниям и тщательно-му анализу. Однако «земное пророчество» учёных неспособно преодолеть безразличие общества, не «чувствующего» на себе скорость происходящих изменений и живущего надеждой на их «волшебный» благоприятный исход. Исторический опыт и литературные примеры не позволяют передать всю тяжесть и сложность разворачивающихся на наших глазах глобальных изменений. Греческие трагедии – это истории людей, которые, казалось бы, начинают понимать, но уже слишком поздно, ведь ошибку с катастрофическими последствиями они уже совершили.

Премьера «Царя Эдипа» состоялась весной 429 года до нашей эры в период между первой и второй волнами чумы, убившей почти треть населения Афин. Греческое общество пыталось понять, насколько неизбежными были понесённые потери, стремилось «выучить» урок из истории о надменных лидерах и упрямой слепоте. В трагедии Эдип — символ той опасности, которую несёт власть и которая кроется в любой политической системе<sup>3</sup>.

«Царь Эдип» — это трагедия важнейшей темы: *знания и незнания*. С одной стороны, Эдип — мудрец, который спас Фивы от Сфинкс, разгадав её загадку. Именно как к мудрецу к нему приходит хор фиванских граждан с просьбой спасти город, и как мудрец Эдип заявляет о необходимости разгадать загадку убийства прежнего царя и разгадывает её. С другой стороны, он и слепец, не знающий самого важного: кто он, кто его отец и мать. В стремлении узнать истину он игнорирует всё, о чём его предупреждают окружающие. Получается, что он мудрец, который не мудр<sup>4</sup>.

*Оппозиция знания и незнания* — это одновременно и оппозиция видения и слепоты. Слепой пророк Тиресий, который разговаривает с видящим Эдипом, все время говорит ему: «Ты слеп». Эдип в этот момент видит, но не знает — в отличие от Тиресия, который знает, но не видит.

По-гречески знать и видеть (οἶδα [oida]), знание и видение — это одно и то же слово (как и в английском *see* – видеть и понимать – одно и то же слово). Это тот же корень, который, с точки зрения греков, заключён в имени Эдипа, и в трагедии это многократно обыгрывается. В конце трагедии, узнав, что он убил своего отца и женился на своей матери, Эдип ослепляет себя, и теряет зрение, став мудрецом.

Трагедия построена на тонкой игре двух тем — знания и видения. Внутри трагедии они часто меняются местами. Смысл трагедии двойственный. С одной стороны, Эдип — самый несчастный человек, с другой стороны, Эдип торжествует в финале трагедии. Он хотел узнать, кто его отец и кто его мать, и узнал. Он хотел узнать, кто убил пастуха Лаия, — и узнал. Он хотел спасти город от чумы и мора — и спас. Эдип обрёл самое важное для него — знание, ценой невероятных страданий, ценой потери собственного зрения. В этой двойственности — литературный и политический смысл трагедии, поскольку она демонстрирует двусторонность власти, связанность власти и знания. Царь Эдип воплотил основной смысл и посыл греческой трагедии - это интеллектуальный опыт, который соотнесён с религиозным,

<sup>3</sup> См. также *Николай Гринцер*. Главная греческая трагедия. 24.05.2016. <https://www.youtube.com/watch?v=-VXC7bDrqZk>, О главной греческой трагедии: <https://cameralabs-org.turbopages.org/cameralabs.org/s/10057-o-glavnoj-grechskoj-tragedii-kem-byt-tsar-edip-i-chto-s-nim-proizoshlo-na-samom-dele>.

<sup>4</sup> В современной истории видим: как не мудры и нынешние политики, устроившие «борьбу» с пандемией; как растёт недоверие к разработанным вакцинам; отсутствует чёткая аргументированная система диагностики вируса и правил жизни в урбанизированном обществе в период пандемии; выдаваемые решения и меры непоследовательны и меняются на глазах. Удивительно, но практически во всех развитых странах наблюдается подобная картина: строгие требования к ношению масок и перчаток сочетаются с проведением массовых мероприятий и различными запретами. В 21-ом технологически развитом веке ставится вопрос о поголовной вакцинации без индивидуального анализа и какой-либо диагностики, без замера наличия антител в организме. Очевидно, что сначала необходимо выяснить путём лабораторного анализа конкретного организма потребность в вакцинации, а потом предлагать её... См. также редакционную статью «Онтология пандемии: реальная и мнимая» в нашем журнале - [https://www.ontology-of-designing.ru/article/2020\\_2\(36\)/1\\_Editorial.pdf](https://www.ontology-of-designing.ru/article/2020_2(36)/1_Editorial.pdf).

литературным и политическим опытом. И чем теснее эти разные смыслы взаимодействуют друг с другом, тем важнее смысл трагедии и сильнее её эффект.

Совсем иной смысл использования знания и незнания был у известного банкира Натана Ротшильда<sup>5</sup>. Известность он приобрёл благодаря фразе: «Кто владеет информацией, тот владеет миром». Банкир воспользовался тем, что, обладая важной информацией и манипулируя ею в своих интересах, смог заработать свой капитал на тех, кто не имел к ней доступа. Реальная информация, ценная раньше других о поражении Наполеона в битве при Ватерлоо в 1815 году, позволила Ротшильдам вести беспроигрышную игру на Лондонской бирже. Впоследствии многие держатели ценных бумаг покончили с собой, а Натан Ротшильд за один день овладел большой долей британской экономики. Такую же операцию на Парижской бирже осуществил и его брат - Якоб Ротшильд...



Натан Ротшильд  
(1777-1836)

В этом номере журнала статьи авторов из Иркутска, Москвы, Самары, Санкт-Петербурга и Ульяновска. Открывает номер статья, которая рассматривается редакцией как призыв к разработке идей формализации смысла и готовность журнала обсуждать эту тему. Эта острая тема для разработчиков сильного искусственного интеллекта. И онтологам, стремящимся докопаться до сути вещей, здесь есть огромное поле деятельности.

В разделе прикладные онтологии проектирования представлены статьи из Ульяновска (о формировании базы знаний для поддержки процесса архитектурного проектирования программных систем) и из Санкт-Петербурга (об управлении мастер-данными в рамках предлагаемого итеративного подхода). По оценкам авторов этих статей, приложения результатов публикуемых работ во многом инвариантны к какой-либо предметной области, т.к. касаются управления программными системами и данными.

В разделе инжиниринг онтологий - статьи из Москвы (об онтологической системе «знания - деятельность»), о подходе к построению онтологического описания рациональной деятельности) и из Иркутска (о подходе к созданию онтологий на основе электронных таблиц с произвольной структурой).

В разделе методы и технологии принятия решений - статья авторов из Москвы и Самары об оптимизации при ограничении числа проектных переменных.

**2021** год - *год науки и технологий в России*. Наш научный журнал уже **10** лет вносит свою лепту в развитие науки о проектировании, подготовив **40** номеров журнала «Онтология проектирования» и проведя **20** заседаний одноимённого семинара. В этом году ядро нашей исполнительной редакции приняло решение подать заявку на Всероссийскую премию «За верность науке» в номинации «Лучшее периодическое издание о науке», снабдив её краткими дополнениями-пояснениями (см. в конце этого номера).

Оптимальное проектирование будущего – наша общая цель. Поэтому хочется верить, что разумные решения реализуются в благоприятные для человечества сценарии развития.

*Dum spiro, spero!*

*Ontologists and designers of all countries and subject areas, join us!*

<sup>5</sup> [https://ru.wikipedia.org/wiki/Ротшильд,\\_Натан\\_Майер](https://ru.wikipedia.org/wiki/Ротшильд,_Натан_Майер)

## ОБЩИЕ ВОПРОСЫ ФОРМАЛИЗАЦИИ ПРОЕКТИРОВАНИЯ: ОНТОЛОГИЧЕСКИЕ АСПЕКТЫ

УДК 004.896

DOI: 10.18287/2223-9537-2021-11-2-144-153

### Формализация смысла<sup>1</sup>. Часть 1

А.Д. Редозубов

Фонд имени академика Натальи Бехтеревой, Санкт-Петербург, Россия

#### Аннотация

Всякая область знаний строит свои описания, используя понятия. При этом распространено определение понятий через задание свойственных им признаков. На этом базисе строятся как основные математические, так и многие философские концепции. Понятия, которыми оперирует человек, подчиняются аналогичным свойствам и их природа – это природа определений. Многочисленные попытки создать сильный искусственный интеллект основаны на соответствующей парадигме. В статье делается попытка обосновать необходимость использования контекстно-смысловой парадигмы для объяснения работы естественного мозга и для создания сильного искусственного интеллекта. Приводится формальная модель, описывающая смысл, и предлагается способ представления её связи с известными данными о функционировании мозга. Показывается, что вокруг каждого понятия может быть создан контекст, который может являться носителем смысла понятия. Контекст позволяет отойти от использования набора признаков для узнавания явления, стоящего за понятием. Контекст оказывается точкой зрения, связанной с понятием, в которой меняется описание окружающего мира. Зная правила этих изменений, можно не только моделировать различные точки зрения, но и определять, какие из них создают адекватные трактовки. При этом наличие адекватной трактовки в контексте явления служит критерием присутствия этого явления.

**Ключевые слова:** понятие, смысл, контекст, мозг, искусственный интеллект, сильный искусственный интеллект.

**Цитирование:** Редозубов, А.Д. Формализация смысла. Часть 1 / А.Д. Редозубов // Онтология проектирования. – 2021. – Т. 11, №2(40). – С.144-153. – DOI: 10.18287/2223-9537-2021-11-2-144-153.

Как сердцу высказать себя?  
Другому как понять тебя?  
Поймёт ли он, чем ты живёшь?  
**Мысль изречённая есть ложь.**

Федор Тютчев, Silentium

<sup>1</sup> Тема статьи архи актуальна в цифровом, информационном мире и тесно связана с искусственным интеллектом (ИИ), с его моделями, онтологиями предметных областей и задачами, в них решаемыми. Мечта исследователей, занятых созданием сильного ИИ, привлекает огромное количество специалистов из разных областей, стремящихся переложить бремя принятия решений на создаваемые «машины». Биологи, физиологи, психологи, математики, специалисты по компьютерным наукам активно вовлечены в процесс создания сильного ИИ. При этом сложность изучаемой проблемы затрудняет её обсуждение учёными из разных предметных областей. За дело берутся специалисты, способные в доступной форме изложить накопленный материал, обобщить его, найти возможность преодолеть стыковые барьеры наук. Автор статьи известен как популяризатор наук о мозге, написавший ряд книг о работе мозга: «Мозг напрокат. Как работает человеческое мышление и как создать душу для компьютера», 2010; «Логика эмоций», 2012; «Логика мышления», 2014. Статья подготовлена по итогам научного семинара «Онтология проектирования», на котором автор выступил с докладом «Формализация смысла: понятия, принципы, модели, примеры». Семинар прошёл в онлайн режиме 25 мая 2021 года. Запись семинара доступна по ссылке <https://bbb.ssau.ru/playback/presentation/2.0/playback.html?meetingId=8fcb5765c6784570796bb5894aae0dec6a08db06-1621939322257>. Редакция выражает надежду на развитие дискуссии и поиск формализмов, способных фиксировать и транслировать смысл информации, искать формальные и иные решения в отделении правдивой и достоверной информации от «фэйков», ставшие бичом начала 21-го столетия. Смелый шаг автора статьи в область формализации смысла и дальнейшей его передачи стоит рассматривать как *призыв к развитию идей* в этой области и поиску путей их реализации. *Прим. ред.*

## Введение. Искусственный интеллект. Тест Тьюринга

Для описания интеллектуальных систем используют два термина — «искусственный интеллект» (ИИ) и «сильный искусственный интеллект» (СИИ) [1, 2]. В английской традиции — *Artificial intelligence (AI)* и *Artificial general intelligence (AGI)*. Первый подразумевает любую деятельность компьютера, имитирующую человеческий интеллект, второй — только такую, которая претендует на что-то универсально общее, похожее на то, как мыслит человек.

В английском названии СИИ не случайно называется общим или универсальным. Суть этого названия в том, что СИИ должен уметь решать не одну конкретную задачу и даже не широкий набор задач, он должен быть «человекоподобным». То есть обладать способностью, начав с нуля, вникнуть в любую область и начать ориентироваться в ней подобно тому, как это способен делать человек. И желательно не хуже человека.

Толкование СИИ можно дать через хорошо известный тест Тьюринга [1]: «Человек взаимодействует с одним компьютером и одним человеком. На основании ответов на вопросы он должен определить, с кем он разговаривает: с человеком или компьютерной программой. Задача компьютерной программы — ввести человека в заблуждение, заставив сделать неверный выбор».

В своей статье «Вычислительные машины и разум» [1], описывая этот тест, Тьюринг задавался вопросом: «Зафиксируем наше внимание на цифровом компьютере А. Действительно ли, изменяя компьютер так, чтобы иметь достаточный объём памяти, что равнозначно увеличению скорости его действий, и обеспечивая его подходящей программой, компьютер А можно сделать таким, чтобы он удовлетворительно играл и роль компьютера А в игре-имитации, и роль человека В?» (см. рисунок 1).

Говоря о СИИ, как о компьютере, способном пройти тест Тьюринга, возникают два главных вопроса. Первый – это вопрос, заданный Тьюрингом: можно ли за счёт увеличения мощности смоделировать на компьютере человеческий интеллект? Второй: будут ли какие-то особые требования к программе, реализующей СИИ?

Второй вопрос требует пояснения. Дело в том, что сегодня функционируют системы ИИ, которые показывают замечательные результаты в различных областях и которые не претендуют на роль СИИ. Можно ли ожидать, что эволюционное развитие этих систем, связанное с увеличением объёмов обрабатываемых данных и ростом компьютерных мощностей, приведёт к созданию СИИ, или для этого потребуется что-то новое в методологии алгоритмов? Чтобы разобраться в этом вопросе, необходимо рассмотреть базовый уровень представления знаний, с которого начинается построение любого алгоритма.

### 1 Понятия

Чтобы описывать окружающий мир, человеческий мозг в результате эволюции приобрёл возможность формировать внутренние понятия. На таком представлении основана, например, когнитивная психология [3]. Главным доводом в пользу такого представления может служить то, что понятия в форме слов проявляются в человеческой речи, которая, в свою очередь, отражает внутренние процессы мозга.

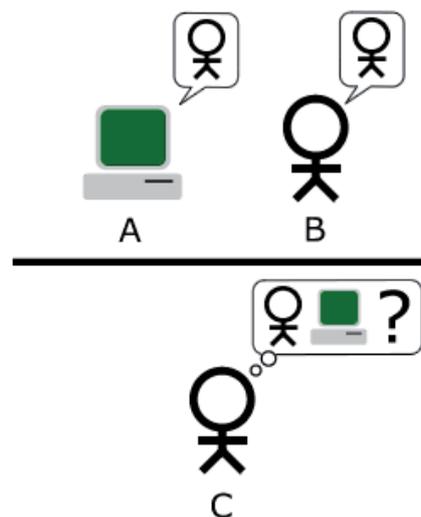


Рисунок 1 - Тест Тьюринга

Формирование внутренних понятий происходит под воздействием опыта, который человек получает из окружающего мира. Наличие в окружающем мире определённой структуры приводит к формированию понятий, способных адекватно эту структуру описать.

В какой-то момент эволюции у человека появилось новое средство общения – язык [4]. Внутренним понятиям были сопоставлены слова. Устные слова — это просто звуки, они не несут смысла сами по себе. Но будучи связанными с понятиями, передают их значения. То есть за содержание отвечает внутреннее понятие, а слово — это знак понятия. При этом надо учитывать, что далеко не всем внутренним понятиям соответствуют какие-либо слова.

Язык развивался, и в нем появлялись новые слова. Кроме обозначения того, что непосредственно видно или слышно, в языке стали появляться слова, обозначающие некие сложные, порой абстрактные явления. Люди открывали для себя эти явления и через язык давали им названия. Например, так появились слова: «понятие», «логика», «определение» и т.п.

Для появления внутренних понятий, в принципе, язык не нужен. Они сформируются и без него. Классический пример – дети, подобные Маугли. Но с момента своего появления язык стал влиять на то, какая именно система внутренних понятий формируется у людей. Образование понятий напоминает рост виноградных лоз. Для лозы направляющая служит опорой и задает направление ее роста. Сколько есть направляющих, столько лоз их обовьет. Для формирования внутренних понятий слова языка во многом служат такими направляющими. Изобретенные когда-то слова, обозначающие некие абстракции, и сегодня позволяют сформироваться у нас понятиям, которые не возникли бы сами по себе без участия языка.

Характерное свойство слов – их многозначность. В зависимости от обстоятельств их применения они могут передавать разные значения. Обстоятельства принято называть контекстом. Текущие значения слов – их трактовками.

В случаях, когда в некой предметной области за словом закрепляется фиксированное значение, такое слово принято называть термином.

## 2 Имплицитность понятий

Важным свойством внутренних понятий является имплицитность их формирования. То есть, формируясь в результате опыта, они выделяются мозгом, но при этом никак не «сообщают» человеку принципы, по которым они сформировались. Так, человек может успешно пользоваться словами языка, будучи при этом не в состоянии объяснить значения многих слов. Знания о значениях понятий скрыты от человека. Но он может судить об этих значениях, наблюдая за тем, как он сам и другие люди используют эти понятия. Попытка наблюдения и объяснения знаний называется их экспликацией.

Убедительно продемонстрировал неосознанное обучение Джордж Миллер [5]. Испытуемым предлагались последовательности букв, образованные по определенным сложным правилам. Затем испытуемые должны были отличить правильные буквенные цепочки от неправильных. Оказалось, что, даже не зная правил, испытуемые успешно справлялись с задачей. Артур Ребер изменил условия, в его опытах испытуемые вообще не знали о наличии правил [6].

Но это не помешало испытуемым выделить последовательности, подчиняющиеся этим правилам. Примечательно, что, когда же была поставлена задача найти правила, ни один из испытуемых не смог этого сделать. Более того, попытка найти правила привела к ухудшению способности дифференцировать грамматические и неграмматические цепочки. Аналогичные результаты показали Дональд Бродбент и Диана Берри на примере задач, оформленных в форме компьютерных игр [7].

Наиболее ярко соотношение имплицитных и эксплицитных знаний было показано у Сократа [8]. Сократ убеждал, что истинные значения понятий всегда скрыты. Что в попытке объяснить эти значения люди дают поверхностные объяснения, основанные на наблюдаемых признаках, и эти объяснения всегда ошибочны. Чтобы понять, что стоит за понятиями, Сократ разработал специальный метод. Метод состоял из двух частей. В каждой из них происходил диалог «человека» и «простака». Сократ полагал, что всякое рассказанное знание бесполезно. И только знание выстраданное чего-то стоит.

Принято считать, что метод Сократа противопоставлялся приемам софистов. Если софисты полагали, что важно с помощью подбора доводов убедить слушателей в своей точке зрения, то Сократ предлагал рассматривать все доводы, включая неудобные, с разных точек зрения и полагать истинным то объяснение и ту точку зрения, в которых все доводы оказывались непротиворечивы.

### 3 Классическая парадигма понятий

Вопрос о внутренней природе понятий всегда был ключевым вопросом для философии, математики и естественных наук. Ответ на этот вопрос многим казался столь очевидным, что воспринимался на уровне аксиом. Этот «очевидный» ответ лёг в основу «классического» подхода, который во многом определил здание современной науки. Такое «классическое» толкование можно встретить, например, в [9]: «Понятие — отображённое в мышлении единство существенных свойств и отношений предметов; мысль, выделяющая и обобщающая предметы некоторого класса по общим и в своей совокупности специфическим для них признакам».

В приведённом толковании изначально заложен алгоритм, который объясняет суть понятий. Этот алгоритм подразумевает, что неоднородность внешнего мира позволяет выделять в нём явления. Первично человек имеет дело с некой сенсорной информацией, которая является для него набором элементарных доступных явлений. Явления, которые человек способен детектировать, называются в этом алгоритме признаками.

Наблюдая за проявлением известных ему признаков, человек выделяет новые явления. Эти явления есть следствия воздействия окружающего мира на человека и являются отражением неких присущих миру сущностей. Выделенные новые явления сами становятся признаками, которые позволяют находить более сложные явления.

Каждое выделенное явление приводит к формированию соответствующего понятия. При этом есть разница между понятием и явлением. Неоднородность окружающего мира приводит к появлению механизма, детектирующего некую особенность его структуры. Так для нас появляется явление. Этому *детектору* присваивается соответствующий знак. Совокупность *детектора явления* и знака создают понятие. В случае человеческого мозга знаком выступает некий нейронный код, в речи знаком служит устное слово, в математике - символьная запись и т.п. По последовательности формирования понятий между ними можно установить определённую иерархию.

Вопрос о значении понятия – это вопрос о функционировании детектора. Классический подход аксиоматически вводит алгоритм работы детектора, основанный на сравнении признаков, считает его неотъемлемой частью понятия и подразумевает его единственно возможным. Согласно классическому подходу явления описываются наборами характеризующих их признаков. Считается, что можно выделить некие главные признаки, которые должны всегда присутствовать в явлении, и вторичные, которые иногда могут отсутствовать, но при этом своим присутствием повышают вероятность правильного детектирования.

Такая конструкция приводит к модели, в которой явлению сопоставляется некое признаковое описание. Часто такое описание задаётся либо перечислением признаков, либо вектором, элементы которого указывают на выраженность определённых признаков. Векторное представление особенно удобно, поскольку легко позволяет ввести метрику. Например, часто используется расстояние Махаланобиса [10], которое является евклидовым расстоянием, скорректированным, исходя из дисперсий признаков и их взаимных ковариаций.

Расстояние Махаланобиса от многомерного вектора  $x = (x_1, x_2, x_3, \dots, x_n)^T$  до множества со средним значением  $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_n)^T$  и матрицей ковариации  $S$  определяется:

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)}$$

Наличие подобной метрики позволяет рассматривать понятие как класс, для которого можно задать прототип класса, то есть его условный центр  $\mu$ . По расстоянию, учитывающему ковариации признаков, можно судить о вероятности отнесения нового события  $x$ , описываемого его набором признаков, к советуемому классу, а значит и к понятию.

Попытка описать понятия, как классы, сталкивается с тем, что само отнесение к классам исходит из того значения, которое человек вкладывает в то или иное понятие. В результате в пространстве признаков классы могут иметь сложную форму, сильно отличную от сфероида. В этом случае используются более сложные методы, суть которых сводится к попыткам описать некие границы, задающие форму явления в пространстве описывающих его признаков.

Такой подход лежит в основе стандартной постановки задачи классификации, где признаковые описания объектов принято отождествлять с самими объектами.

Подобным образом решаются задачи кластеризации, которая имеет дело либо с признаковыми описаниями и заданной метрикой, либо с матрицами расстояний или близости объектов.

Этот же подход используется в нейронных сетях, где каждый нейрон в процессе обучения за счёт настройки своих весов становится детектором определённого свойства. Процесс обучения во многом напоминает процесс создания границ, описывающих те понятия, что требуется детектировать нейронной сети.

Полагая, что искусственные нейроны, описываемые моделью Маккаллока — Питтса [11], похожи на реальные биологические нейроны, представление о работе искусственных нейронных сетей часто используют для попыток объяснить работу естественного мозга. При этом по умолчанию переносят классическое представление о понятиях на понятия, формируемые мозгом. Несмотря на то, что современные представления о работе естественных нейронов далеки от модели искусственного нейрона, подобного рода объяснения остаются крайне распространёнными. Ярче всего они проявляются в представлениях о гипотетических «нейронах бабушки» [12] - нейронах, избирательно реагирующих на определённый образ.

Некоторые сложности в классическом подходе возникают в связи многозначностью естественных понятий, когда в зависимости от контекста одни те же понятия начинают приобретать разные значения, то есть получают разную трактовку. Для объяснения изменчивости понятий в контекстах в классическом подходе используется анализ формальных понятий (АФП) [13].

Контекстом в АФП называют тройку  $K = (G, M, I)$ , где  $G$  — множество объектов,  $M$  — множество признаков, а отношение  $I \subseteq G \times M$  говорит о том, какие объекты какими признаками обладают. Внутри контекста каждому объекту сопоставляется его признаковое описание. При переходе к другому контексту признаковое описание меняется на новое.

Для произвольных  $A \subseteq G$  и  $B \subseteq M$  определены операторы Галуа:

$$A' = \{m \in M \mid \forall g \in A (g \mid m)\},$$

$$B' = \{g \in G \mid \forall m \in B (g \mid m)\}.$$

Первый оператор описывает множество всех признаков  $A'$ , общих для всех объектов множества  $A$ . Второй - множество всех объектов  $B'$ , содержащих все признаки множества  $B$ . Пара множеств  $(A, B)$ , таких, что  $A \subseteq G$ ,  $B \subseteq M$ ,  $A' = B$  и  $B' = A$ , называется формальным понятием контекста  $K$ . Множество  $A$  при этом называют объёмом, а множество  $B$  - содержанием формального понятия  $(A, B)$ .

Идея контекстов используется в АФП как попытка устранить вариативность проявления реальных понятий. Введение множества контекстов позволяет внутри каждого контекста отождествить объект с его признаковым описанием и получить для каждого из понятий неизменные в рамках заданного контекста объём и содержание.

Обобщая главную идею классического подхода, можно сказать, что он сводится к попыткам тем или иным способом дать понятиям определения, основанные на перечислении характерных для них признаков.

#### 4 Комбинаторный взрыв классических понятий

Классический подход успешно применяется для решения многих задач. Нейронные сети в ряде случаев справляются с интеллектуальными задачами лучше человека. Но при всех успехах этого подхода ему присущ неустранимый изъян. Попытка задать понятия через их определения в задачах, связанных с реальным миром, всегда приводит к невозможности описать их так, чтобы они соответствовали тем понятиям, которые способен сформировать человек.

На первый взгляд это не кажется большой проблемой. Формируя свою систему понятий, ИИ так же, как и человек, получает возможность отражать картину окружающего мира и оперировать ей. Но практика показывает, что понятия, заданные через определения, не только отличаются от «естественных» понятий, но и содержат неизбежные ошибки в своих определениях. Можно сказать, что любое определение, задающее границу понятия, является неизбежной ошибкой<sup>2</sup>. То есть, когда происходит попытка описать некое явление реального мира, создавая ему определение, результат всегда получается неточен. Чтобы уточнить его приходится использовать больше признаков и создавать больше контекстов. Но в случае реального мира это всегда приводит к комбинаторному взрыву. На практике это выглядит так, что достаточно быстро удаётся получить некую «инженерную» точность, однако попытки улучшить результат ведут к непропорционально быстрому росту требований к ресурсам и объёмам обучающих данных. За счёт экспоненциального, а порою факториального роста сложности это делает невозможным получение надёжного результата.

С одной стороны, неизбежность комбинаторного взрыва очевидна. Но, с другой стороны, у многих исследователей складывается ощущение, что, повышая вычислительные мощности, рано или поздно удастся добиться в задачах, требующих описания реального мира, результатов, сопоставимых со способностями человека. Истоки этой веры кроются в убеждении, что человеческий мозг использует для формирования своих понятий ту же классическую парадигму, основанную на определениях. А значит, сравнившись в мощности с мозгом, ИИ удастся сравняться по способностям с человеком. По сути, вопрос, заданный в начале статьи относительно возможности создания СИИ на базе классической парадигмы, упирается в «веру», что мозг использует именно эту парадигму.

<sup>2</sup> См. заключительную фразу в эпиграфе к статье. *Прим. ред.*

## 5 СМЫСЛЫ ПОНЯТИЙ

Попытки описать природу понятий рано или поздно приводят к разговору о смысле. Смысл понятия воспринимается как нечто, отвечающее за его содержание. При этом традиционно содержание противопоставляется форме.

Для смысла, содержания и для формы существует множество философских определений. Но изначально — это слова естественного языка, относительно которых у каждого есть свое имплицитное понимание. Всякое определение исходит из некой модели понимания этих явлений, которая может отличаться от имплицитного знания. Если же исходная суть понятия неверно или понятия не до конца, то определение приводит к замене исходного явления на конструкцию, задаваемую определением. И тогда последующая формализация хотя и может быть интересна, но она относится уже не к исходному явлению.

Можно попытаться дать общее представление о природе смысла. Пусть есть нечто, наблюдаемое, но неузнанное наблюдателем. Тогда можно изменить точку зрения и посмотреть на этот предмет с другой стороны. Если исследуемый предмет в принципе знаком, то должна найтись такая точка зрения, с которой его можно узнать. Важно, что в этот момент кроме предмета обнаруживается ещё и точка зрения, в которой сложилась ранее знакомая картина.

Можно допустить, что любое понятие может быть точкой зрения. За всяким понятием стоит явление, которое меняет восприятие мира человеком. Одно и то же видится по-разному в зависимости от того, через призму какого явления сейчас смотрит наблюдатель [14].

«Рассмотрение с разных сторон» в буквальном подходе требует перемещения наблюдателя. Но это же перемещение может быть реализовано и виртуально. Наблюдая за явлением, можно попытаться определить правила, по которым текущее описание заменяется на описание с новой точки зрения.

Точку зрения, соответствующую явлению, можно назвать контекстом явления. Новое описание, возникающее при рассмотрении исходной картины в контексте явления, можно назвать *трактовкой*. Правила, по которым текущее описание переходит в трактовку в контексте явления, — это правила трактовки. Знания о том, что уже знакомо человеку — это память предыдущих событий, или событийная память.

В такой терминологии получается, что смысл понятия можно задать, создав контекст, связанный с этим понятием. Для этого же потребуется сформировать правила трактовки и наполнить событийную память.

Когда контекст создан, он может служить детектором понятия. Входная информация получает трактовку в контексте понятия. И если оказывается, что полученная трактовка похожа на что-то уже содержащееся в памяти, то можно говорить о присутствии и самого явления, соответствующего контексту.

### Заключение

Сделана попытка описать понятия через соответствующие им контексты и приблизиться к созданию контекстно-смыслового подхода, который может позволить по-новому взглянуть на работу человеческого мозга и перспективы создания СИИ.

В последующих частях планируется предложить формализацию<sup>3</sup> этого подхода и показать, что существующие методы обработки данных могут быть описаны как частные случаи, реализующие ограниченный функционал работы со смыслом.

---

<sup>3</sup> Редакция решила добавить «масло в огонь» дискуссии отметив, что в Словаре Кембриджа нет определения для термина *формализация* (*formalization isn't in the Cambridge Dictionary yet. You can help!*) и в нём приводятся лишь примеры употреб-

## СПИСОК ИСТОЧНИКОВ

- [1] *Turing, A.* Computing Machinery and Intelligence / A. Turing // *Mind*, vol.LIX, no.236. October 1950. P.433-460.
- [2] *Redozubov, A.* The meaning of things as a concept in a strong AI architecture / A. Redozubov, D. Klepikov // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, 12177 LNAI. P.290-300.
- [3] *Фаликман, М.В.* Когнитивная психология: история и современность / М.В. Фаликман, В. Спиридонов // – М.: Ломоносовъ, 2011.- 384 с.
- [4] *Corballis, M.C.* From hand to mouth. The origins of language / M.C. Corballis // Princeton University Press. 2002. – 257 p.
- [5] *Miller, A.G.* Project Gramarama / A.G. Miller // New York, NY: Basic Books., 1967.
- [6] *Reber, A.S.* Implicit learning of artificial grammars / A.S. Reber // *Journal of Verbal Learning and Verbal Behavior*, 1967, 6. P.855-863.
- [7] *Berry, D.C.* The combination of explicit and implicit learning processes in task control / D.C. Berry, D.E. Broadbent // *Psychol. Res.* 1987. 49. P.7-15.
- [8] *Платон.* Теэтет / Платон // Перевод с греч., и прим. В. Сerezникова М.-Л.: СОЦЭКГИЗ, 1936. - 192 с.
- [9] Понятие. Словари и энциклопедии на Академикe. - <https://dic.academic.ru/dic.nsf/ruwiki/16579>.
- [10] *Mahalanobis, P.C.* On the generalised distance in statistics / P.C. Mahalanobis // *Proceedings of the National Institute of Sciences of India.* 1936. V.2, No 1. P.49-55.
- [11] *McCulloch, W.S.* A logical Calculus of Ideas Immanent in Nervous Activity / W.S. McCulloch, W. Pitts // *Bulletin of Mathematical Biophysics.* 1943. 5. P.115-133.
- [12] *Gross, C.* Genealogy of the Grandmother Cell / C. Gross // *Neuroscientist.* 2002. 8 (5). P.512-518.,
- [13] *Wille, R.* Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies / R. Wille // *Formal Concept Analysis. Foundations and Applications.* Springer Science & Business Media, 2005.
- [14] *Redozubov, A.* Holographic Memory: A Novel Model of Information Processing by Neuronal Microcircuits / A. Redozubov // In: Opris I., Casanova M. (eds) *The Physics of the Mind and Brain Disorders.* Springer Series in Cognitive and Neural Systems, vol 11. Springer, 2017. Cham. [https://doi.org/10.1007/978-3-319-29674-6\\_13](https://doi.org/10.1007/978-3-319-29674-6_13).

## Сведения об авторе



*Редозубов Алексей Дмитриевич*, 1968 г. рождения. Прикладной математик. Учился в Ленинградском политехническом институте прикладной математике. Длительное время занимался изучением механизмов восприятия художественных произведений, математическим анализом явлений и построением объясняющих моделей. Соучредитель Фонда поддержки научных исследований механизмов работы мозга, лечения его заболеваний, нейромоделирования имени академика Натальи Петровны Бехтеревой. Область интересов – создание сильного искусственного интеллекта, основанного на нейрофизиологии. Author ID (Scopus): 57218267840. [galdrd@gmail.com](mailto:galdrd@gmail.com).

Поступила в редакцию 07.06.2021, после рецензирования 19.06.21. Принята к публикации 22.06.2021.

ления этого термина (<https://dictionary.cambridge.org/dictionary/english/formalization>), что соответствует позиции автора об ограниченности подхода, связанного с определением понятий через набор свойств (признаков).

При этом на сайте английского словаря-тезауруса (*English Thesaurus Dictionary* <https://thesaurus.plus/>) отмечается наличие 207 синонимов этого термина (<https://thesaurus.plus/synonyms/formalization>), что подчёркивает важность контекста в употреблении и понимании термина, в передаваемом смысле, в границах и содержании понятия. В этом же словаре-тезаурусе *формализация и утверждение смысла* рассматриваются как связанные термины, где «*Formalization – The act of making formal (as by stating formal rules governing classes of expressions)*», а «*Statement of meaning and formalization are semantically related in description topic. In some cases you can use "Statement of meaning" instead a "Formalization"*» ([https://thesaurus.plus/related/formalization/statement\\_of\\_meaning](https://thesaurus.plus/related/formalization/statement_of_meaning)).

Самая свежая работа Курта Кейгла о роли контекста в данных добавляет интереса к смыслу и его формализации (*Kurt Cagle. The role of context in data.* 19/06/2021. <https://www.bbntimes.com/science/the-role-of-context-in-data>.)

Редакция настроена активно обсуждать затронутую тему формализации смысла и ждёт от читателей материалов, способных приблизиться к решению проблем, актуальных для создателей СИИ. *Прим.ред.*

## Formalization of the meaning<sup>4</sup>. Part 1

**A.D. Redozubov**

*Foundation named after Academician Natalia Bekhtereva, St. Petersburg, Russia*

### Abstract

Every area of knowledge builds its descriptions using concepts. At the same time, the definition of concepts given through their characteristic features is widespread. On this basis, both basic mathematical and many philosophical concepts are built. The concepts which a person uses are subject to similar properties, and their nature is the nature of definitions. Numerous attempts to create strong artificial intelligence are based on the corresponding paradigm. The article attempts to substantiate the need to use the contextual-semantic paradigm to explain the work of the natural brain and to create a strong artificial intelligence. A formal model describing the meaning is presented, and its connection with the known data on the functioning of the brain is given. It is shown that a context can be created around each concept, which can be the bearer of the concept's meaning. The context allows one to move away from using a set of features to recognize the phenomenon behind a concept. The context turns out to be a point of view associated with the concept, in which the description of the surrounding world changes. Knowing the rules of these changes, one can not only model different points of view, but also determine which of them create adequate interpretations. At the same time, the presence of an adequate interpretation in the context of the phenomenon serves as a criterion for the presence of this phenomenon.

**Keywords:** *concept, meaning, context, brain, artificial intelligence, strong artificial intelligence.*

**Citation:** *Redozubov AD.* Formalization of the meaning. Part 1 [In Russian]. *Ontology of Designing.* 2020; 11(2): 144-153. DOI: 10.18287/2223-9537-2021-11-2-144-153.

### List of figure

Figure 1 – Turing test

### References

- [1] **Turing A.** Computing Machinery and Intelligence. *Mind*, vol.LIX, no.236. October 1950. P.433-460.
- [2] **Redozubov A, Klepikov D.** The meaning of things as a concept in a strong AI architecture. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, 12177 LNAI. P.290-300.
- [3] **Falikman M, Spiridonov V.** Cognitive psychology: history and modernity [In Russian]. Moscow: Lomonosov, 2011.- 384 p.
- [4] **Corballis MC.** From hand to mouth. The origins of language. Princeton University Press. 2002. – 257 p.

---

<sup>4</sup> The topic of the article is archly relevant in the digital, informational world and is closely related to artificial intelligence (AI), with its models, ontologies of subject areas and tasks solved in them. The dream of researchers engaged in the creation of strong AI attracts a huge number of specialists from different fields, seeking to shift the burden of decision-making to the "machines" that are being created. Biologists, physiologists, psychologists, mathematicians, computer scientists are actively involved in the process of creating strong AI. At the same time, the complexity of the problem under study makes it difficult to discuss it by scientists from different subject areas. Specialists who are able to present the accumulated material in an accessible form, to generalize it, to find an opportunity to overcome the joint barriers of sciences, take up the task. The author of the article is known as a popularizer of brain science, having written a number of books on the work of the brain: "Brain for rent. How human thinking works and how to create a soul for a computer", 2010; "The logic of emotions", 2012; "Logic of thinking", 2014. The article was prepared following the results of the scientific seminar "Ontology of designing", at which the author made a presentation "Formalization of meaning: concepts, principles, models, examples." The seminar was held online on May 25, 2021. The recording of the seminar is available at <https://bbb.ssau.ru/playback/presentation/2.0/playback.html?meetingId=8fcb5765c6784570796bb5894aae0dec6a08db06-1621939322257>. The editorial board expresses hope for the development of discussion and the search for formalisms capable of recording and transmitting the meaning of information, looking for formal and other solutions in the selection of truthful and reliable information from "fakes", which became the scourge of the beginning of the 21st century. The bold step of the author of the article in the field of formalization of meaning and its further transfer should be considered as a call for the development of ideas in this area and the search for ways to implement them. *Editorial note.*

- 
- [5] **Miller AG**. Project Grammmarama. New York, NY: Basic Books., 1967.
- [6] **Reber AS**. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*. 1967; 6: 855-863.
- [7] **Berry DC, Broadbent DE**. The combination of explicit and implicit learning processes in task control. *Psychol. Res*. 1987; 49: 7-15.
- [8] **Plato**. Theetetus [In Russian]. Trans. from Greek, and approx. V. Serezhnikova. Moscow: SOTSEKGIZ, 1936. - 192 p.
- [9] Concept [In Russian]. Dictionaries and eccyclopedias on Academician.. - <https://dic.academic.ru/dic.nsf/ruwiki/16579>.
- [10] **Mahalanobis PC**. On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*. 1936; 2(1): 49-55.
- [11] **McCulloch WS, Pitts W**. A logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*. 1943; 5: 115-133.
- [12] **Gross C**. Genealogy of the Grandmother Cell. *Neuroscientist*. 2002. 8 (5). P.512-518.,
- [13] **Wille R**. Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies. *Formal Concept Analysis. Foundations and Applications*. Springer Science & Business Media, 2005.
- [14] **Redozubov A**. Holographic Memory: A Novel Model of Information Processing by Neuronal Microcircuits. In: Opris I, Casanova M. (eds) *The Physics of the Mind and Brain Disorders*. Springer Series in Cognitive and Neural Systems, vol 11. Springer, 2017. Cham. [https://doi.org/10.1007/978-3-319-29674-6\\_13](https://doi.org/10.1007/978-3-319-29674-6_13).
- 

### About the author

**Alexey Redozubov** (b. 1968), Applied mathematician. Studied applied mathematics at the Leningrad Polytechnic Institute. For a long time he studied the mechanisms of perception of works of art, mathematical analysis of the phenomenon and the construction of an explanatory model. Co-founder of the Foundation for the Support of Scientific Research of the Mechanisms of the Brain, Treatment of Its Diseases, Neuro-modeling named after Academician Natalya Petrovna Bekhtereva. His area of interest is the creation of strong artificial intelligence based on neurophysiology. Author ID (Scopus): 57218267840. [galdrd@gmail.com](mailto:galdrd@gmail.com).

---

*Received June 7, 2021. Revised June 19, 2021. Accepted June 22, 2021.*

---

## ПРИКЛАДНЫЕ ОНТОЛОГИИ ПРОЕКТИРОВАНИЯ

УДК 004.89

DOI: 10.18287/2223-9537-2021-11-2-154-169

### Формирование базы знаний для поддержки процесса архитектурного проектирования программных систем

Г.Ю. Гуськов, А.М. Наместников, А.А. Романов, А.А. Филиппов

Ульяновский государственный технический университет, Ульяновск, Россия

#### Аннотация

Рассмотрен подход к формированию базы знаний для автоматизации процесса архитектурного проектирования программных систем на основе опыта предыдущих проектов. Под архитектуризацией понимается представление программных систем в форме описания архитектуры, состоящей из множества артефактов проектирования. При разработке новой программной системы возможно повысить её качество за счёт привлечения опыта предыдущих проектов - удачных проектных и архитектурных решений, содержащихся в базе знаний проектной организации. Такая база знаний должна формироваться в процессе анализа артефактов проектирования, полученных в процессе работы над предыдущими проектами: исходный код проекта, диаграммы проектирования, модели данных, структурированные информационные ресурсы и т. д. В статье рассмотрена модель базы знаний проектной организации, позволяющая учитывать опыт предыдущих проектов. Представлена модель прикладного решения платформы 1С: Предприятие в качестве примера артефакта проектирования. Представлен метод для формирования фрагментов базы знаний в процессе анализа прикладного решения для платформы 1С и метод генерации диаграмм вариантов использования на основе содержимого базы знаний. Приведены результаты экспериментов оценки качества по точности (наличие в сгенерированной диаграмме элементов экспертной диаграммы) и полноте (наличие в сгенерированной диаграмме элементов, отсутствующих в экспертной диаграмме). Полученные оценки значений точности в среднем составляет 0.875, а полноты – 0.6.

**Ключевые слова:** архитектуризация, программная система, артефакт проектирования, база знаний, проектный опыт.

**Цитирование:** Гуськов, Г.Ю. Формирование базы знаний для поддержки процесса архитектурного проектирования программных систем / Г.Ю. Гуськов, А.М. Наместников, А.А. Романов, А.А. Филиппов // Онтология проектирования. – 2021. – Т.11, №2(40). - С.154-169. – DOI: 10.18287/2223-9537-2021-11-2-154-169.

#### Введение

В настоящее время сложность разработки и поддержки программных систем (ПС) достигла уровня, при котором необходимо применять принципы и процедуры процесса архитектуризации [1]. Под *архитектуризацией* понимается представление ПС в форме описания архитектуры, состоящей из множества артефактов проектирования. Свойства ПС определяют поведение системы, её состав и подходы к развитию, влияющие на выполнимость, полезность и жизнеспособность системы [2-4]. Описание архитектуры ПС, полученное в процессе архитектурного проектирования, используется для координации действий сторон, вовлечённых в процессы разработки, внедрения, поддержки и использования системы. В процессе архитектурного проектирования ПС формируются требования к системе на всех стадиях жизненного цикла, выделяются особенности процессов разработки, внедрения, эксплуатации и сопровождения ПС.

Архитектуру ПС можно рассматривать как целостную концепцию основных свойств системы, воспринимаемую наилучшим образом через деловое, физическое и техническое представления этой архитектуры [1]. Описание архитектуры ПС определяется как совокупность артефактов проектирования, полученных в процессе разработки ПС. Артефакты проектирования – элементы, составляющие описание архитектуры ПС.

Представление архитектуры ПС в виде множества артефактов проектирования позволяет систематизировать описание архитектуры ПС с позиций заинтересованных сторон в рамках предметных областей (ПрО). Особенности ПрО формируют контекст, в рамках которого разрабатывается и функционирует ПС. Контекст ПС определяет множество архитектурных требований и ограничений, непосредственно влияющих на состав и характеристики артефактов проектирования [5]:

- определение основных компонентов и элементов ПС;
- подходы к организации и взаимодействию элементов ПС;
- принципы организации ПС;
- принципы организации проекта ПС;
- принципы развития ПС в рамках её жизненного цикла.

Качество описания архитектуры ПС влияет на качество процесса разработки и качество системы [6, 7]. Чем на более позднем этапе разработки ПС обнаружены ошибки проектирования, тем больше ресурсов требуется для их исправления.

Описание архитектуры ПС также может быть получено в процессе переработки описания архитектуры предыдущих проектов [1, 4]. Удачные проектные решения, полученные при работе над предыдущими проектами, должны быть использованы в процессе работы над новыми проектами.

## 1 Постановка задачи

Современная проектная организация обладает значительным по объёму проектным заделом, состоящим из артефактов проектирования, полученных в процессе работы над предыдущими проектами. Артефакты проектирования содержат в себе опыт и знания (единицы проектного опыта, ЕПО) большого числа специалистов, которые на протяжении многих лет занимались разработкой ПС. Под ЕПО понимаются удачные проектные и архитектурные решения, полученные при работе над предыдущими проектами.

ЕПО зафиксированы в различных артефактах проектирования:

- проектная документация, представленная в виде слабоструктурированного текста на естественном языке;
- проектные диаграммы, обычно представленные в нотации *UML*;
- исходные коды;
- модели данных объектов ПрО, представленные в виде сущностей и связей реляционных баз данных;
- различные структурированные информационные ресурсы, описывающие особенности разработки, развёртывания и настройки окружения проекта, обычно представленные в виде вики-ресурсов.

При увеличении количества гетерогенных неструктурированных данных затрудняется поиск необходимых сведений. В результате опыт предыдущих разработок, зафиксированный в артефактах проектирования, остаётся невостребованным, и, как следствие, увеличивается время выполнения цикла разработки ПС, снижается качество конечного продукта.

В работах [8, 9] отмечается, что на начальных этапах разработки новых проектов важным является использование опыта предыдущих разработок. Решение указанной проблемы может

основываться на применении интеллектуальных методов и алгоритмов анализа артефактов проектирования с целью построения базы знаний (БЗ) проектной организации.

Разработка ПС состоит из нескольких этапов, качество выполнения которых влияет на качество системы [6, 7]. Основным этапом разработки новой ПС является проектирование, в рамках которого формируется описание архитектуры и модель разрабатываемой системы. На данном этапе должны быть решены следующие задачи:

- поиск критических участков проекта;
- формирование окончательного описания архитектуры ПС;
- объектное моделирование и проектирование основных элементов ПС.

Форматы представления и методы хранения артефактов проектирования различны, что затрудняет их анализ и формирование фрагментов БЗ. Полученная в результате такого анализа БЗ проектной организации может быть применена для:

- накопления множества проектных решений, используемых в проектной организации;
- извлечения часто используемых ЕПО;
- анализа качества полученных ЕПО;
- определения предпочтений в выборе ЕПО разработчиками;
- определения показателя изменения динамики интереса к различным ЕПО во времени;
- прототипирования проекта на основе анализа технического задания или спецификации и множества ЕПО, содержащихся в БЗ.

Учёт специфики проектного опыта, полученного при работе над предыдущими проектами, приводит к необходимости формирования БЗ проектной организации особой структуры, основанной на системе онтологий.

## 2 База знаний проектной организации

Исследователи в области инженерии знаний отмечают актуальность исследований, основанных на онтологическом подходе [10-13], а в работах [14-18] - важность онтологического инжиниринга в процессах проектирования и разработки ПС. Так, в [14, 17] предлагается использовать вместо традиционных языков моделирования ПС (например *UML*) онтологии, которые позволяют контролировать логическую целостность и непротиворечивость полученной модели. Однако существующие методы формирования БЗ в виде набора онтологий для поддержки и автоматизации процесса архитектуризации ПС предполагают привлечение экспертов ПрО и специалистов в области инженерии знаний, что требует существенных временных затрат на формирование такой БЗ.

Следовательно, актуальной является разработка моделей и методов построения БЗ проектной организации в процессе автоматизированного анализа артефактов проектирования, полученных в процессе работы над предыдущими проектами. Сложность здесь заключается в необходимости унификации артефактов проектирования для формирования БЗ, являющейся основой информационного обеспечения комплекса средств автоматизации архитектурного проектирования ПС, а также в необходимости включения в БЗ сведений о различных ПрО, в рамках которых осуществлялась работа над предыдущими проектами.

### 2.1 Онтология ПрО

Онтология ПрО позволяет адаптировать БЗ проектной организации к особенностям ПрО, в рамках которых проводилась работа над предыдущими проектами. Онтология ПрО выступает семантическим базисом, позволяющим сформировать единый понятийный аппарат, объединяющий и унифицирующий сущности различных ПрО.

Онтология ПрО представляется в виде множества:

$$O^{dom} = \{O^{dom}, \dots, O_i^{dom}, \dots, O_n^{dom}\}, \quad (i = 1, \dots, n), \quad (1)$$

где  $n$  – количество ПрО, в рамках которых проводилась работа над предыдущими проектами;  $O_i^{dom}$  –  $i$ -я онтология ПрО, содержащая описание особенностей  $i$ -й ПрО. Формальное представление любой  $O^{dom}$  запишется в виде кортежа:

$$O^{dom} = \langle C^{dom}, T^{dom}, R^{dom}, F^{dom} \rangle, \quad (2)$$

где  $C^{dom}$  – множество понятий ПрО;

$T^{dom}$  – множество терминов ПрО, описывающих понятия ПрО;

$F^{dom}$  – множество функций интерпретации, определяющих множество отношений  $R^{dom}$ ;

$R^{dom}$  – множество отношений вида:

$$R^{dom} = \{R_{CC}^{dom}, R_{CT}^{dom}\},$$

где  $R_{CC}^{dom}$  – множество отношений обобщения, определяющих иерархию понятий ПрО («понятие-понятие»);

$R_{CT}^{dom}$  – множество отношений ассоциации, формирующих связи между понятиями и терминами, описывающими данные понятия («понятие-термин»).

## 2.2 Онтология артефактов проектирования

Онтология артефактов проектирования содержит формализованные описания артефактов проектирования в виде фрагментов БЗ проектной организации. Базовыми элементами данной онтологии являются:

- *сущность* – объект некоторой ПрО, смоделированный с помощью артефакта проектирования, например, класс, таблица и т.д.;
- *атрибут сущности* – свойство сущности, например, поле класса, столбец таблицы и т.п.;
- *ограничение атрибута сущности* – свойства атрибута сущности, например, тип поля класса, размер поля класса с типом «строка» и т.д.;
- *бизнес-процесс* – процесс в некоторой ПрО, в который вовлечена некоторая сущность, например, метод класса, вариант использования и т.д.

Онтология артефактов проектирования представляется следующим кортежем:

$$O^{exp} = \langle E^{exp}, A^{exp}, C^{exp}, R^{exp}, F^{exp} \rangle, \quad (3)$$

где  $E^{exp}$  – множество сущностей;

$A^{exp}$  – множество атрибутов сущностей;

$C^{exp}$  – множество ограничений атрибутов сущностей;

$R^{exp}$  – множество бизнес-процессов, в которые вовлечены сущности;

$F^{exp}$  – множество функций интерпретации, определяющее множество отношений  $R^{exp}$ ;

$R^{exp}$  – множество отношений вида:

$$R^{exp} = \{R_{EE}^{exp}, R_{EA}^{exp}, R_{AC}^{exp}, R_{EP}^{exp}\},$$

где  $R_{EE}^{exp}$  – множество отношений обобщения, определяющие иерархию сущностей («сущность-сущность»);

$R_{EA}^{exp}$  – множество отношений ассоциации, определяющие связь между сущностью и её атрибутом («сущность-атрибут»);

$R_{AC}^{exp}$  – множество отношений ассоциации, определяющие связь между атрибутом сущности и его ограничением («атрибут-ограничение»);

$R_{EP}^{exp}$  – множество отношений ассоциации, определяющие связь между сущностью и бизнес-процессом, в который она вовлечена («сущность - бизнес-процесс»).

### 2.3 Модель базы знаний проектной организации

Рассмотренные в подразделах 2.1 и 2.2 онтологии формируют компоненты БЗ проектной организации. Модель БЗ проектной организации можно представить в виде кортежа:

$$O = \langle O^{dom}, O^{exp}, R, F \rangle, \quad (4)$$

в котором  $O^{dom}$  – онтология для учёта сведений о различных ПрО, в рамках которых производилась работа над предыдущими проектами (1);

$O^{exp}$  – онтология для представления артефактов проектирования, формируемая в процессе анализа артефактов проектирования (3);

$R$  – множество отношений ассоциации между понятиями онтологии ПрО и артефактами проектирования;

$F$  – множество функций интерпретации, определяющих множество отношений  $R$ .

### 3 Формирование фрагментов базы знаний в процессе анализа артефактов проектирования

Формирование фрагментов БЗ в процессе анализа структуры прикладного решения разработано с применением платформы 1С: Предприятие 8 [19]. Платформа 1С широко используется для автоматизации различных типов учёта: управленческий, оперативный, бухгалтерский и т.д. Для каждого типа учёта разрабатываются типовые прикладные решения, выполняемые в рамках платформы 1С. Прикладное решение платформы 1С представляет собой комплексный артефакт проектирования, содержащий опыт и знания большого числа высококвалифицированных специалистов. Прикладное решение платформы 1С можно представить в виде следующего выражения:

$$Conf = \{S_1^{Conf}, \dots, S_i^{Conf}, \dots, S_n^{Conf}\}, i = 1, \dots, n,$$

где  $S_i^{Conf}$  –  $i$ -я подсистема прикладного решения. Подсистема используется для группировки объектов прикладного решения в рамках некоторого контекста, например: «Закупки», «Кадры», «Склад» и т. д. Подсистему можно представить в виде следующего выражения:

$$S_i^{Conf} = (Name, E^{Conf}, P^{Conf}), \quad (5)$$

где  $Name$  – уникальное имя подсистемы;

$E^{Conf} = \langle C^{Conf}, N^{Conf} \rangle$  – множество сущностей прикладного решения; для представления сущностей ПрО в рамках прикладного решения 1С используется множество справочников  $C^{Conf}$  и множество перечислений  $N^{Conf}$ ;

$P^{Conf} = \langle D^{Conf}, R^{Conf} \rangle$  – множество бизнес-процессов прикладного решения, в которые вовлечены сущности прикладного решения. Множество бизнес-процессов  $P^{Conf}$  содержит множество документов  $D^{Conf}$  и множество регистров  $R^{Conf}$ . Документы используются для изменения состояния сущностей прикладного решения. Регистры прикладного решения используются для хранения истории состояний сущностей. Записи регистров формируются на основе данных документов.

Каждый элемент из множеств  $C^{Conf}$  и  $D^{Conf}$  можно представить в виде следующего выражения:

$$C_i^{Conf}, D_i^{Conf} = \langle Name, V^{Conf}, T^{Conf} \rangle,$$

где  $V^{Conf} = \{V_1^{Conf}, \dots, V_i^{Conf}, \dots, V_m^{Conf}\}$  – множество атрибутов сущности;

$V_i^{Conf} = \langle Name, Type, Constraint \rangle$  – каждый атрибут элемента характеризуется именем  $Name$ , типом данных  $Type$  и ограничениями  $Constraint$ . При этом тип данных  $Type$  может представлять собой не только простой тип данных, но и ссылку на другой объект прикладного решения (элемент справочника, документ и т. д.);

$T^{Conf} = \{T_1^{Conf}, \dots, T_i^{Conf}, \dots, T_m^{Conf}\}$  – множество таблиц, связанных с сущностью. Таблицы в рамках платформы 1С используются для формирования связей вида «один ко многим» для группировки данных рамках одной сущности, например, элемент справочника «Сотрудник» может содержать таблицу со списком его детей;  $T_i^{Conf} = \langle Name, V^{Conf} \rangle$  – каждая таблица сущности содержит уникальное имя и множество атрибутов.

Элемент множества  $R^{Conf}$  можно записать как  $R_i^{Conf} = \langle Name, V^{Conf} \rangle$ , а элемент множества  $N^{Conf}$  имеет вид  $N_i^{Conf} = \langle Name \rangle$

Таблица 1 содержит соответствие объектов –  $i$ -й подсистемы прикладного решения и аксиом онтологии артефактов проектирования.

Алгоритм формирования фрагментов БЗ представлен на примере анализа прикладного решения платформы 1С следующей структуры:

- подсистемы: «Торговля».
- справочники: «Товары», «ЕдиницыИзмерения», «Склады», «Поставщики».
- документы: «ПоставкаТовара».
- атрибуты справочника «Товары»: «Код» (строка), «Наименование» (строка), «ЕдиницаИзмерения» (элемент справочника «ЕдиницыИзмерения»).
- атрибуты справочника «ЕдиницыИзмерения»: «Код» (строка), «Наименование» (строка).
- атрибуты справочника «Склады»: «Код» (строка), «Наименование» (строка).
- атрибуты справочника «Поставщики»: «Код» (строка), «Наименование» (строка).
- атрибуты документа «ПоставкаТовара»: «Склад» (элемент справочника «Склады»), «Поставщик» (элемент справочника «Поставщики»), «Товары» (таблица).
- атрибуты таблицы «Товары»: «Товар» (элемент справочника «Товары»), «Цена» (число), «Количество» (число), «Сумма» (число).

Алгоритм формирования фрагментов БЗ в процессе анализа прикладного решения платформы 1С можно записать в виде следующей последовательности шагов.

- 1) для каждой подсистемы прикладного решения сформировать сущность онтологии артефактов проектирования:

$$S_i^{Conf} \rightarrow E_i^{exp},$$

например, Торговля  $\in E^{exp}$ .

- 2) для каждой сущности текущей подсистемы прикладного решения создать сущность онтологии артефактов проектирования с учётом связи с подсистемой:

$$C^{Conf} \rightarrow \hat{E}^{exp} \in E^{exp}, \hat{R}_{EE}^{exp} \in R_{EE}^{exp},$$

$$N^{Conf} \rightarrow \hat{E}^{exp} \in E^{exp}, \hat{R}_{EE}^{exp} \in R_{EE}^{exp},$$

например, Товары  $\in E^{exp}$ , ЕдиницыИзмерения  $\in E^{exp}$ ,  
Склады  $\in E^{exp}$ , Поставщики  $\in E^{exp}$ ,  
Товары  $R_{EE}^{exp}$  Торговля, ..., Поставщики  $R_{EE}^{exp}$  Торговля.

Таблица 1 – Соответствие объектов подсистемы прикладного решения и аксиом онтологии  $O^{exp}$

Прикладное решение	Онтология
$S_i^{Conf}.Name$	$E_i^{exp}$
$C_i^{Conf}.Name$	$E_i^{exp}, R_{EE_i}^{exp}$
$N_i^{Conf}.Name$	$E_i^{exp}, R_{EE_i}^{exp}$
$T_i^{Conf}.Name$	$E_i^{exp}, R_{EE_i}^{exp}$
$D_i^{Conf}.Name$	$P_i^{exp}, R_{EE_i}^{exp}$
$R_i^{Conf}.Name$	$P_i^{exp}, R_{EE_i}^{exp}$
$V_i^{Conf}.Name$	$A_i^{exp}, R_{EA_i}^{exp}$
$V_i^{Conf}.Type, V_i^{Conf} \in C_j^{Conf}$	$(C_i^{exp}, R_{AC_i}^{exp})$
$V_i^{Conf}.Type, V_i^{Conf} \in D_j^{Conf}$	$R_{EP_i}^{exp}$
$V_i^{Conf}.Constraint$	$C_i^{exp}, R_{AC_i}^{exp}$

- 3) если текущая сущность прикладного решения является справочником, создать множество атрибутов, ограничений атрибутов и отношений для данной сущности онтологии артефактов проектирования:

$$V^{Conf} \in C_i^{Conf} \rightarrow \hat{A}^{exp} \in A^{exp}, \hat{R}_{EA}^{exp} \in R_{EA}^{exp},$$

$$V^{Conf} \in C_i^{Conf} \rightarrow \hat{C}^{exp} \in C^{exp}, \hat{R}_{AC}^{exp} \in R_{AC}^{exp},$$

например, Код  $\in A^{exp}$ , Наименование  $\in A^{exp}$ , ЕдиницаИзмерения  $\in A^{exp}$ ,  
Товары  $R_{EA}^{exp}$  Код, Товары  $R_{EA}^{exp}$  Наименование, Товары  $R_{EA}^{exp}$  ЕдиницаИзмерения,

...,  
Поставщики  $R_{EA}^{exp}$  Код, Поставщики  $R_{EA}^{exp}$  Наименование,  
СтрокаТип  $\in C^{exp}$ , ЧислоТип  $\in C^{exp}$ , СсылкаТип  $\in C^{exp}$ ,  
Код  $R_{AC}^{exp}$  СтрокаТип, Наименование  $R_{AC}^{exp}$  СтрокаТип,  
ЕдиницаИзмерения  $R_{AC}^{exp}$  СсылкаТип.

- 4) если текущая сущность прикладного решения имеет таблицы (справочник), то для каждой таблицы сформировать сущность онтологии артефактов проектирования, множество атрибутов, ограничений атрибутов и отношений для данной сущности:

$$T^{Conf} \in C_i^{Conf} \rightarrow \hat{E}^{exp} \in E^{exp}, \hat{R}_{EE}^{exp} \in R_{EE}^{exp},$$

$$V^{Conf} \in T_j^{Conf}, T_j^{Conf} \in C_i^{Conf} \rightarrow \hat{A}^{exp} \in A^{exp}, \hat{R}_{EA}^{exp} \in R_{EA}^{exp},$$

$$V^{Conf} \in T_j^{Conf}, T_j^{Conf} \in C_i^{Conf} \rightarrow \hat{C}^{exp} \in C^{exp}, \hat{R}_{EC}^{exp} \in R_{EC}^{exp}.$$

- 5) для каждого бизнес-процесса текущей подсистемы прикладного решения создать бизнес-процесс онтологии артефактов проектирования с учётом связи с подсистемой:

$$D^{Conf} \rightarrow \hat{P}^{exp} \in P^{exp}, \hat{R}_{EE}^{exp} \in R_{EE}^{exp},$$

$$R^{Conf} \rightarrow \hat{P}^{exp} \in E^{exp}, \hat{R}_{EE}^{exp} \in R_{EE}^{exp},$$

например, ПоставкаТовара  $\in P^{exp}$ , ПоставкаТовара  $R_{EE}^{exp}$  Торговля.

- 6) создать множество отношений для связи текущего бизнес-процесса и сущностей онтологии артефактов проектирования на основе анализа атрибутов текущего бизнес-процесса прикладного решения:

$$V^{Conf} \in D_i^{Conf} \rightarrow \hat{R}_{EP}^{exp} \in R_{EP}^{exp},$$

$$V^{Conf} \in R_i^{Conf} \rightarrow \hat{R}_{EP}^{exp} \in R_{EP}^{exp},$$

например, Поставщики  $R_{EP}^{exp}$  ПоставкаТовара, Склады  $R_{EP}^{exp}$  ПоставкаТовара.

- 7) если текущий бизнес-процесс прикладного решения имеет таблицы (документ), то для каждой таблицы сформировать множество отношений для связи текущего бизнес-процесса и сущностей онтологии артефактов проектирования:

$$V^{Conf} \in T_j^{Conf}, T_j^{Conf} \in D_i^{Conf} \rightarrow \hat{R}_{EP}^{exp} \in R_{EP}^{exp},$$

например, Товары  $R_{EP}^{exp}$  ПоставкаТовара.

Для представленного прикладного решения платформы 1С формируется онтология ПрО. Например, такая онтология может включать следующие элементы:

Товар  $\in C^{dom}$ , Склад  $\in C^{dom}$ , Поставщик  $\in C^{dom}$ ,  
ЕдиницаИзмерения  $\in C^{dom}$ , ПоставкаТовара  $\in C^{dom}$ ,  
Товар  $\in T^{dom}$ , Склад  $\in T^{dom}$ , Поставщик  $\in T^{dom}$ ,  
Единица измерения  $\in T^{dom}$ , ..., Поставка товара  $\in T^{dom}$ ,  
Товар  $R_{CT}^{dom}$ , ..., ПоставкаТовара  $R_{CT}^{dom}$  Поставка товара.

Полученный в результате анализа прикладного решения платформы 1С фрагмент БЗ в виде онтологии представлен на рисунке 1.

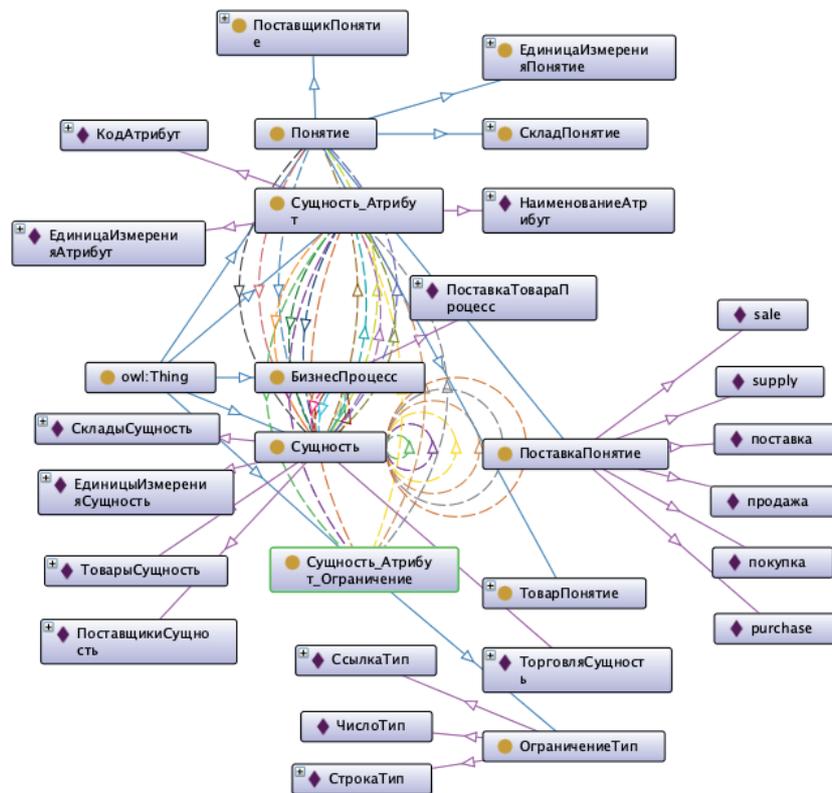


Рисунок 1 – Пример сформированной онтологии

#### 4 Генерация диаграмм вариантов использования на основе базы знаний

Диаграмму вариантов использования можно записать в виде следующего выражения:

$$UCD = \langle A^{UCD}, P^{UCD}, R^{UCD} \rangle, \tag{6}$$

где  $A^{UCD}$  – множество акторов диаграммы вариантов использования;

$P^{UCD}$  – множество прецедентов (вариантов использования);

$R^{UCD}$  – множество отношений диаграммы вариантов использования:

$$R^{UCD} = \{R_C^{UCD}, R_I^{UCD}, R_E^{UCD}\},$$

в котором  $R_C^{UCD}$  – множество отношений обобщения;

$R_I^{UCD}$  – множество отношений включения;

$R_E^{UCD}$  – множество отношений расширения.

Тогда постановка задачи генерации диаграммы вариантов использования заключается в преобразовании онтологии  $O^{exp}$  (3) в элементы модели диаграммы (6).

Таблица 2 содержит описание сопоставления структурных компонентов диаграммы вариантов использования с сущностями онтологии  $O^{exp}$ . В настоящий момент при формировании диаграммы вариантов использования на основе содержимого БЗ формируется

Таблица 2 – Соответствие компонентов онтологии артефактов проектирования и диаграммы вариантов использования

Онтология	Диаграмма вариантов использования
$P_i^{exp}$	Прецедент $P_i^{UCD}$
$P_i^{exp}, P_j^{exp}$	Отношение включения $R_I^{UCD}$

только отношение включения. Для формирования отношений расширения и обобщения требуется дополнительная проработка метода.

Алгоритм генерации диаграммы вариантов использования на основе содержимого БЗ состоит из следующих шагов:

- 1) выбор бизнес-процессов онтологии в качестве основы диаграммы вариантов использования, например,  $\text{ПоставкаТовара} \in P^{exp}$ ;
- 2) формирование списка сущностей онтологии, с которыми связаны выбранные бизнес-процессы, например:  $\text{Товары} \in E^{exp}$ ,  $\text{ЕдиницыИзмерения} \in E^{exp}$ ,  $\text{Склады} \in E^{exp}$ ,  $\text{Поставщики} \in E^{exp}$ ;
- 3) формирование иерархии сущностей онтологии для генерации диаграммы вариантов использования на основе анализа отношений вида  $R_{EE}^{exp}$  онтологии, например:  $\text{Товары} R_{EE}^{exp}$   $\text{Торговля}$ , ...,  $\text{Поставщики} R_{EE}^{exp}$   $\text{Торговля}$ ;
- 4) замена сущностей полученной иерархии на связанные с ней бизнес-процессы (если такого бизнес-процесса ещё нет в диаграмме);
- 5) обход иерархии с корневого узла. Формирование прецедента для каждого узла, организация отношений включения между прецедентами, например:  $\text{Торговля} R_I^{UML}$   $\text{Поставка товара}$  и т.д.

После выполнения данного алгоритма формируется набор команд для системы *PlantUML* [20], например:

```
@startuml
    :user: - (Торговля)
    (Торговля)..>(Поставка товара):include
    (Поставка товара)..>(Внести данные Поставщик):include
    (Поставка товара)..>(Внести данные Склад):include
    (Поставка товара)..>(Внести данные Товар):include
    (Поставка товара)..>(Оформить Поставка товара):include
@enduml
```

На рисунке 2 для рассмотренного примера представлен результат генерации диаграммы вариантов использования.

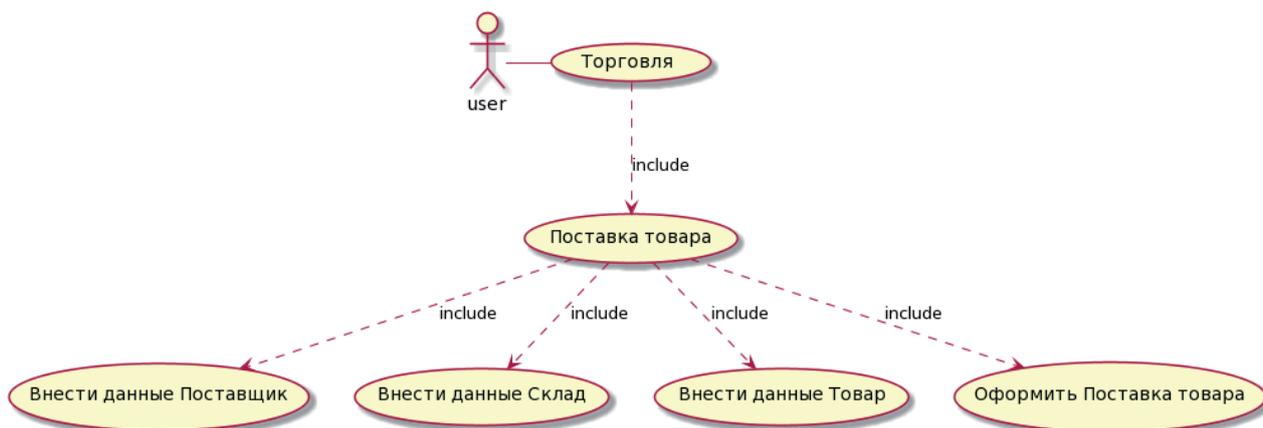


Рисунок 2 – Пример сгенерированной диаграммы вариантов использования

## 5 Эксперименты

Для оценки адекватности предложенных методов и моделей проведён ряд экспериментов, план которых включает:

- формирование содержимого БЗ проектной организации в виде *OWL* онтологии в процессе анализа прикладного решения для платформы 1С;
- проверка полученной *OWL* онтологии на корректность с применением механизма логического вывода редактора онтологий *Protégé* [21];
- генерация набора команд системы *PlantUML* для формирования диаграмм вариантов использования;
- оценка качества полученных диаграмм вариантов использования путём их сравнения с аналогичными диаграммами, построенными экспертом.

Для оценки качества полученных диаграмм вариантов использования применялись следующие метрики.

*Точность:*

$$Precision = \frac{|Gen \cap Exp|}{|Exp|},$$

где  $|Gen \cap Exp|$  – количество совпадающих элементов в сгенерированной диаграмме и диаграмме, построенной экспертом;

$|Exp|$  – количество элементов в диаграмме, построенной экспертом.

*Полнота:*

$$Recall = \frac{|Gen - Exp|}{|Gen|},$$

где  $|Gen - Exp|$  – количество элементов в сгенерированной диаграмме, которые отсутствуют в диаграмме, построенной экспертом;

$|Gen|$  – количество элементов в сгенерированной диаграмме.

В отличие от аналогичных метрик качества в информационном поиске, при оценке качества диаграмм вариантов использования, полученных с применением предложенного подхода, точность и полнота не являются взаимоисключающими значениями.

## 6 Оценка качества сгенерированных диаграмм вариантов использования

В качестве входных данных для формирования содержимого БЗ использовалось типовое прикладное решение для платформы 1С – «Управление торговлей».

Формирование *OWL* онтологии проводилось при анализе объектов прикладного решения, входящих в подсистему «Расхождения»:

- подсистемы «Закупки» (родительская подсистема) и «Расхождения»;
- регистр накопления «ГрафикиОтгрузки»;
- справочники «Организации», «Валюты», «Партнеры», «Контрагенты», «Пользователи», «Номенклатура», «Назначения», «Склады»;
- документы «АктОРасхожденияхПослеПеремещенияТоваров», «АктОРасхожденияхПослеПриемкиТоваров» и «АктОРасхожденияхПослеОтгрузкиТоваров».

Проверка полученной *OWL* онтологии в редакторе *Protégé* закончилась успешно.

Для проведения данного эксперимента в редакторе *Protégé* была составлена онтология *PrO*  $O^{dom}$  (2), содержащая названия прецедентов для объектов прикладного решения платформы 1С. Например, для понятия «Закупки» установлено описание «Закупка товаров», для понятия «Склады» установлено описание «Данные о складах», а для понятия «АктОРасхожденияхПослеПеремещенияТоваров» описание – «Акт о расхождениях после перемещения».

В сгенерированной на основе содержимого *OWL* онтологии диаграмме вариантов использования (рисунок 3) содержится 14 прецедентов, из которых 9 не содержится в диаграмме, построенной экспертом (рисунок 4). Количество прецедентов диаграммы вариантов использования, построенной экспертом – 5. Таким образом, получены следующие значения метрик качества:  $Precision = 5/5=1$ ,  $Recall = 9 / 14 = 0.643$ .

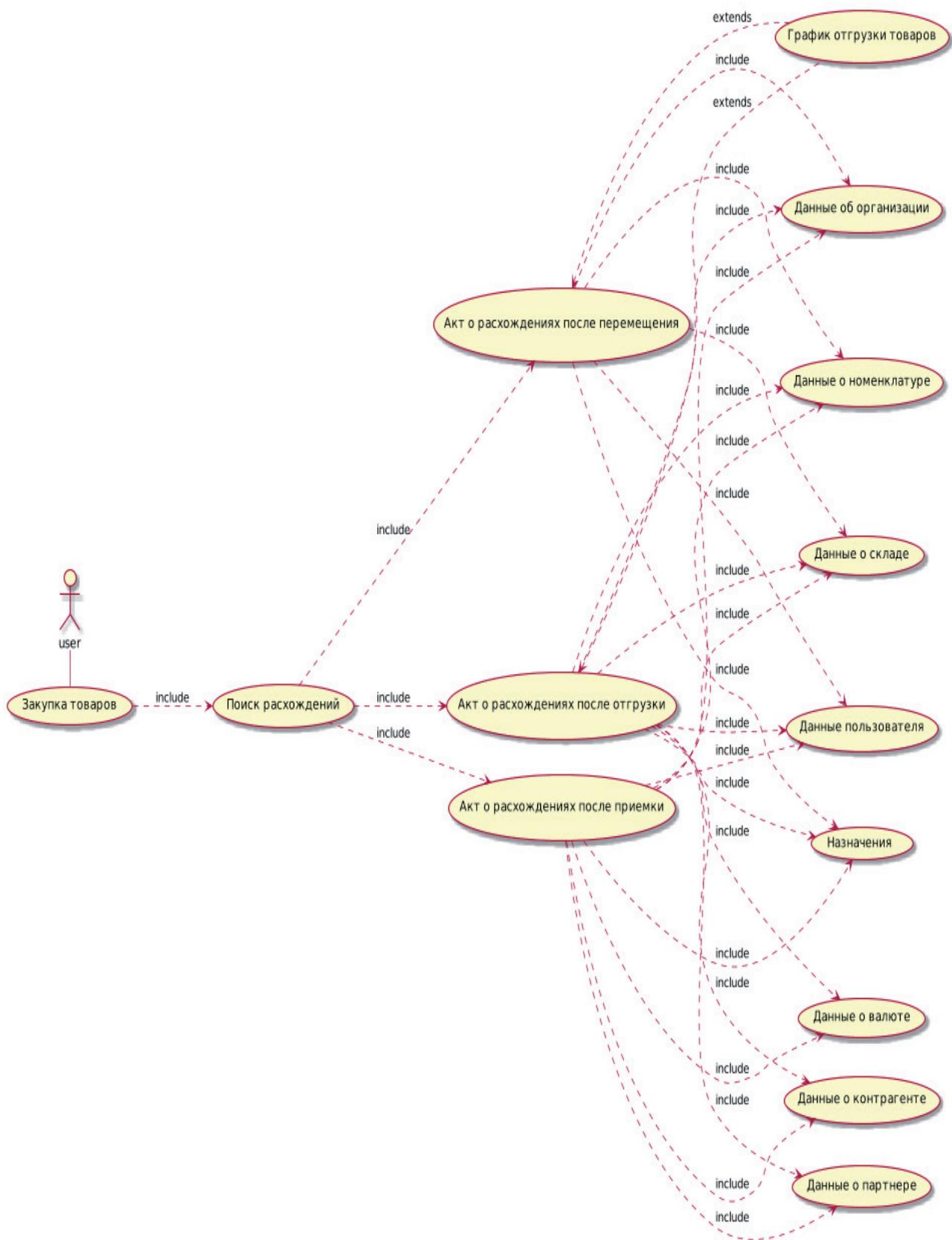


Рисунок 3 – Сгенерированная диаграмма вариантов использования для прикладного решения на платформе 1С «Управление торговлей»

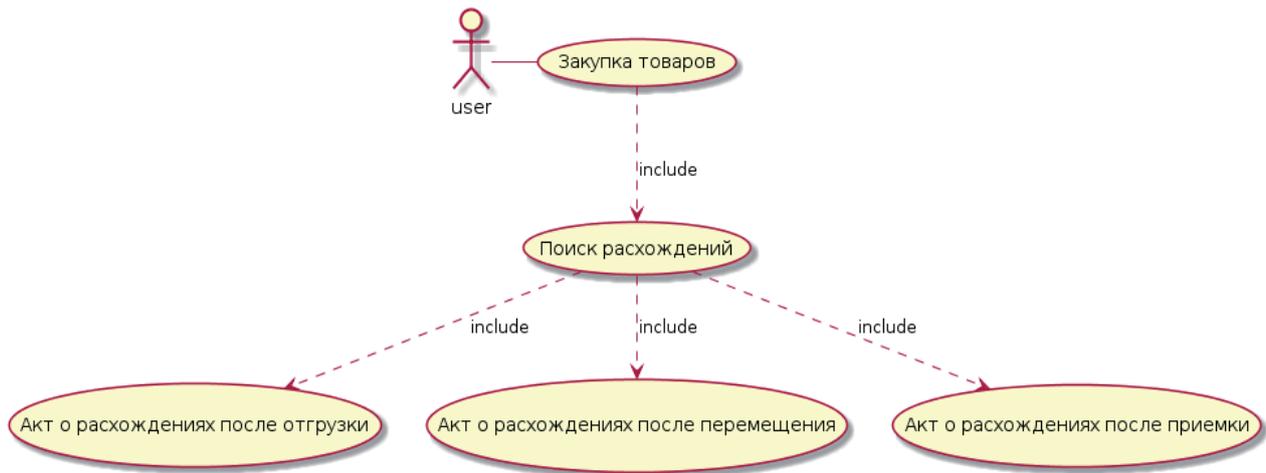


Рисунок 4 – Диаграмма вариантов использования для прикладного решения на платформе 1С «Управление торговлей», построенная экспертом

Время, затраченное на автоматическую генерацию диаграммы вариантов использования, составило 3 минуты (без учёта времени формирования онтологии Про  $O^{dom}$ ). Экспертом на решение аналогичной задачи было потрачено 17 минут.

Таблица 3 содержит результаты экспериментов по формированию диаграмм вариантов использования на основе содержимого БЗ, сформированной на основе анализа различных прикладных решений для платформы 1С.

Таблица 3 – Результаты проведённых экспериментов

Конфигурация	Подсистема	Кол-во элементов			Точность	Полнота
		Метод	Эксперт	Совпадающие		
Управление торговлей	Расхождения (закупки со сверкой товаров)	14	5	9	1	0.643
Зарплата и управление персоналом	Кадры (регистрация переработок)	8	4	4	1	0.5
Комплексная автоматизация	Склад (излишки, недостачи, порча продукции)	12	6	4	0.667	0.667
Управление торговлей	Банк и касса (работа с кассой)	13	6	5	0.833	0.615

Полученные оценки значений точности в среднем составляет 0.875, а полноты – 0.6. Как видно из результатов экспериментов, сгенерированные артефакты проектирования имеют высокий показатель точности, но при этом такие артефакты проектирования содержат большее количество структурных элементов, что повышает показатель полноты. Высокий показатель полноты связан с учётом связей между объектами анализируемого прикладного решения для платформы 1С, которые получены в процессе их жизненного цикла и не были спроектированы изначально.

## Заключение

Предложен подход к автоматизации формирования фрагментов БЗ для поддержки архитектурного проектирования ПС. БЗ представляет собой систему онтологий для учёта особенностей Про и хранения формализованного представления артефактов проектирования, полу-

ченных при работе над предыдущими проектами. Дано описание теоретико-множественной модели БЗ проектов. В качестве примера артефакта проектирования использовано прикладное решение на платформе 1С.

Представлены результаты экспериментов оценки качества предложенных моделей и методов путём сравнения сгенерированных диаграмм вариантов использования с построенными экспертом.

Дальнейшее развитие предложенного подхода видится в работе с различными артефактами проектирования и в выделении ЕПО из содержимого БЗ с применением механизмов логического вывода. Выделение ЕПО в процессе анализа содержимого БЗ проектной организации позволит сократить пространство и время поиска удачных проектных решений.

## Благодарности

Исследование выполнено в рамках государственного задания № 075-00233-20-05 по проекту «Исследование интеллектуального предиктивного мультимодального анализа больших данных и извлечения знаний из разных источников», а также при финансовой поддержке РФФИ и Правительства Ульяновской области в рамках научных проектов № 19-47-730006, 19-47-730005.

## Список источников

- [1] ГОСТ Р 57100-2016/ISO/IEC/IEEE 42010:2011. Системная и программная инженерия. Описание архитектуры. М.: Стандартинформ, 2019. - <http://docs.cntd.ru/document/1200139542>.
- [2] **Heesch van, U.** A documentation framework for architecture decisions / U. van Heesch, P. Avgeriou, R. Hilliard // *Journal Syst. Softw.* 2012. vol. 85, 4, p.795-820. DOI:10.1016/j.jss.2011.10.017.
- [3] **Hilliard, R.** Using Aspects in Architectural Description / R. Hilliard // *Lecture Notes in Computer Science*, 2007. vol. 4765. P.65-68.
- [4] **Sosnin, P.** Substantially evolutionary theorizing in designing software-intensive systems / P. Sosnin // *Information (Switzerland)*, 2018. 9 (4). P.91.
- [5] **Sosnin, P.** Ontological controlling the lexical items in conceptual solution of project tasks / P. Sosnin, A. Pushkareva // *Lecture Notes in Computer*, 2017. Vol. 10409. P.31-46.
- [6] ISO/IEC 25010 2011. ISO/IEC 25010:2011: Systems And Software Engineering – Systems And Soft-Ware Quality Requirements And Evaluation (Square) – System And Software Quality Models.
- [7] **Макконнелл, С.** Совершенный код: Практическое руководство по разработке программного обеспечения / С. Макконнелл. СПб.: Русская редакция/БХВ, 2017.
- [8] **Маклаев, В.А.** Инструментально-технологическая среда формирования и использования опыта проектной организации / В.А. Маклаев, П.И. Соснин // *Автоматизация процессов управления*. 2009. № 2 (16). С.8-13.
- [9] **Соснин, П.И.** Архитектурный подход к прецедентно-ориентированному решению задач в разработке автоматизированных систем / П.И. Соснин, С.С. Шумилов, А.Е. Ивасев // *Автоматизация процессов управления*. 2019. № 1 (55). С.49-56.
- [10] **Загорулько, Ю.А.** Использование паттернов онтологического проектирования для разработки онтологий предметных областей / Ю.А. Загорулько, О.И. Боровикова, Г.Б. Загорулько // *Сборник тр. конф. «Знания-Онтологии-Теории» (ЗОНТ-2017)*. Новосибирск, 2017. С.139-148.
- [11] **Муромцев, Д.И.** Интеграция wiki-технологии и онтологического моделирования в задаче управления знаниями предприятия / Д.И. Муромцев [и др.] // *Сборник тр. конф. «КИИ-2008»*. Дубна, 2008. С.360-368.
- [12] **Самойлов, Д.Е.** Паттерны структурной организации системы измеряемых свойств в онтологическом анализе данных / Д.Е. Самойлов, В.А. Семенова, С.В. Смирнов // *Сборник тр. конф. «Проблемы управления и моделирования в сложных системах»*. Самара, 2018. С.358-366.
- [13] **Shaaban, A.M.** Ontology-based security tool for critical cyber-physical systems / A.M. Shaaban, T. Gruber, C. Schmittner // *Proceedings of the 23rd International Systems and Software Product Line Conference*. Vol. B. 2019. P.207-210.
- [14] **Bhatia, M.P.S.** Ontologies for software engineering: past, present and future / M.P.S. Bhatia, A. Kumar, R. Beniwal // *Indian Journal of Science and Technology*. 2016. Vol. 9. P.1-16.

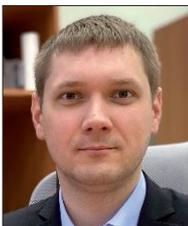
- [15] *Falbo, R.A. et al.* An ontology pattern language for service modeling / R.A. Falbo et al. // Proceedings of the 31st Annual ACM Symposium on Applied Computing. 2016. P.321-326.
- [16] *Ilyas, Q.M.* Ontology Augmented Software Engineering / Q.M. Ilyas // Software Development Techniques for Constructive Information Systems Design. IGI Global, 2013. P.406-413.
- [17] *Isotani, S. et al.* Ontology driven software engineering: a review of challenges and opportunities / S. Isotani et al. // IEEE Latin America Transactions. 2015. Vol.13. P.863-869.
- [18] *Pan, J.Z. et al.* Ontology-driven software development / J.Z. Pan et al. Springer Science & Business Media, 2012.
- [19] What is 1C:Enterprise? - [https://1c-dn.com/1c\\_enterprise/what\\_is\\_1c\\_enterprise/](https://1c-dn.com/1c_enterprise/what_is_1c_enterprise/).
- [20] PlantUML. UML Diagram Generator. - <https://plantuml.com>.
- [21] Protégé. Free, open-source ontology editor and framework for building intelligent systems. - <https://protege.stanford.edu>.

## Сведения об авторах



**Гуськов Глеб Юрьевич**, 1992 г. рождения, кандидат технических наук, окончил факультет информационных систем и технологий Ульяновского государственного технического университета (УлГТУ). Доцент кафедры «Информационные системы» УлГТУ. Имеет около 30 работ в области онтологического инжиниринга и интеллектуального анализа данных. Author ID (RSCI): 812005; Author ID (Scopus): 57021635300. [g.guskov@ulstu.ru](mailto:g.guskov@ulstu.ru).

**Наместников Алексей Михайлович**, 1974 г. рождения, доктор технических наук, доцент, окончил радиотехнический факультет УлГТУ. Профессор кафедры «Информационные системы» УлГТУ. Имеет более 80 работ в области автоматизированного проектирования и интеллектуальных систем. Author ID (RSCI): 392690; Author ID (Scopus): 9277806100. [nam@ulstu.ru](mailto:nam@ulstu.ru).



**Романов Антон Алексеевич**, 1986 г. рождения, кандидат технических наук, окончил факультет информационных систем и технологий УлГТУ, доцент кафедры «Информационные системы» УлГТУ. Имеет около 60 работ в области интеллектуальных систем хранения и обработки информации. ORCID: 0000-0001-5275-7628; Author ID (RSCI): 684949; Author ID (Scopus): 55903279200. [romanov73@gmail.com](mailto:romanov73@gmail.com).

**Филиппов Алексей Александрович**, 1987 г. рождения, кандидат технических наук, окончил факультет информационных систем и технологий УлГТУ, доцент кафедры «Информационные системы» УлГТУ. Имеет около 80 работ в области онтологического инжиниринга и интеллектуального анализа данных. ORCID: 0000-0001-5275-7628; Author ID (RSCI): 708454; Author ID (Scopus): 57191472723. [al.filippov@ulstu.ru](mailto:al.filippov@ulstu.ru).



Поступила в редакцию 27.05.2021, после рецензирования 17.06.2021. Принята к публикации 23.06.2021.

## Formation of a knowledge base to support the process of architectural design of software systems

G.Y. Guskov, A.M. Namestnikov, A.A. Romanov, A.A. Filippov

*Ulyanovsk State Technical University, Ulyanovsk, Russia*

### Abstract

This article describes an approach to knowledge base (KB) formation for automating the process of architectural design of software systems (SS) based on the experience of previous projects. Software architecting is the presentation of software systems in the form of design artifacts and their architecture. When developing a new SS it is possible to improve its quality based on the experience of previous projects. The experience of previous projects is successful architectural

solutions contained in the knowledge base of the design organization. Such a KB should be formed in the process of analyzing design artifacts extracted from previous projects: source code, project diagrams, data models, structured text resources, etc. This article describes a KB model of a design organization and a model of the 1C: Enterprise 8 (1C) application solution as an example of a design artifact. The article also presents a method for generating fragments of a KB in the process of analyzing an applied solution for the 1C application and a method for generating use-case diagrams based on the KB content. A set of experiments was executed to evaluate the adequacy of the proposed models and methods. The results of experiments for assessing quality in terms of accuracy (the presence of elements of the expert diagram in the generated diagram) and completeness (the presence of elements in the generated diagram that are absent in the expert diagram) are presented. According to the results of the experiments, the average value of accuracy is 0.875, and the completeness is 0.6.

**Key words:** *architecting, software system, design artifact, knowledge base, design experience.*

**Citation:** *Guskov GY, Namestnikov AM, Romanov AA, Filippov AA. Formation of a knowledge base to support the architecting process of software systems [In Russian]. Ontology of designing. 2021; 11(2): 154-169. DOI: 10.18287/2223-9537-2021-11-2-154-169.*

**Acknowledgment:** The authors acknowledge that the work was supported by the framework of the state task of the Ministry of Science and Higher Education of the Russian Federation No.075-00233-20-05 "Research of intelligent predictive multimodal analysis of big data, and the extraction of knowledge from different sources". The reported study was funded by RFBR and the government of Ulyanovsk region according to the research projects: 19-47-730006 and 19-47-730005.

## List of figures and tables

Figure 1 - Example of a generated ontology

Figure 2 - Example of generated use case diagram

Figure 3 - Use case diagram for the "Trade Management" 1C application generated by the system

Figure 4 - Use case diagram for the "Trade Management" 1C application built by the expert

Table 1 - Correspondence between the 1C solution objects and the  $O^{exp}$  ontology axioms

Table 2 - Correspondence between design artifact ontology components and use-case diagram components

Table 3 - Experimental results

## References

- [1] ISO/IEC/IEEE 42010:2011 Systems and Software Engineering - Architecture Description. 2011. <https://www.iso.org/standard/50508.html>.
- [2] **Heesch U, Avgeriou P, Hilliard R.** A documentation framework for architecture decisions. *Journal Syst. Softw.* 2012; 85(4): 795-820. DOI:10.1016/j.jss.2011.10.017.
- [3] **Hilliard R.** Using Aspects in Architectural Description. *Lecture Notes in Computer Science*, 2007; 4765: 65-68.
- [4] **Sosnin P.** Substantially evolutionary theorizing in designing software-intensive systems. *Information* (Switzerland), 2018; 9 (4): 91.
- [5] **Sosnin P, Pushkareva A.** Ontological controlling the lexical items in conceptual solution of project tasks. // *Lecture Notes in Computer*, 2017; 10409: 31-46.
- [6] ISO/IEC 25010 2011. ISO/IEC 25010:2011: Systems And Software Engineering – Systems And Soft-Ware Quality Requirements And Evaluation (Square) – System And Software Quality Models.
- [7] **McConnell S.** Code complete. – Pearson Education, 2004.
- [8] **Maklaev VA, Sosnin PI.** Instrumental and technological environment for design organization experience formation and usage [In Russian] // *Automation of Control Processes*. 2009; 2(16): 8-13.
- [9] **Sosnin PI, Shumilov SS, Ivasev AE.** An architectural approach to the precedent-oriented problem solving in the designing of automated systems [In Russian]. *Automation of Control Processes*. 2019; 1(55): 49-56.
- [10] **Zagorulko YA, Borovikova OI, Zagorulko GB.** Using Ontological Design Patterns to Domain Ontologies Development [In Russian]. Proceedings of the Conference "Knowledge-Ontology-Theory" (3OHT-2017). Novosibirsk, 2017. P.139-148.
- [11] **Muromtsev DI et al.** Integration of wiki technology and ontological modeling in the enterprise knowledge management [In Russian]. Proceedings of the RCAI Conference (КИИ-2008). Dubna, 2008. P.360-368.

- [12] **Samoylov DE, Semenova VA, Smirnov SV.** Patterns of the structural organization of the system of measured properties in ontological data analysis [In Russian]. Proceedings of the Conference «Control and modeling problems in complex systems». Samara, 2018. P.358-366.
  - [13] **Shaaban AM, Gruber T, Schmittner C.** Ontology-based security tool for critical cyber-physical systems // Proceedings of the 23rd International Systems and Software Product Line Conference. 2019; B: 207-210.
  - [14] **Bhatia MPS., Kumar A, Beniwal R.** Ontologies for software engineering: past, present and future. Indian Journal of Science and Technology. 2016; 9: 1-16.
  - [15] **Falbo RA et al.** An ontology pattern language for service modeling. Proceedings of the 31st Annual ACM Symposium on Applied Computing. 2016. P.321-326.
  - [16] **Ilyas QM.** Ontology Augmented Software Engineering. Software Development Techniques for Constructive Information Systems Design. IGI Global, 2013. P.406-413.
  - [17] **Isotani S. et al.** Ontology driven software engineering: a review of challenges and opportunities. IEEE Latin America Transactions. 2015; 13: 863-869.
  - [18] **Pan JZ. et al.** Ontology-driven software development. Springer Science & Business Media, 2012.
  - [19] What is 1C:Enterprise? [https://1c-dn.com/1c\\_enterprise/what\\_is\\_1c\\_enterprise/](https://1c-dn.com/1c_enterprise/what_is_1c_enterprise/).
  - [20] PlantUML. UML Diagram Generator. <https://plantuml.com>.
  - [21] Protégé. Free, open-source ontology editor and framework for building intelligent systems. <https://protege.stanford.edu>.
- 

## About the authors

**Gleb Yurievich Guskov** (b. 1992), graduated from Ulyanovsk State Technical University in 2014, PhD (2018). He is an Associate Professor at Ulyanovsk State Technical University (Department of information systems). He is a co-author of about 30 scientific articles and abstracts in the field of ontology engineering and data mining. Author ID (RSCI): 812005; Author ID (Scopus): 57021635300. [g.guskov@ulstu.ru](mailto:g.guskov@ulstu.ru).

**Aleksey Mihaylovich Namestnikov** (b. 1974), graduated from Ulyanovsk State Technical University in 1996, D. Sc. Eng. (2018). He is a Professor at Ulyanovsk State Technical University (Department of information systems). He is a co-author of about 80 scientific articles and abstracts in the field of CAD and AI. Author ID (RSCI): 392690; Author ID (Scopus): 9277806100. [nam@ulstu.ru](mailto:nam@ulstu.ru).

**Anton Alekseevich Romanov** (b. 1986), graduated from Ulyanovsk State Technical University in 2009, PhD (2013). He is an Associate Professor at Ulyanovsk State Technical University (Department of information systems). He is a co-author of about 60 scientific articles and abstracts in the field of intelligent systems. ORCID: 0000-0001-5275-7628; Author ID (RSCI): 684949; Author ID (Scopus): 55903279200. [romanov73@gmail.com](mailto:romanov73@gmail.com).

**Aleksey Aleksandrovich Filippov** (b. 1987), graduated from Ulyanovsk State Technical University in 2009, PhD (2013). He is an Associate Professor at Ulyanovsk State Technical University (Department of information systems). He is a co-author of about 80 scientific articles and abstracts in the field of ontology engineering and data mining. ORCID: 0000-0001-5275-7628; Author ID (RSCI): 708454; Author ID (Scopus): 57191472723. [al.filippov@ulstu.ru](mailto:al.filippov@ulstu.ru).

---

Received 27 May, 2021. Revised June 17, 2021. Accepted 23 June, 2021.

---

## Управление мастер-данными в рамках итеративного подхода

С.В. Кузнецов<sup>1</sup>, Д.В. Кознов<sup>2</sup>

<sup>1</sup>ООО «Юнидата», Санкт-Петербург, Россия

<sup>2</sup>Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

### Аннотация

Управление мастер-данными (*Master-Data Management, MDM*) – область бизнес-информатики, нацеленная на консолидацию и централизованное управление важнейшими данными бизнес-организаций, которые распределены по разным информационным системам. Ведущие мировые ИТ-компании (*IBM, Oracle, Informatica* и др.) предлагают широкий спектр готовых продуктов по управлению мастер-данными (*MDM-продукты*). Внедрение *MDM* сопряжено с большим количеством технических и организационных трудностей. В работе рассматривается итеративная стратегия внедрения *MDM*, подразумевающая поэтапную реализацию управления мастер-данными на основе реальных нужд организации-заказчика. Вводится понятие *MDM-решения*, которое является результатом внедрения *MDM* в организацию и включает в себя адаптированный к нуждам и особенностям организации *MDM-продукт*, новые регламенты по работе с данными обученных сотрудников, налаженный процесс управления мастер-данными. Основным результатом статьи является функциональная модель управления мастер-данными, предназначенная для ранних стадий разработки *MDM-решения* в рамках итеративной стратегии. Потребности организации могут быть похожи на *MDM*, однако требуют использования других инструментов. Цель данной модели – выявить реальные потребности организации в *MDM* и установить, какие *MDM-компоненты* должны быть реализованы в первую очередь. В работе приводится описание компонент модели, представлено портфолио из шести реальных *MDM-проектов* и их анализ с позиций предложенной модели.

**Ключевые слова:** управление мастер-данными, бизнес-информатика, прикладная онтология, итеративный подход, информационная система.

**Цитирование:** Кузнецов, С.В. Управление мастер-данными в рамках итеративного подхода / С.В. Кузнецов, Д.В. Кознов // Онтология проектирования. – 2021. – Т.11, №2(40). – С.170-184. – DOI: 10.18287/2223-9537-2021-11-2-170-184.

### Введение

Переход к цифровой экономике предъявляет особые требования к управлению данными в организациях [1]. При этом имеется особый вид данных, без которых трудно себе представить нормальную работу организаций: юридические данные, клиентская база, сведения о поставщиках и контрагентах и пр. Пользователи этих данных рассчитывают на их согласованность в пределах организации, т.е. они ожидают получить одну и ту же информацию об одном и том же объекте из разных источников (информационных систем, ИС) организации. Разночтения и противоречия здесь порождают проблемы: задержки, коллизии, финансовые и имиджевые потери организации [2]. Такие консолидированные данные принято называть мастер-данными (*Master Data*, далее МД)<sup>1</sup>, а процесс их консолидации и сопровождения – управлением МД (*Master Data Management, MDM*) [1, 8].

В области *MDM* существует обширный спектр стратегий и методов, структурированных в рамках энциклопедии *DAMA-DMBOK (Data Management Body of Knowledge)* [2]. Существует большое количество готовых *MDM-продуктов* от таких компаний как *IBM, Oracle,*

<sup>1</sup> В России МД часто называют основными данными, при этом выделяют особый подвид МД – справочные данные (реестры, классификаторы, справочники и пр.), для которых существует много различных стандартов и решений [3-7].

*Informatica* и др. (см., например, [9]). На сегодняшний день внедрение *MDM* не является типовой, проработанной задачей, поскольку организации, особенно крупные, имеют большое количество индивидуальных черт.

На практике существуют две принципиально разные стратегии введения *MDM*: сверху-вниз и итеративная стратегия [2]. *Стратегия сверху-вниз* подразумевает следующую последовательность действий: создание стратегической *MDM*-концепции для организации, формирование требований к *MDM*-решению, внедрение и доработка существующего на рынке *MDM*-инструментария, выполнение необходимых организационно-административных работ, эксплуатация и сопровождение *MDM*-решения<sup>2</sup>. *Итеративная стратегия* подразумевает внедрение *MDM* для решения конкретной задачи с дальнейшим наращиванием *MDM*-функционала и/или реализацией *MDM* для других сегментов данных организации, т.е. для решения других задач<sup>3</sup>. Эти две стратегии коррелируют с концепцией внедрения инноваций в организацию, предложенной в [10]. Первую стратегию можно сопоставить с технологией *push*: внедрение инновации происходит на основе некоторой передовой технологии, которая должна решить различные, в том числе и не известные на данный момент проблемы организации. Вторую стратегию можно сопоставить с технологией *pull*: драйвер внедрения инноваций – сама организация, точнее, определённые её задачи. Не отвергая первой стратегии, авторы ориентируются на вторую, как менее рисковую и позволяющую достичь конкретных практических результатов в обозримые сроки.

При реализации *MDM*-проекта в рамках итеративной стратегии возникает задача перевода требований заказчика на язык *MDM*, получивший значительное развитие [2]. Если задача, имеющаяся у организации-заказчика, хорошо переводится в термины *MDM*, значит для её решения можно использовать имеющийся на рынке *MDM*-инструментарий [9], что существенно сокращает затраты на такой проект. При этом на практике оказывается, что заказчик, как правило, не владеет *MDM*-терминологией и часто под видом *MDM*-проектов пытается представить проекты иного класса или заказать реализацию *MDM*-решения «с нуля». Ошибки здесь приводят к коллизиям, растянутым срокам и денежным потерям.

В статье предлагается функциональная модель МД, которая должна помочь в первичной оценке и согласовании работ в итеративных *MDM*-проектах на самых ранних стадиях. Внимание сосредоточено на создании/развертке ИТ-инфраструктуры по поддержке *MDM*, её наладке и выполнении необходимых аналитических работ (очистка и консолидация данных, классификация и иерархизация и т.д.). В дальнейшем эти аналитические работы должны выполняться в организации на постоянной основе с помощью внедренной ИТ-инфраструктуры, и задачей-максимум является полная автоматизация этих работ. Ориентируясь на программно-аналитические аспекты *MDM*, здесь намеренно оставляются в стороне изменение бизнес-процессов и практик работы с данными в организации, обучение персонала и пр. Обычно эти вопросы лучше решаются, когда уже выполнены программно-аналитические работы по внедрению *MDM*. Кроме того, существует специальная область, в сферу которой попадают подобные вопросы – это управление данными (*Data Governance*, [11]). В статье приводятся несколько реальных индустриальных *MDM*-проектов, описанных в рамках предложенной модели.

<sup>2</sup> Как правило, это осуществляется посредством серии проектов, которые выполняются различными внешними компаниями.

<sup>3</sup> Внедрение *MDM* за пределами одного конкретного *MDM*-проекта зависит от многих условий, в том числе от оправдания ожиданий организации от данного проекта, от наличия других задач, свободных средств, заинтересованных людей и т.д. Именно такая цепочка *MDM*-проектов и даёт название всей стратегии – *итеративная стратегия* внедрения *MDM*.

## 1 Мастер-данные

Следуя глоссарию [8], МД являются специальным видом данных организации, описывающих её важнейшие характеристики, в том числе действующих и потенциальных клиентов, поставщиков, потребляемую/выпускаемую продукцию, офисные и производственные площади организации, платёжные реквизиты и различную информацию о счетах юридических и физических лиц, с которыми организация работает. Смысл выделения МД как отдельного объекта управления заключается в том, что в организациях существует большое количество различных ИС и/или используется значительное число внешних разнородных источников данных. В результате оказывается, что одна и та же информация выражена разными данными и в разных форматах. Более того, в разных источниках имеются разные атрибуты для одних и тех же данных, а сами данные могут противоречить друг другу. В организации имеется подмножество критически важных данных, разночтение в которых препятствует нормальному функционированию и наносит организации значительный ущерб. Именно эти данные становятся объектом специальной заботы организации, превращаясь в МД.

Для эффективного создания, использования и сопровождения МД в организации должна быть налажена специальная деятельность, посредством которой организация фокусируется на своих бизнес-процессах, вопросах качества и интеграции данных, а также на стандартизации существующих и используемых ИС [12]. *MDM* ориентирован на сбор и накопление данных из различных ИС – источников данных (ИД), консолидацию данных и их распределение (доставку) ИС – потребителям данных (ПД).

В [2] выделены следующие ключевые процессы *MDM* в организации: управление моделью данных, сбор и накопление данных, проверка, стандартизация и обогащение данных, разрешение конфликтов данных и использование данных.

*MDM* в организации должен быть поддержан специальным ИТ-решением, созданным и внедренным в организацию (далее – *MDM-решение*). Важнейшей составляющей *MDM-решения* является центральный репозиторий данных (хаб, от англ. *hub*). В него собираются данные, являющиеся кандидатами в МД, они надлежащим образом обрабатываются и доставляются ПД. Определены четыре варианта архитектуры хаба данных [9].

- Индексная архитектура: на хабе хранятся только соответствующие ссылки (индексы) данных; это актуально для данных, которые нельзя копировать или перемещать.
- Консолидирующая архитектура: данные регулярно загружаются в хаб, обрабатываются, и при этом обеспечивается доступ ПД к этим данными.
- Централизованная архитектура во всём подобна предыдущей, но в этом случае дополнительно хаб выполняет задачу разового ввода данных, и далее все изменения данных делаются непосредственно на хабе (ИС попадают в разряд ПД).
- Смешанная архитектура реализует сочетание консолидирующей и централизованной архитектур для различных МД организации. Если какие-то фрагменты данных организации запрещено перемещать, то для них может использоваться индексная архитектура.

Существует большое количество готового программного инструментария по созданию *MDM-решений*. Прежде всего, это такие продукты как *SAP MDG*, *Informatica MDM*, *IBM InfoSphere MDM* и др., которые ориентированы на решение стандартных задач *MDM*. Однако разнообразие практических задач столь велико, что некоторые производители (например, *Informatica MDM*, Юнидата и пр.) предлагают программные «конструкторы», которые позволяют получить *MDM-решения* для конкретных потребностей организаций.

## 2 MDM-проекты

На практике *MDM*-проекты часто принимают за обычные ИТ-проекты. И требуется определить, что задачи организации имеют ярко выраженную *MDM-специфику*, что позволяет применить готовый *MDM*-инструментарий, а также привлечь компании-интеграторы, специализирующиеся на таких проектах.

Запрос организации имеет *MDM-специфику*, когда организации необходим сбор, обогащение и консолидация данных из различных ИД, а также выдача этих данных различным ПД, использующим их далее для обеспечения жизнедеятельности организации. ИД могут находиться как в самой организации, так и вне её. Например, требуется обогатить данные о клиентах организации информацией, собранной в соцсетях. Требование наличия нескольких ПД для созданных МД является менее жёстким и иногда не выполняется. Как правило, в организации существует критический бизнес-процесс, для эффективного исполнения которого требуются качественные обогащённые и консолидированные данные из различных ИД. Например, речь может идти о процедуре проверки новых клиентов или спорных транзакций в банке.

С другой стороны, организации может быть и не нужен *MDM*-проект. Во-первых, когда речь идёт об обработке однородных данных – например, сформированных посредством ввода одним или несколькими операторами (операторный ввод). Такие проекты могут потребовать: создания логической модели данных, валидацию и очистку данных, обеспечение доступа к данным в различных режимах и т.д. Но при этом отсутствует главная задача *MDM* – консолидация данных из разных источников. Если для такой задачи операторный ввод данных заменить автоматизированным поступлением тех же данных из разных ИД, то задача приобретает *MDM-специфику*. Во-вторых, «вне юрисдикции» *MDM*-проектов лежат запросы на реализацию сложного бизнес-функционала. Такой функционал должен быть вынесен из *MDM*-решения в отдельные ИС [9], а *MDM*-проект заканчивается доставкой МД ПД.

Следует отметить, что *MDM-специфика* должна быть основной в проекте. Если этой специфики нет или она лишь частично присутствует, но является не главной, то данный ИТ-проект не является *MDM*-проектом.

## 3 MDM-решение

В результате *MDM*-проекта у организации-заказчика появляется *MDM*-решение, которое включает в себя: *MDM*-систему (программную часть решения); новые регламенты по работе с данными; обученных сотрудников организации, которые умеют использовать *MDM*-систему в соответствии с новыми регламентами; налаженный и запущенный процесс управления МД. Последний пункт важен, поскольку всё может иметься в наличии, но деятельность по *MDM* в организации отсутствует. Например, из-за нерешённых вопросов безопасности или противоречий внутренних регламентов, или в силу большой загруженности сотрудников, которые должны участвовать в *MDM*. Таким образом, *MDM*-решение является осуществляющейся в организации деятельностью, налаженной и обеспеченной всем необходимым.

*MDM-система* - это развёрнутое в организации готовое программное обеспечение (ПО), которое реализует основные функции *MDM* – хаб данных, консолидацию и пр. Главной частью этого ПО является базовый *MDM*-продукт (см. в [9]), а также, возможно, дополнительный набор ПО для частных вопросов (например, очистки данных). Наличие многофункционального готового ПО, которое следует лишь настроить и развернуть у организации-заказчика, существенно снижает стоимость и риски *MDM*-проекта. Однако некоторую часть

*MDM*-системы приходится дорабатывать в рамках *MDM*-проекта, чтобы отразить особенности организации, которые не удаётся покрыть стандартным инструментарием<sup>4</sup>.

#### 4 Описание модели

После первой оценки потребностей заказчика необходимо провести их детальный анализ в терминах *MDM* [8, 9]. Для этого в статье предлагается специальная функциональная модель. Она описывает типовое *MDM*-решение, включая в себя максимальный объём функциональности с тем, чтобы можно было выбрать необходимые компоненты, которые нужно реализовать в данной ситуации<sup>5</sup>.

Для удобства использования модель представлена с помощью «метафоры» полного жизненного цикла (ЖЦ) МД и состоит из трёх пакетов (этапов): сбор данных, обработка данных и доставка данных. Пакеты содержат функциональные компоненты, каждая из которых описывает блок работ по наладке/управлению МД. Таким образом, функциональные компоненты модели включают в себя работы по наладке *MDM* и работы, выполняемые в рамках дальнейшего функционирования *MDM*-решения.

Например, при реализации одной из главных компонент модели «Консолидация данных» нужно сделать следующее:

- выполнить работы по наладке: наладить/конфигурировать/дописать ПО, поддерживающее соответствующее рабочее место аналитика, определить правила для разрешения конфликтов и выполнить первый раз консолидацию данных из ИД организации;
- осуществлять консолидацию данных в рамках дальнейшего функционирования *MDM*, поскольку далее данные из ИД будут продолжать поступать на хаб данных.

Предлагаемая модель ориентируется на внедрение новой ИТ-системы в организацию на основе готовых инструментов, которые могут быть настроены и доработаны под особенности задач организации-заказчика. Поэтому каждая компонента модели имеет *программную* и *аналитическую* части. Часть функционала компоненты выполняет соответствующее ПО, а часть – человек (аналитик). Работы по наладке компоненты целесообразно разделить на наладку/реализацию некоторого ПО и выполнение аналитических функций. Например, создание логической модели МД является аналитической функцией, а очистка данных – программно-аналитической. В последнем случае речь идёт о создании и программной реализации специальных правил очистки, которые применимы именно для этих данных, именно для этой организации, и применении этих правил, включая анализ результатов, возможно создание новых правил и т.д. При этом используется готовое ПО, но отдельные специальные правила, ориентированные на специфику данных организации, могут быть реализованы в виде дополнительного ПО, созданного в рамках *MDM*-проекта.

Основные пакеты модели.

- *Сбор данных*. В этот пакет включены компоненты, отвечающие за идентификацию данных-кандидатов в МД, так называемых «сырых» данных, а также за их дальнейший анализ и предварительную обработку. Сюда же входит доступ к различным ИД.

<sup>4</sup> Здесь можно привести аналогию с рынком *ERP*-решений: вначале эти решения, также как и *MDM*-решения сегодня, предназначались для больших организаций и также создавались на основе базовых *ERP*-продуктов со значительными доработками. Далее стали появляться специализированные решения для отдельных отраслей (энергетики, добывающих отраслей, логистики и пр.). Аналогично, в области *MDM* уже сейчас выделяется отдельный класс продуктов, ориентированных на работу с конкретной предметной областью (данные продукты принято называть *Product Information Management, PIM*).

<sup>5</sup> Предложенную модель можно также назвать идеальной моделью «*to be*» – эта терминология принята в области структурного [13] и объектно-ориентированного анализа [14], а также в области реинжиниринга бизнес-процессов [15].

- **Обработка данных.** В этот пакет включается функционал по созданию и хранению МД в хабе данных, включая создание и поддержку логической модели данных, а также выполнение классификации, иерархизации и пр. В хаб поступают предварительно обработанные «сырые» данные, полученные из ИД. Здесь они обрабатываются, становясь МД.
- **Доставка данных.** В этот пакет включены функциональные компоненты, отвечающие за доставку МД системам ПД. ИД и ПД могут совпадать полностью или частично. При этом оказывается важным решение вопросов разделения прав доступа к МД, а также реализация различных режимов доставки данных для ПД. Выделяют следующие режимы доставки МД: пакетный, режим реального времени и подписочный режим.

Предложенная модель ориентирована на сценарий не только одноразовой загрузки данных, а на повторяющееся обновление МД в хабе с учётом новых «сырых» данных из ИД.

Пакеты и функциональные компоненты модели представлены на рисунке 1.

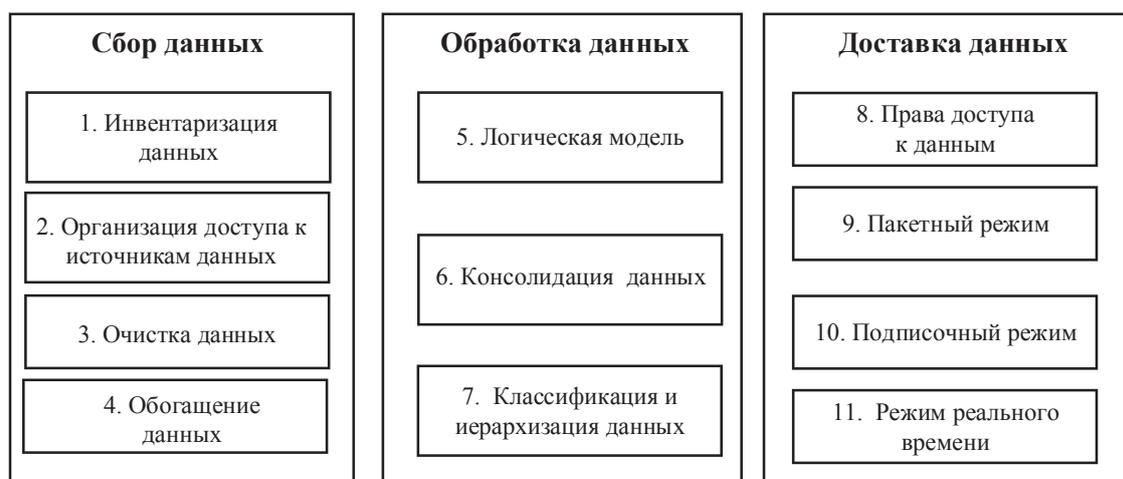


Рисунок 1 - Схема функциональной модели типового MDM-решения

#### 4.1. Инвентаризация данных

В рамках этого пакета производится идентификация ИД, а также определяется, какие именно данные из этих ИД нужно преобразовывать в МД. Необходимо определить точный состав МД. Чем больше разнообразие данных, тем сложнее (и, следовательно, дороже) будет MDM-проект. При этом собирают лишь те атрибуты, которые будут востребованы ПД. Важно провести типизацию данных, выяснить реальную заполняемость каждого существенного атрибута, его типовые значения и пр. В этом пакете также определяется уровень доверия к различным ИД. Возможна ситуация, что некоторый ИД имеет очень низкую степень доверия, в частности, он может хранить данные, которые давно не обновлялись. Обращаться к нему следует лишь в крайнем случае. Эта функциональная компонента является преимущественно аналитической.

#### 4.2. Организация доступа к источникам данных

Поскольку «сырые» данные, которые должны использоваться для создания МД, находятся в различных ИД организации-заказчика, то для создания MDM-решения необходимо организовать программный доступ к этим данным. В большинстве случаев загрузка данных является многократной процедурой и должна выполняться регулярно во время функционирования MDM-решения. Для автоматизации этой процедуры необходимо решить технические задачи, т.к. ИД часто реализованы на разных платформах и могут не иметь программных ин-

терфейсов доступа. Эта функциональная компонента является преимущественно программной. Объём работ здесь во многом зависит от того, насколько обмен данными налажен в организации (например, уже могут быть внедрены технологии передачи данных между различными ИС организации).

#### 4.3. Очистка данных

Под очисткой данных понимается устранение ошибок в данных и нормализация данных из различных ИД перед их загрузкой в хаб. Эта компонента необходима, поскольку в противном случае будет непросто искать в данных дубликаты, а также выполнять их консолидацию. Очистка данных является трудоёмким процессом. Первичная очистка, включая нормализацию и приведение значения всех атрибутов к единому формату, не является затратной, однако она значительно упрощает дальнейшие шаги по консолидации данных и восстановлению связей между сущностями.

Эта компонента является программно-аналитической. На практике часто требуется программная доработка таких инструментов для корректной работы с различными форматами данных организации, либо реализовывать специальную логику по очистке информации. Например, ИД может хранить несколько значений в одном атрибуте, и тогда требуется разделить эти значения по соответствующим полям. Для этого нужно разработать специальное ПО, которое выполнит это разделение.

#### 4.4. Обогащение данных

Может оказаться, что для пользователей МД требуется дополнить имеющиеся в ИД данные, например, информацией из открытых источников.

Данная компонента является программно-аналитической, поскольку требует анализа и разметки соответствующих данных из ИД, а также реализации программного доступа к публичным источникам для обогащения данных.

#### 4.5. Логическая модель

Эта компонента предназначена для создания и сопровождения логической модели МД. Такая модель должна отражать структуру консолидированных данных со всеми атрибутами, собранными из различных ИД организации. Модель необходима для дальнейшей обработки МД, а также их доставки ПД. Одним из важнейших шагов при создании логической модели является восстановление/обнаружение различных связей в данных, которые отсутствовали в ИД, но появляются при консолидации.

Деятельность по созданию логической модели является аналитической. Она должна быть поддержана соответствующим ПО для создания такой модели, средствами визуализации, перечислением атрибутов и связей между сущностями, а также программной связи созданной мастер-модели данных с соответствующим отражением её элементов в ИД и/или ПД. При этом некоторые аспекты этого инструментария требуется дорабатывать под конкретный проект: например, в качестве ИД и ПД может выступать старая ИС, в которой модель данных жёстко задана (типичный случай – *ERP*-система), и тогда доставка новых значений для существующих атрибутов будет требовать специальной программной реализации.

#### 4.6. Консолидация данных

Эта функциональная компонента отвечает за загрузку данных из разных ИД на хаб и выполнение консолидации реальных данных в соответствии с созданной логической моделью.

Процесс загрузки производится автоматически, с использованием соответствующих инструментов. При его выполнении возникают конфликты, которые разрешаются следующими способами.

- «Вручную» – эксперт предметной области разрешает конфликт; этот способ применяется для критических данных (например, юридических), где ошибки недопустимы и поэтому автоматические алгоритмы разрешения конфликтов неприемлемы.
- Семантический (онтологический) подход, который применяется для данных, которые хранятся в виде онтологий. Если фрагмент данных из ИД попадает с другими фрагментами в одну онтологию, то эти фрагменты являются консолидированными.
- Методы искусственного интеллекта (ИИ), в частности, методы машинного обучения, которые обучаются на типичных ситуациях с тем, чтобы разрешать возникающие в процессе консолидации конфликты автоматически.
- Смешанные стратегии, например, с помощью алгоритмов ИИ экспертам представляется на одобрение предварительные варианты разрешения конфликтов. Такой подход может снизить трудоёмкость процедуры разрешения конфликтов без снижения качества.

Загрузка данных из ИД может осуществляться однократно, например, в случае централизованной архитектуры хаба данных или при наличии ИД, которые прекратили свою работу, но содержат ценные данные. Иначе помимо первичной загрузки требуется организовать регулярное обновление МД.

Данная компонента является программно-аналитической. Программной частью является доработка ПО деятельности аналитика по консолидации данных для работы со специфическими данными, а также для реализации уникальных правил консолидации и правил разрешения конфликтов. Если используются алгоритмы ИИ, то они должны быть адаптированы под конкретную задачу. Например, это могут быть обновляемые или самонастраиваемые правила для разбора конфликтов данных при консолидации.

#### 4.7. Классификация и иерархизация данных

Организация нуждается в максимальной систематизации и упорядочении МД (например, основных активов, поставщиков, клиентов). Требуется разделить МД на группы, выделить значимые атрибуты для каждой группы и т.д. Необходимо связать данные организации с различными внешними классификаторами – государственными стандартами, отраслевыми классификаторами и т.д., а во многих случаях также иерархизировать данные. В качестве примера можно привести ситуацию, когда новый заказчик организации оказывается в том же организационном сегменте большой корпорации, что и предыдущий; в таком случае нет необходимости получать новое разрешение у службы безопасности, что экономит время и ресурсы. Деятельность по классификации и иерархизации данных производится над неструктурированными данными и может быть связана с обогащением данных.

Данная компонента является аналитической в части создания правил иерархизации и классификации. Она должна быть поддержана соответствующим ПО, которое позволяет эти правила отлаживать на малом количестве данных и далее применять их к полному объёму данных. Для подсказок аналитику и создания полуавтоматической классификации данных всё больше используют различные методы ИИ, в частности, методы машинного обучения.

#### 4.8. Права доступа к данным

ПД могут находиться в различных бизнес-контурах организации и иметь разные права доступа к данным. В рамках этой компоненты требуется определить стратегию прав доступа к МД и выполнить её реализацию. Здесь необходимо полагаться на существующие в органи-

зации роли и связанные с ними права, взаимодействуя со службой безопасности организации.

Данная компонента включает работы, которые не являются трудоёмкими и оказываются преимущественно аналитическими: политика разграничения прав доступа к МД реализуется средствами администрирования ИС. Однако создание соответствующей спецификации (какие данные и кому должны быть доступны) является ответственной работой, требующей глубокого знания данных и бизнес-процессов, а также структуры организации.

#### 4.9. Пакетный режим

Эта компонента отвечает за загрузку и обновление МД в ПД в соответствии с некоторым расписанием. Многие ПД ориентированы на получения пакетных выгрузок МД в промежуточные базы данных («витрины данных»), с которыми они работают в своём режиме. При этом каждая витрина использует, как правило, свой фрагмент мастер-модели данных. Целесообразно реализовать отдельный механизм управления такими витринами для отслеживания их своевременного обновления актуальными МД, а также для журналирования запросов к МД разными ПД. Таким образом отслеживается, какие именно МД потребляются какими ПД и в каком режиме; какие конфликты данных появляются из каких ИД и как это соотносится с потреблением данных.

Данная компонента имеет программную часть по наладке/реализации интерфейса *MDM*-системы с соответствующими ПД. Аналитическая часть заключается в определении тех ПД и тех частей МД, которые нуждаются именно в такой стратегии.

#### 4.10. Подписочный режим

В рамках этого режима каждый ПД подписывается на определённый фрагмент МД (часть логической модели или множество сущностей и их атрибутов). Далее формируется одна или несколько очередей, куда выгружаются наиболее актуальные МД после их очередного обновления. После этого все ПД считывают свои обновления из данной очереди согласно своей подписке. Сложность реализации подписочной модели состоит в том, что необходимо либо повторно использовать существующий механизм очередей, которым уже пользуются ПД, либо доработать соответствующие ПД для использования очередей *MDM*-решения.

Данная компонента является программно-аналитической.

#### 4.11. Режим реального времени

Данная компонента включает функционал по доставке МД ПД в режиме реального времени, т.е. непосредственно после изменения данных. Такой режим часто трудно реализовать из-за конфликтов на стороне потребителя, поскольку система ПД может временно блокировать доступ к фрагменту данных из-за выполнения некоторой операции, и это приведёт к задержкам с обновлением записи в рамках *MDM*. Например, на сайтах телекоммуникационных компаний часто есть функция проверки возможности подключения той или иной услуги по адресу. Потенциальный клиент заходит на сайт компании, вводит свой адрес (возможно с ошибками) и выбирает интересующую его услугу, например, широкополосный доступ в Интернет. Промедление с ответом сайта для такого клиента может быть опасно, т.к. клиент может уйти к конкуренту. Поэтому такая проверка адреса становится задачей поиска данного адреса (очистка от опечаток) и указанной услуги в режиме реального времени.

Данная компонента является программно-аналитической.

## 5 Примеры

Использование предложенной модели можно показать на примере *MDM*-проектов, выполненных при непосредственном участии авторов (см. таблицу 1).

Таблица 1 – Характеристики *MDM*-проектов

№	Акроним	Название	Сектор индустрии
1.	КМТР	Каталог материально-технических ресурсов	Энергетика
2.	ПК	Продуктовый каталог	Телекоммуникации
3.	КТУ	Каталог товаров и услуг	Транспорт
4.	СКБ	Сегментирование клиентской базы	Розничная торговля
5.	ЛКГ	Личный кабинет горожанина	Государственное управление
6.	ПДК	Проверка данных клиента	Энергетика и тяжелая промышленность

В описании выполненных проектов курсивом выделены компоненты функциональной модели, которые были в фокусе разработчиков.

- КМТР. Данный проект направлен на создание системы для управления данными о материально-технических ресурсах крупной организации в энергетическом секторе. Система предназначалась для решения следующих задач: обеспечить качественными данными бизнес-процессы технического обслуживания, ремонта и управления запасами; консолидировать различную информацию за счёт создания технологии стандартизации и унификации данных. В рамках проекта рассматривались данные о сырье и материалах, используемом оборудовании, запасных частях и комплектующих изделиях, необходимых для обеспечения деятельности организации. К особенностям проекта можно отнести автоматизацию сложных регламентов организации по работе с информацией, затрагивающих более десяти различных подразделений, реализацию классификатора материальных ресурсов (*классификация и иерархизация данных*), построение *логической модели данных*.
- ПК. Проект выполнялся для крупной телекоммуникационной организации и был нацелен на консолидацию информации по следующим направлениям: по продуктовым предложениям (услугам) для различных сегментов заказчиков; по проверке технических возможностей подключения услуг; по объединению финансовой информации из систем биллинга и бухгалтерской отчетности. Основной акцент был сделан на *инвентаризации данных* о продуктах компании из различных ИД, а также на создании единой *логической модели* МД с последующей *консолидацией*. Построено итоговое дерево продуктов компании с различными характеристиками, включая финансовые, для дальнейшего анализа отделом продаж и финансистами (*классификация и иерархизация данных*).
- КТУ. Основной задачей проекта была консолидация товаров и услуг, закупаемых крупной транспортной организацией. Было необходимо объединить информацию из различных классификаторов товаров и составить перечень услуг подрядчиков. Внутренним заказчиками этого *MDM*-решения стала служба закупки организации. В ходе проекта были идентифицированы по своему атрибутивному составу товары и услуги, имеющие различные цены. В результате были созданы монетарные метрики, т.е. подсчитана итоговая экономия организации по закупкам ввиду того, что требуемые товары и услуги стали закупаться по гарантированным минимальным доступным ценам автоматически. Фокус проекта был на *консолидации данных* о закупаемых товарах, *разграничении прав* и обеспечении доступа к данным в рамках *подписочного режима*.
- СКБ. Проект разрабатывался для организации, занимающейся продажей модных товаров, и предназначался для сегментирования клиентской базы и поддержки продаж в премиальном сегменте. Целью проекта было выявление клиентов из клиентской базы организации, которые активны в социальных сетях и имеют много подписчиков. Организа-

ция хотела «заручиться их лояльностью» с помощью дополнительных скидок и других мотивационных акций с целью получить больше потенциальных покупателей – подписчиков этих клиентов. В рамках проекта был сделан акцент на *обогащение* и *консолидацию данных*.

- ЛКГ. Данный проект разрабатывался для городской государственной службы управления с целью создания «умного» личного кабинета горожанина. Требовалось выполнить интеграцию личного кабинета с многочисленными ИС федерального и регионального уровней с целью извлечения профильной информации о горожанине, например, сведений о его транспортных средствах, недвижимости, банковских счетах и пр. Важными особенностями проекта была информационная безопасность и разделение *прав доступа к данным*, а также получение МД *в режиме реального времени* и в рамках *подписочной модели*.
- ПДК. Проект разрабатывался для многопрофильной международной организации из сектора энергетики и тяжелой промышленности. Организация имеет сотни тысяч клиентов в разных странах мира, поэтому процедура создания нового клиента оказывается трудоёмкой. До создания *MDM*-решения она занимала 21 день, а после – всего лишь 8 дней. *MDM*-решение позволило автоматизировать различные проверки, поиск конечных бенефициаров юридических лиц в корпоративных иерархиях, а также реализовать централизованный ввод информации. В рамках данного проекта основной акцент был сделан на *инвентаризации данных*, создании *логической модели* МД, позволившей решить задачу поиска дубликатов юридических лиц и поиска аффилированных лиц, а также реализации доступа к МД *в режиме реального времени* с целью ускорить целевой бизнес-процесс.

Для указанных проектов в таблице 2 показано, какие функциональные компоненты были реализованы в соответствующих *MDM*-проектах. При этом использовалась шкала:

- *High* – компонента является одной из основных в проекте, т.е. она бизнес-критична или технологически сложна;
- *Med(ium)* – компонента важна для проекта, но не является приоритетной или трудоёмкой;
- *Low* – компонента реализована в облегчённом варианте: она либо уже существует к началу проекта и требует лишь доработки, либо полная реализация компоненты была вынесена в отдельный проект;
- *N/A* – данная компонента в рамках этого проекта не востребована.

Таблица 2 – Описание индустриальных *MDM*-проектов в рамках предложенной функциональной модели

Элементы функциональной модели	Проекты					
	КМТР	ПК	КТУ	СКБ	ЛКГ	ПДК
<b>Сбор данных</b>						
Инвентаризация данных	Med	High	Med	Low	Med	High
Организация доступа к источникам данных	Low	Med	Low	Low	Low	Med
Очистка данных	Med	Low	Med	Med	Med	Med
Обогащение данных	N/A	Med	Med	High	n/a	Low
<b>Обработка данных</b>						
Логическая модель	High	High	Med	Low	Med	High
Консолидация данных	Med	High	High	High	Med	Med
Классификация и иерархизация данных	High	High	Med	N/A	N/A	High
<b>Доставка данных</b>						
Права доступа к данным	Low	Low	High	Low	High	Med
Подписочный режим	Med	Low	High	Low	High	Med
Режим реального времени	Med	Low	High	Low	High	Med
Пакетный режим	Med	N/A	Med	Med	High	High

## 6 Сопоставление существующих и описанного подходов

За последние годы создано значительное количество стандартов и методологий управления данными в организациях [2, 16, 17] и др.

Другой вид работ сфокусирован на ПО в сфере *MDM*, см., например [18]. Предлагаемый в данной статье подход отличается от этих методологий тем, что не зависит от конкретного базового *MDM*-продукта.

В [19] предложена модель для построения *MDM* в организации, которая включает семь блоков: концепцию, стратегию, метрики, информационное управление, оргвопросы и роли, ЖЦ информации, а также инфраструктуру. Эта модель предназначена для ранних стадий внедрения *MDM*, однако она ориентирована на стратегию сверху-вниз, покрывая всю деятельность организации по внедрению *MDM*. Предложенная в статье функциональная модель предназначена для итеративной стратегии, ориентированной на удовлетворение конкретных потребностей организации, которые выражаются в терминах *MDM*.

В [20] предлагается модель для анализа ЖЦ МД в организации с целью определить недостающие виды деятельности. Основными компонентами модели являются: портфолио данных; проектирование данных и системы; управление данными, поддержка данных. Эта модель слабо связана с программной частью *MDM*-решения, а также не рассматривает уникальные задачи организации по внедрению *MDM*.

### Заключение

В работе предложена функциональная модель *MDM*-решения, ориентированная на разработку таких решений в рамках итеративной стратегии. Модель предназначена для ранних стадий разработки и призвана перевести потребности организации на язык *MDM* для оценки доли *MDM*-специфики. Модель позволяет выполнить планирование функционала *MDM*-решения и перейти к выбору базового инструментария, а также созданию технического задания на разработку. В работе представлен анализ выполненных *MDM*-проектов на основе предложенной модели.

В качестве дальнейшего продолжения исследования интерес представляет детальная разработка методик оценки функционала *MDM*-проектов на ранних стадиях, а также создание детальных метрик сложности *MDM*-решений. Планируется провести перевод (отображение) функциональности типового *MDM*-решения на различные *MDM*-продукты, а также осуществить более тесную интеграцию подхода с областью управления знаниями [21, 22].

### Список источников

- [1] *Khatri, V.* Designing data governance / V. Khatri V., C.V. Brown // Communications of the ACM. – 2010. – 53(1).
- [2] DAMA-DMBOK: Data Management Body of Knowledge, 2017.
- [3] *Андриченко, А.Н.* Тенденции и состояние управления справочными данными в машиностроении / А.Н. Андриченко // Онтология проектирования. – 2012. – 2 (4). – С. 25-35.
- [4] *Немцов, Э.Ф.* ИСУЖТ и нормативно-справочные данные / Э.Ф. Немцов // Автоматика, связь, информатика. – 2020. – № 2. – С. 15-18.
- [5] *Голубев, С.С.* Отраслевая система государственной службы стандартных справочных данных нефтегазового комплекса / С. С. Голубев, А. Н. Лоцманов, А. Ю. Кузин, В. Г. Соловьев, А. Д. Козлов, Б. А. Григорьев // Законодательная и прикладная метрология. – 2020. – № 3. – С. 12-16.
- [6] *Янченко, Г.А.* К вопросу о стандартизации справочных данных плотностных свойств горных пород / Г.А. Янченко // Горный информационно-аналитический бюллетень (научно-технический журнал). – 2011. – № 8. – С. 111-115.

- [7] **Чигиринский, Ю.Л.** Методика повышения надежности справочных данных / Ю.Л. Чигиринский // Известия Волгоградского государственного технического университета. – 2011. – № 13 (86). – С. 55-61.
  - [8] Gartner Glossary, <https://www.gartner.com/en/glossary>.
  - [9] **Walker, S.** Magic Quadrant for Master Data Management. / S.Walker, A. Dayley, S. Parker, M. Hawker // Gartner. – 2021.
  - [10] **Zmud, R.W.** An Examination of ‘Push-Pull’ Theory Applied to Process Innovation in Knowledge Work / R.W. Zmud // Management Science. – 1984. – 30 (6). – P. 727–738.
  - [11] **Ladley, J.** Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program / J.Ladley. Academic Press; 2nd edition 2019. 350 p.
  - [12] **Silvola, R.** Managing one master data – Challenges and preconditions / R. Silvola, O. Jääskeläinen, H. Kropus-Vehkaperä, H. Naapasalo//Industrial Management & Data Systems. – 2011. – 111(1). – P. 146–162.
  - [13] **Yourdon, E.** Structured Design: Fundamentals of a Discipline of Program and Systems Design / E. Yourdon, L.L. Constantine. Yourdon Press. 1975.
  - [14] **Jacobson, I.** Object-Oriented Software Engineering / I. Jacobson. ASM press. 1992. 528 p.
  - [15] **Ould, M.A.** Business Processes: Modelling and Analysis for Re-Engineering and Improvement / M. A. Ould. Wiley, 1995
  - [16] CMMI Data Management Maturity Model (DMM). CMMI Institute (website). <http://bit.ly/1Vev9xx>.
  - [17] IBM Data Governance Council Maturity Model <https://ibm.co/2sRfBIn>
  - [18] Velocity Methodology. Best Practices. Informatica. 2008.
  - [19] **O’Kane, B.** The Seven Building Blocks of MDM: A Framework for Success / B. O’Kane, M. P. Moran. August 2016. Gartner. ID: G00311161
  - [20] **Ofner, M.H.** Management of the master data lifecycle: a framework for analysis / M. H. Ofner, K. Straub, B. Otto, H. Österle // J. Enterp. Inf. Manag. – 2013. – 26(4). – P. 472-491.
  - [21] **Гаврилова, Т.А.** Представление знаний в экспертной диагностической системе АВТАНТЕСТ / Т.А. Гаврилова // Известия Академии наук СССР. Техническая кибернетика. – 1984. – № 5. – С. 165-173.
  - [22] **Гаврилова, Т.А.** Управление знаниями: от слов к делу / Т.А. Гаврилова, Д.В. Кудрявцев // Intelligent Enterprise: RE (Корпоративные системы). – 2004. – № 12-13 (101). – С. 48.
- 

## Сведения об авторах



**Кузнецов Сергей Викторович**, 1979 г. рождения. Окончил с отличием математико-механический факультет Санкт-Петербургского государственного университета (2001), аспирантуру (2004). Возглавлял научно-исследовательские направления в ряде компаний в области управления данными. Возглавил российское представительство компании Informatica (2010), создал компании ООО «ТаскДата» и ООО «Юнидата» (2014), разработчика одноимённой платформы управления данными. В 2020 и 2021 году система MDM Юнидата была отмечена в ежегодных отчётах консалтингового агентства Gartner по MDM. [sergey@unidata-platform.ru](mailto:sergey@unidata-platform.ru).



**Кознов Дмитрий Владимирович**, 1971 г. рождения. Окончил с отличием математико-механический факультет Санкт-Петербургского государственного университета (1994), к.ф.м.н. (2000), д.т.н. (2016). Профессор кафедры системного программирования Санкт-Петербургского государственного университета. Член ACM. Приглашённый профессор Национального университета Сингапура (Сингапур, 2000, 2001), Технического университета Лаапентанты (Финляндия, 2015), Южно-Китайского Технологического университета (Китай, 2018). Автор более 100 работ в области программной инженерии информационных систем. Author ID (RSCI): 126261. Author ID (Scopus): 8885649400. [d.koznov@spbu.ru](mailto:d.koznov@spbu.ru).

---

Поступила в редакцию 23.04.2021, после рецензирования 7.05.2021. Принята к публикации 14.06.2021.

---

## Master data management in an iterative approach

S.V. Kuznetsov<sup>1</sup>, D.V. Koznov<sup>2</sup>

<sup>1</sup>Unidata LLC, Saint-Petersburg, Russia

<sup>2</sup>Saint-Petersburg State University, Saint-Petersburg, Russia

### Abstract

Master Data Management (MDM) is a young area of business informatics that concerns consolidation and centralized control of highly important business data distributed over different information systems. Leading IT companies such as IBM, Oracle, Informatica and others offer a wide range of ready-made products for master data management (MDM products). MDM product deployment involves many technical and organizational complications: it is necessary to adapt these products for the specifics of the business, modify business processes, create new data policies, solve security questions, etc. A popular approach to this task is the iterative strategy of MDM deployment, which supposes a step-by-step implementation of master data management based on the real needs of the business organization. In this paper, the notion of an MDM solution is introduced, which is the result of the deployment of MDM in an organization. It includes a specifically adapted MDM product, new regulations for working with data, trained employees, and an up-and-running process of master data management. The main result of the paper is a functional model of master data management intended for the early stages of the development of an MDM solution within the iterative deployment strategy. The purpose of this model is the representation of real business needs in terms of MDM. It is important to understand which MDM components should be implemented first. The paper describes a detailed description of the model components, as well as a portfolio of six real MDM projects analyzed from the viewpoint of the proposed model.

**Key words:** master data management, business informatics, applied ontology, iterative approach, information systems.

**Citation:** Kuznetsov SV, Koznov DV. Master data management in an iterative approach [In Russian]. *Ontology of designing*. 2021; 11(2): 170-184. DOI: 10.18287/2223-9537-2021-11-2-170-184.

### List of figures and tables

Figure 1 – Functional Model of typical MDM solution

Table 1 – Features of MDM projects

Table 2 – Industrial MDM projects in terms of functional model

### References

- [1] **Khatri V, Brown CV.** Designing data governance. *Communications of the ACM*. 2010; 53(1).
- [2] DAMA-DMBOK: Data Management Body of Knowledge, 2017.
- [3] **Andrichenko AN.** Tendencies and condition in the field of reference data management in the engineering industry [In Russian]. *Ontology of designing*. 2012; 2 (4): 25-35.
- [4] **Nemtsov EF.** Reference data in the intelligent railway transport management system [In Russian]. *Automation, communication and Informatic*. 2020; 2: 15-18.
- [5] **Golubev SS, Lotsmanov AN, Kuzin AY, Soloviev VG, Kozlov AD, Grigoriev BA.** The branch system of the National Standard Reference Data Service for the Oil and Gas Complex [In Russian]. *Legislative and applied metrology*. 2020; 3: 12-16.
- [6] **Yanchenko GA.** On the standard reference data density properties of rocks [In Russian]. *Mining informational and analytical bulletin*. 2011; 8: 111-115.
- [7] **Chigirinsky YL.** Methodology for improving the reliability of reference data [In Russian]. *Izvestia Volgogradskogo universitets*. 2011; 13 (86): 55-61.
- [8] Gartner Glossary, <https://www.gartner.com/en/glossary>.
- [9] **Walker S, Dayley A, Parker S, Hawker M.** Magic Quadrant for Master Data Management. Gartner. 2021.
- [10] **Zmud RW.** An Examination of ‘Push-Pull’ Theory Applied to Process Innovation in Knowledge Work. *Management Science*. 1984; 30 (6): 727–738.
- [11] **Ladley J.** Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program. Academic Press; 2nd edition 2019. 350 p.
- [12] **Silvola R, Jääskeläinen O, Kropsu-Vehkaperä H, Haapasalo H.** Managing one master data – Challenges and preconditions. *Industrial Management & Data Systems*. 2011; 111(1): 146–162.

- [13] **Yourdon E, Constantine LL.** Structured Design: Fundamentals of a Discipline of Program and Systems Design. Yourdon Press. 1975.
  - [14] **Jacobson I.** Object-Oriented Software Engineering. ASM press. 1992. 528 p.
  - [15] **Ould MA.** Business Processes: Modelling and Analysis for Re-Engineering and Improvement. Wiley, 1995
  - [16] CMMI Data Management Maturity Model (DMM). CMMI Institute (website). <http://bit.ly/1Vev9xx>.
  - [17] IBM Data Governance Council Maturity Model <https://ibm.co/2sRfBIn>
  - [18] Velocity Methodology. Best Practices. Informatica. 2008.
  - [19] **O'Kane B, Moran MP.** The Seven Building Blocks of *MDM*: A Framework for Success. August 2016. Gartner. ID: G00311161
  - [20] **Ofner MH, Straub K, Otto B, Österle H.** Management of the master data lifecycle: a framework for analysis. *J. Enterp. Inf. Manag.* 2013; 26(4): 472-491.
  - [21] **Gavrilova TA.** Knowledge presentation in expert system ABTAHTECT [in Russian]. *Izvestia Acadimii Nauk SSSR.* 1984; 5: 165-173.
  - [22] **Gavrilova TA, Kudryvtsev DV.** Knowledge management: from words to business [in Russian]. *Intelligent Enterprise: RE.* 2004; 12-13(101): 48.
- 

### About the authors

**Sergey Viktorovich Kuznetsov** (b. 1979) graduated with honors from the Faculty of Mathematics and Mechanics at Saint Petersburg State University (2001). He was the head of R&D departments in several data management startups, and the chief of the Russian representative office of Informatica (2010). Founded TaskData, LLC and Unidata, LLC, concentrated on development of data management platform (Unidata platform, released in 2014). Unidata master data management system was acknowledged by Gartner in their prominent annual Magic Quadrant reports on *MDM* (2020, 2021). [sergey@unidata-platform.ru](mailto:sergey@unidata-platform.ru)

**Dmitry Vladimirovich Koznov** (b. 1971) graduated with honors from the Faculty of Mathematics and Mechanics at Saint Petersburg State University (1994), received PhD degree (2000), and a Doctor of Technical Sciences degree (2016). Professor of the Software Engineering Department at Saint Petersburg State University since 2016. Member of the ACM. Visiting professor of the National University of Singapore (Singapore, 2000, 2001), Lappeenranta University of Technology (Finland, 2015), and South China University of Technology (China, 2018). Author of over 100 papers in the field of software engineering of information systems, including more than 30 Scopus-indexed articles. Author ID (RSCI): 126261. Author ID (Scopus): 8885649400. [d.koznov@spbu.ru](mailto:d.koznov@spbu.ru).

---

*Received April 23, 2021. Revised May 7, 2021. Accepted June 14, 2021.*

---

## ИНЖИНИРИНГ ОНТОЛОГИЙ

УДК 004.82

DOI: 10.18287/2223-9537-2021-11-2-185-211

### Онтологическая система «знания-деятельность»

**Н.В. Максимов, А.А. Лебедев***Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ), Москва, Россия*

#### Аннотация

Рассматривается подход к построению онтологического описания рациональной деятельности. Подобное описание может служить дополнительным источником для входа в информационное пространство, расширяет класс возможных индексов для описания документов и позволяет осуществлять контекстный поиск. Рассмотрены назначение и особенности использования онтологий при описании предметной области, как с теоретической, так и с прикладной точки зрения. Обозначены критерии для выделения типов онтологий. Исходя из принципов схематизма познания, разработана система онтологий представления знаний, объединяющая язык, формы представления знаний и схемы процессов. Показано, что такая система онтологий даёт возможность практического использования онтологий в вычислительных средах. Для комплексного описания производственной (рациональной) деятельности представлены онтологии артефакта и процессов, разрабатываемые на основе анализа государственных и международных стандартов. Онтология артефакта представляет собой описание объекта – средства и результата целенаправленной деятельности, онтология процессов – формализованное описание жизненного цикла деятельности. С помощью онтологии артефакта становится возможным представить логическую структуру деятельности, а с помощью онтологии процессов – временную структуру, которые совместно определяют две стороны описания сложной деятельности. Предложенная онтология процессов включает иерархически связанную совокупность этапов жизненного цикла деятельности, объекты и субъекты деятельности каждого этапа и исполняет роль «навигационной карты» для субъектов деятельности, позволяя им ориентироваться в сложных организационных структурах.

**Ключевые слова:** онтология, знания, деятельность, онтология артефакта, онтология процессов.

**Цитирование:** Максимов, Н.В. Онтологическая система «знания - деятельность» / Н.В. Максимов, А.А. Лебедев // Онтология проектирования. – 2021. – Т.11, №2(40). - С.185-211. – DOI: 10.18287/2223-9537-2021-11-2-185-211.

#### Введение

Промышленность и наука характеризуются большим массивом накопленной информации по истории функционирования, решениям и их обоснованиям, технологиям и методикам работы и другой информации. Это приводит к изменению характера труда, изменению форм взаимодействий/взаимосвязей участников деятельности, росту числа элементов, образующих систему предприятия. Человек, как часть этой системы, ограничен в возможностях восприятия и обработки поступающих данных. Его возможности и следствия технического развития приводят к задачам унифицированного представления и описания имеющихся знаний в форме, удобной для восприятия и обработки, в т.ч. в вычислительной среде.

Новые стандарты предлагают переход от функциональных моделей к процессно-проектным моделям управления. Важную роль в таких моделях занимают вопросы управления знаниями, в частности, актуальными являются задачи улучшения поиска информации, в том числе за счёт учёта специфики деятельности и представления документации.

В этой связи перспективным направлением для построения унифицированного представления, учитывающего организационную и семантическую составляющую деятельности, является использование онтологий. Унификация здесь основывается на типизировании сущностей и отношений, а также на наблюдаемом схематизме процессов деятельности, обусловленном устойчивой тенденцией, отражаемой, в частности, разработкой отраслевых, государственных и международных стандартов, обобщающих передовой опыт. Онтологический подход к представлению семантики документов позволяет представить не только описываемый в них объект, но и процесс его получения в виде «контурной карты», на которой пользователь реализует траекторию поиска, причём целесообразность ключевых понятий для поиска обеспечивается за счёт визуализируемых путей между рассматриваемыми сущностями онтологии [1, 2].

Целью данной работы является определение онтологии, позволяющей осуществлять отражение знаний о действительности и служить инструментом и стимулом для их развития. Входящие в состав онтологии компоненты должны иметь практическую направленность, быть в достаточной степени универсальными и отражать принятые в различных предметных областях (ПрО) стандарты.

Переход к схематизированному, цельному, концептуализированному (онтологическому) описанию деятельности и перенос их содержания в вычислительную среду позволяет решать задачи в ПрО. В работе [3] отмечается, что «...опыт предприятия является ценным ресурсом, который требуется сохранять, развивать и передавать в наиболее удобной форме для специалистов предприятия. Система накопления опыта в виде нормативной документации имеет ряд существенных недостатков для повышения эффективности творческой деятельности...».

Использование онтологий для решения задач ситуационного управления предприятиями в режиме реального времени рассматривается в [4]. Автором предложено объединять отдельные онтологии, представляющие собой зафиксированные и концептуализированные знания субъектов деятельности (акторов) о своей ПрО. В итоге должна образоваться единая онтология деятельности предприятия, состав и структура которой может изменяться в реальном времени, что способствует своевременному информированию участников деятельности о возникших изменениях в процессах или возникших нештатных ситуациях.

Подход к построению единой теории комплексной деятельности представлен в [5], где деятельность рассматривается с точек зрения различных областей знаний: психологической теории деятельности, общей теории систем, теории организации и др. Указывается, что важными требованиями для обобщённой структурной модели комплексной деятельности являются её иерархичность (для отражения логической структуры) и фрактальность (для отражения временной структуры), а сложность деятельности заключается в проявлениях неопределённости и её преодолении.

Обобщённая функционально-ориентированная схема интегрированного процесса целенаправленной деятельности представлена в [6]. Эта модель – последовательность этапов, поддерживающих специфические работы, которая начинается с исследования и заканчивается производством и коммерциализацией. Каждый этап заканчивается целевым результатом и его оценкой. Графовое представление документов, учитывает не только их содержание, но и контекст, например, место и роль документа в жизненном цикле (ЖЦ) проекта [1].

Человеческое мышление ориентировано на использование специальной организации знаний об отношениях объектов реального мира в виде абстрактных схем. Разнообразие видов деятельности и ограниченность в восприятии и обработке приводят к необходимости унификации возможных схем и выработке стандартов для их представления. В статье рассмотрены: принципы формирования схем деятельности, связь схем с онтологическими моделями действительности, онтологическая модель процессов ЖЦ деятельности.

## 1 О схематизме деятельности и знаний

Упорядочение знаний даёт возможность устанавливать причинно-следственные связи, минимизирует длину причинно-следственных цепей. В частности, задача науки – поиск теорий, удовлетворяющих экспериментальным данным [7].

Одним из способов упорядочивания накопленных знаний субъектом является схема, которая задаёт связь между объектом и понятием. Согласно [8], Кант определяет схематизм как деятельность по соотношению объектов реального мира к мысленным категориям, представляющим собой формы пространственно-временных отношений. Схемы определяют правила синтеза воображения, благодаря схеме и подобно ей становятся возможными образы, но связываться с понятиями они всегда должны только при посредстве обозначаемых ими схем, и сами по себе они совпадают с понятиями не полностью. Рассмотрению роли схем как средства мышления субъекта посвящены работы Ж. Пиаже [9], в отечественной науке исследование роли схем и их связи с деятельностью представлено в работах Г. Щедровицкого и Московского методологического кружка [10]. Подробный анализ использования схем в когнитивных процессах приведён в [11].

Схемы выполняют несколько функций: помогают понять происходящее, организуют деятельность человека, собирают смыслы, способствуют выявлению новых объектов [12]. Структурные схемы «сжимают» содержание, выделяют причинно-следственные и функциональные связи. Схема отображает не форму, а отношения и действия предметов.

Нормативная документация организует деятельность, формируя для субъекта схему деятельности; схемы, организующие деятельность субъекта, сами становятся нормой.

Элементарная модель, иллюстрирующая схему рациональной деятельности и построенная на основе анализа нормативной документации и обобщения субъективного опыта, может быть представлена как продолжающиеся и повторяющиеся взаимодействия, а также обратные связи между ними. Функциональные блоки на каждой итерации имеют одинаковую структуру: получение исходного материала, управляемое параметрами выполнение целевой функции, оценка и передача результата.

Подобная схема, определяющая только базовые моменты, при применении в реальности дополняется конкретностью, позволяет заранее выявить условия возможных проблемных ситуаций, подобрать требуемую документацию, определить критерии успешного завершения цикла деятельности. Анализируя процессы конкретных предприятий и используемую на предприятии нормативную документацию, можно синтезировать на их основе схемы деятельности с различных точек зрения субъектов производства.

Графическая форма визуализации, характерная для схем, удобна для восприятия и позволяет находить взаимосвязи между объектами, обнаруживать совпадения между объектами и событиями, выявлять аномальные объекты. Онтологии могут играть роль предопределённой потенциальной траектории расширения поискового пространства и использоваться как шаблон для смежных тем проекта [1].

Рациональная деятельность включает в себя цели, результаты и средства их достижения. Субъект, вступая во взаимодействие с окружающей действительностью, используя инструментальные средства, работает с некоторым множеством объектов для достижения поставленной цели. «Поведение человека ... подразумевает какую-то цель. Человеческое стремление приписывать действиям и событиям смысл и цель подразумевает расчленённость реальности на некоторые условные сегменты ... Осмысление связано с сегментацией недискретного пространства» [13]. С этим коррелирует позиция В.В. Налимова: «Пространство, на котором заданы (существуют) смыслы, имеет физический характер, т.е. непрерывно, и должно допускать вариабельность. Таким образом, семантическое поле «проявляется» (а не членится) и фиксируется через задаваемую на нём величину (метрику)» [14].

Наличие цели, как объективизированной через свойства потребности, приводит к определению ограничений на результат. Наличие свойства ценности – к определению критерия достижимости цели. Наличие метрик выделенного семантического поля позволяет измерять и опосредованно сопоставлять величины, а также фиксировать события - контрольные точки, обеспечивающие возможность эффективного управления, в которых можно получать измеримые результаты для сравнения с заданными значениями. Отсюда следует, что деятельность может «разбиваться» на отдельные работы с отдельными множествами объектов и заканчивающиеся отдельными результатами.

Рациональная деятельность в своей основе имеет природу информационных взаимодействий: операционными объектами деятельности (управления) являются образы – информационные объекты со свойствами. «Проявление» семантического поля (выделение объектов, причинно-следственных связей, определение свойств) возможно, поскольку реальный мир преломляется сознанием субъекта на основе известных ему схем, путём «встраивания» получаемых данных в имеющиеся модели реальности, или их дополнение/перестраивание.

Особенностью деятельности в обществе является распределённость и неоднозначность смыслов (полисемия): объёмы и смыслы понятий у субъектов могут не совпадать, поскольку субъекты могут иметь разный контекст, опыт и мотивации. Условием успешности взаимодействия между субъектами является требование соотносимости множеств объектов, понятий и знаков, а также операций над элементами внутри и между множествами в сознаниях субъектов. Перенос знаний между субъектами происходит посредством языка, где объекты реального мира замещаются объектами знаковой природы, а схемы взаимодействий между реальными объектами заменяются схемами взаимодействий между знаками. Частным случаем переноса знаний являются процессы рефлексии и синтеза знаний внутри субъекта. Подобную соотносимость возможно обеспечить путём типизации и унификации используемых понятий, а также явного учёта контекста их использования.

## 2 Процессы и ЖЦ

Деятельность, как система, в своём развитии проходит через определённые состояния (фазы, стадии, этапы). Концепция ЖЦ представляет собой основной вариант реализации системного подхода к сложным техническим объектам, направленный на отображение измененных состояний этих объектов в некоторый период времени [15].

Рост стоимости производства приводит к задачам управления процессами как на отдельных этапах ЖЦ, так и в целом. Здесь онтологические модели выступают в роли инструмента интеграции разнородных знаний о процессах, ресурсах и объектах, задействованных на разных этапах производства:

- «...онтология позволяет не просто аккумулировать наши знания, а зафиксировать опыт так, чтобы была возможность реализации его представления в разных формах об объектах рассматриваемой ПрО, и зафиксировать так, чтобы этими знаниями мог воспользоваться и интерпретировать человек... и вычислительная техника...» [16];
- «...онтологии для управления знаниями предприятия на базе модели процессов предприятия позволяют сотрудникам одной отрасли или корпорации использовать общую терминологию и избежать взаимных недоразумений...» [17];
- «...энергетика - это та область, где необходимость онтологического моделирования является более чем очевидной...» [18].

Общий подход к процессному описанию ЖЦ систем, представленный в [19], специфицирует типы процессов: процессы соглашения, организационного обеспечения проекта, технического управления и технические процессы. Подобная классификация тесно связана с про-

стейшей моделью системы, где внешнему наблюдателю доступны только входные и выходные связи системы со средой. Высокий уровень абстракции модели этого стандарта, определяя широкие возможности для применения, позволяет свести многообразие возможных видов деятельности к ограниченному множеству объектов и отношений [20].

Для конкретных производств характерно наличие специфических процессов, полное знание которых доступно только специалистам. Отсюда возникает задача типизации разнородных процессов, что позволяет отразить как специфику ПрО, так и общий характер процессов. Например, процессы ЖЦ для программных средств (ПС), кроме процессов ЖЦ системы включают также специфические для данной области процессы реализации ПС, поддержки ПС и повторного применения ПС [21]. Процессы повторного применения ПС включают процессы проектирования доменов, менеджмента повторного применения активов и программ.

Единое описание общих и специализированных процессов ЖЦ должно опираться на нормативно-регламентирующую базу, описывающую данный вид деятельности, где каждый стандарт является онтологией некоторой ПрО.

Похожее решение предложено в [22]: «Каждый стандарт рассматривается как онтология соответствующей ПрО, являющаяся основой для автоматизированного решения ряда задач, включая информационное обслуживание сотрудников, формальную оценку соответствия предприятия этим стандартам и т.д... В основе системы формальных онтологий для предприятий рецептурного производства лежит формализация стандарта ISA-88, который представляется в виде системы онтологий».

Рассматривая деятельность на различных организационных масштабах с учётом цели и характера производимых работ, можно прийти к некоторой взаимосвязанной сети, позволяющей свести имеющееся разнообразие к конечному набору типовых элементов (рисунок 1).

Процеируя различные процессы и технологические операции на этот фиксированный базис, имеется возможность перейти к их типизированному и схематизированному описанию. На таком базисе можно отобразить любую рациональную деятельность по характеру и степени общности, и проследить её временную трансформацию.

Двигаясь снизу-вверх можно ответить на вопросы «что делать?», «когда делать?», «в каких процессах используется полученный результат?», «каков этап ЖЦ деятельности?», «каков вид деятельности?», «какой характер должен быть у полученного результата?». Появляется возможность для информационного поиска в производственной и проектной документации. Для ответа на вопросы «кто/что является объектом деятельности?», «кто/что является субъектом деятельности», «какими средствами осуществляется деятельность?», «что является результатом деятельности?», «что является целью деятельности?» можно использовать онтологию артефакта. Например, операция «подписать техническое задание» относится к процессам соглашения и может быть произведена на этапах замысла или разработки. При этом вид деятельности (если он не указан) необходимо определять по содержанию технического задания.

Использование стандартов позволяет отождествить каждый элемент с набором типовой лексики (*Заказчик, Исполнитель, План, Разработать, Утвердить, Подписать* и т.п.), сформировать тезаурус, который становится основой для автоматического соотнесения возникающей в ходе деятельности документации к этапам ЖЦ и исполнителям. Система выделенных понятий и отношений, к которым применены правила композиции, определённые согласно рисунку 1, образуют онтологию процессов.

Поскольку отдельные процессы в рамках деятельности также являют собой деятельность, очевидно, что онтологическое описание процессов ЖЦ деятельности будет представлять собой сложную конструкцию, где каждая из множества составляющих граф вершин может состоять из подмножества других вершин.

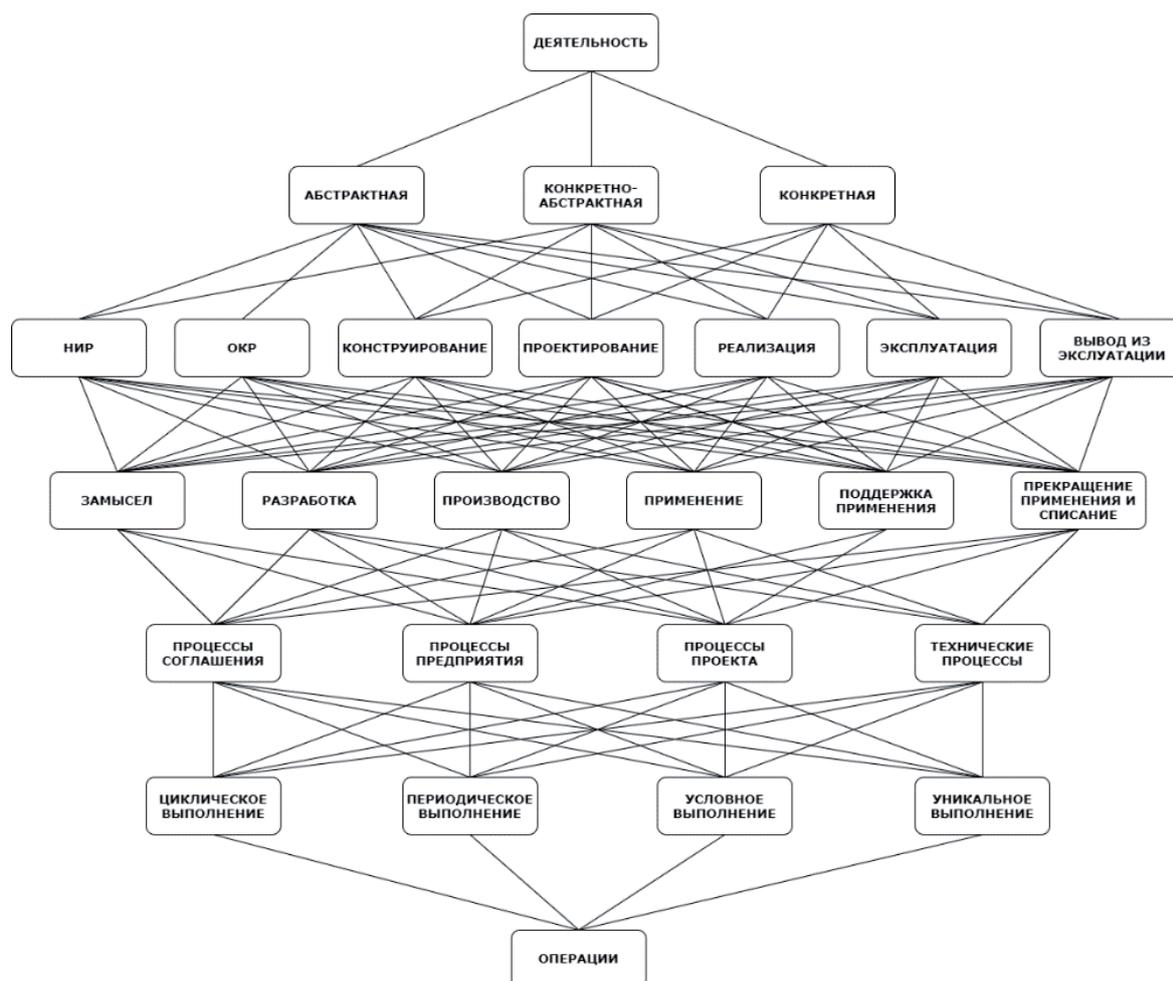


Рисунок 1 — Связь различных аспектов деятельности с точки зрения характера работ, цели и времени выполнения

В таком подходе, в отличие от модели, использующейся, например, в стандарте *Process Specification Language* [23] или онтологии бизнес-процессов *COBRA* [24], нет элементов, отвечающих за временные соотношения начала и окончания процессов (что было бы полезно для построения диаграмм Ганта или сетевых графиков).

### 3 Обобщённая онтологическая модель

Известны разные определения и типологии онтологий, обоснованность которых не всегда очевидна. Известны также практически значимые и долгоживущие проекты онтологий, в частности, *DOLCE*, *BFO*, *SUMO*, онтология *J.F. Sowa*. Чтобы конструктивно подойти к типологизации онтологий, необходимо явно постулировать их назначение, а также семантику и прагматику построения и применения. Это позволит выявить основные типологические признаки конкретных онто-продуктов и онто-конструкций.

- Онтология является «*промежуточной*» семантической формой, занимающей место между конструктивной теорией и описаниями на естественном языке, представляющими содержание этой теории.
- Онтология как концептуальный образ фрагментов действительности играет роль «*принципиальной схемы*», представляющей *существо (смысл)* фрагмента в базе понятий и отношений, соответствующих семантике ПрО.

- Фрагментом действительности (объектом онто-описания) может быть:
  - конкретная ситуация – состояние абстрактных или конкретных объектов и их поведение, связанные с целевой деятельностью субъекта;
  - абстрактно-конкретная понятийно-знаковая система, как инструмент познания.
- Онтология как структура, отражающая относительность и конструктивность знания, построена по схеме «язык-объект - метаязык», т.е. обязательно включает метакомпонент.
- Онтология является вторичным объектом и может использоваться как:
  - формализованный образ текста некоторого описания, обеспечивающий содержательный анализ, в частности, методами графов;
  - формализованный образ схемы навигации по понятиям и аспектам, дополняемым семантической окрестностью, формируемой при полнотекстовом семантическом поиске;
  - комбинаторное пространство для синтеза путей при поиске возможных зависимостей.

### 3.1 О типологии онтологий

Можно выделить следующие виды онтологий: прикладные онтологии, онтологии области знания, онтологии метауровня. В подходе, предложенном в [25], онтология - это наборы метаданных, идентифицирующие содержания ресурса и сервисов в нескольких аспектах: описание объектов, их свойств и отношений ПрО; описание задач и процессов, их свойств и отношений; описание атрибутов знания; описание причин и оценок использования, а также характеристики значимости контента и история обращений отдельного субъекта к информационным ресурсам.

В [26] в составе онтологии выделены следующие категории (типы объектов, представляемых онтологией): *Наблюдаемые объекты/Свойства, Типы задач, Знания/Гипотезы, Методы решения, Примеры решаемых задач.*

На рисунке 2 приведена систематизация онтологий по классификационным признакам: уровню представляемых знаний, масштабу применимости и средствам представления [27].

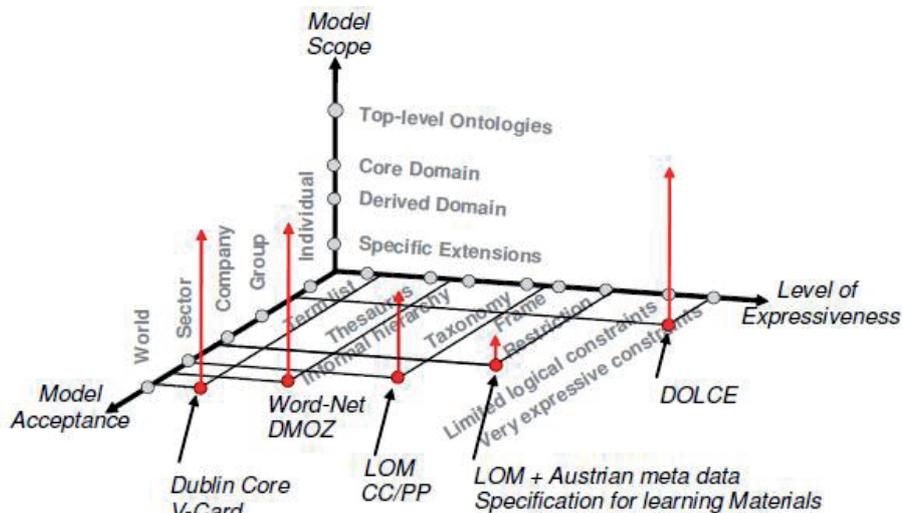


Рисунок 2 — Систематизация онтологий по уровню представляемых ими знаний (*scope*), масштабу применимости (*acceptance*), средствам представления (*expressivity*) [27]

Отдельно выделяют лингвистические онтологии (*WordNet*, *MikroKosmos*, *PyTез* и другие), фиксирующие понятия (слова и словосочетания) вместе с их языковыми свойствами и отношениями (синонимия, гипонимия и т.п.).

Представляется полезным обратиться также к «качественным» определениям. Например, в [28] упоминаются следующие уровни онтологии:

- формальная, включающая формальную логику суждений и аксиоматику;

- описательная – сведения о множестве объектов и их взаимосвязях;
- формализованная – кодифицированное представление онтологий формального или описательного уровней.

Здесь первые два вида выделяются по *типу используемых переменных*: в первом случае – это математические, во втором – лингвистические переменные. Третий вид соотносится с двумя первыми по *способу представления*: первые представлены в непосредственной (естественной) форме, последняя – в опосредованной «закодированной» форме.

В рамках задач извлечения/синтеза знаний [29] онтология, как качественное представление ПрО, включает:

- систему понятий ПрО, в которой формулируется и интерпретируется эмпирическая теория;
- свойства, признаки, величины и соответствующие измерительные процедуры, интерпретируемые в системе понятий;
- априорные и экспертные знания;
- знания, интерпретируемые в системе понятий ПрО, получаемые в процессе построения логической, количественной и конструктивной эмпирических теорий.

Порядок упоминания объектов отражает скорее эмпирическую сторону построения модели ПрО: вначале, исходя из прагматики проблемной ситуации, определяется семантика ПрО, которая и будет средством интерпретации построенной теории.

Обобщение разных видов онтологий, отражающее соответствие типам деятельности, масштабам применимости и типам используемых концептов, представлено на рисунке 3.

Характер (сфера)	Тип деятельности	Уровни сложности объектов	Концепты	Вид онтологий
Теоретический (модель)	Философский (метафизический)	Общенаучный универсум	Категории сущности/отношений	Метаонтология Онтология верхнего уровня
	Фундаментальные научные исследования	Отрасль	Величины/единицы измерений Понятия Свойства Законы	Онтология артефакта Онтология отрасли
	Прикладные научные исследования	Область	Теории Функции Параметры	Онтология ядра (фундаментальная) Прикладная онтология
Практический (практика)	Опытно-конструкторские работы	Технологические системы	Объекты Процессы	... модели предметной области
	Производство, эксплуатация	Компоненты	Экземпляры Связи	... практики предметной области ... интернета вещей

Рисунок 3 — Виды онтологий по типам деятельности

*Онтология мета-уровня* представлена концептуальным каркасом, определяющим классификационную взаимосвязь производных понятий.

*Онтология верхнего уровня* включает фундаментальные понятия и отношения, такие как пространство, время, материя, событие, действие и т. д. Такие онтологии строятся на концептуальном каркасе, задаваемом на мета-уровне. Отдельно надо указать онтологию величин и единиц измерения, фундаментальную для естественных наук и производства. Аналогично, для языка фундаментальными будут таксономии сущностей и отношений.

Поскольку может существовать несколько онтологий верхнего уровня, это означает, что различные онтологии верхнего уровня должны быть сопоставимы, основой для чего и может служить концептуальный каркас.

*Онтологии уровня отрасли* и области знаний включают понятия/отношения, как общие, так и специфические для конкретной ПрО. В соответствии с методологией и организацией процессов познания, по уровню сложности объекта концептуализации выделяются:

- *фундаментальная онтология*, специфицирующая через «ядерные» понятия и отношения семантику отрасли, общую для ПрО, выделяемых в отрасли знаний;
- *прикладная онтология ПрО*, определяющая на ядре (через «родительские» фундаментальные отраслевые понятия) производные специальные понятия и отношения, которые в рамках этой области будут конкретны и специфичны.

Объектами онтологий прикладного уровня могут быть не только абстрактные понятия/отношения, которым соответствуют *онтологии моделей ПрО*, но и конкретные экземпляры (их образы) – реальные экземпляры и связи, которым соответствуют *онтологии практики ПрО* – онтологии конкретных объектов/процессов/проектов. Такие онтологии обычно создаются для Интернета вещей, больших систем и проектов в энергетике, авиастроении и т.п.

Прикладная онтология может включать компоненты по признаку фазы знаний или вида объекта концептуализации. Например, онтологии задач, методов, эффектов, свойств и мер.

В целом, отдельная онтология соотносится с вышестоящей онтологией как «язык-объект» и «мета-язык». Изначально понятия ПрО определяются через понятия мета-языка, но после этапа становления области большинство понятий определяется через понятия ПрО, которые были введены и существуют вполне самостоятельно. Понятие относится к «лингвистическим» переменным: его смысл определяется не только наследуемым содержанием, но и контекстом – обстоятельствами определения и употребления. Для соотнесения понятий или оценки семантической целостности необходимо иметь возможность «приведения» понятий или отношений к мета-языку, что в пределе приведёт к «проявлению» его категорийного образа. Поскольку категории являются общим базисом, то это позволит определять вычислимые меры и семантические расстояния.

Иерархия онтологий представляет собой «стек онтологий», где каждая онтология наследует семантику «родителя» [30, 31]. Это позволяет возвращать результат применения, на основе обратной связи создавать основу для развития стека.

### 3.2 Формы представления онтологий

Онтологии ассоциируются с разными видами и формами их представления: словарями, тезаурусами, таксономиями и классификациями, фреймами, графами, семантическими сетями (см., в частности, рисунок 1), имеющими разные структуры и обладающими разной смысловой выразительностью. Перечисленные формы антропоцентричны и ассоциируются с экспертными способами их формирования и использования с непосредственным участием человека. Автоматизированные же способы предполагают строго формализованное представление и использование вычислительных методов. В связи с этим предлагается различать три формы представления: декларативную (уровень представления), логико-семиотическую (логический уровень), даталогическую (физический уровень).

**Декларативная форма.** Это классические формы концептуализации, относительно простые, достаточно формализованные, наглядные терминологические структуры, отражающие, прежде всего, имманентные отношения между устойчивыми, общепринятыми в ПрО понятиями. Они являются, с одной стороны, базисом для концептуализации конкретных ситуаций, решений и выводов, с другой – инструментом смыслового анализа.

Наиболее полно семантика описываемого объекта/процесса представляется семантическими сетями. Семантические сети представляют факты, ситуации, решения ПрО через граф сущностей и функциональные отношения.

Таксономии и классификации представляют членение доменов ПрО в соответствии с концептуальной решеткой, что обеспечивает наследование свойств, обычно через древовидную форму представления, и позволяет определять меры семантической близости. Классификации и таксономии идентифицируют существо объекта неявно - через свойства, характерные для соответствующего класса и относительно других классов.

Словари и тезаурусы представляют в традиционном и простом виде понятийный базис – средство построения описания объекта/ситуации.

Таксономии и тезаурусы с определёнными оговорками могут быть отнесены к лингвистическим онтологиям.

**Логико-семиотическая форма.** Формализованное структурное представление онтологии, как семиотически целостное образование, определяется в [32] в рамках общей теории систем [33] как совокупность трёх взаимосвязанных систем

$$O = \langle S_f, S_c, S_t, \equiv \rangle, \text{ где}$$

$S_f$  – функциональная система, представляющая объекты и связи действительности и определённая как  $S_f = \langle M_f, A_f, R_f, Z_f \rangle$ , где  $M_f$  – множество объектов (сущностей),  $A_f$  – множество характеристических свойств,  $R_f$  – множество функциональных отношений, представленных типизированными ситуативными связями, характерными для ПрО,  $Z_f$  – закон композиции, т.е. правила и схемы упорядочения объектов (например, мерономия ПрО);

$S_c$  – понятийная система, определённая как  $S_c = \langle M_c, A_c, R_c, Z_c \rangle$ , где  $M_c$  – множество понятий ПрО,  $A_c$  – множество признаков систематизации понятий,  $R_c$  – множество отношений (прежде всего, парадигматических),  $Z_c$  – закон композиции (схема представления, в частности иерархическая таксономия);

$S_t$  – терминологическая система, определённая как  $S_t = \langle M_t, A_t, R_t, Z_t \rangle$ , где  $M_t$  – множество терминов,  $A_t$  – множество свойств,  $R_t$  – множество отношений (эквивалентности, включения, а также лингвистические отношения),  $Z_t$  – закон композиции (грамматика, правила выделения понятий/отношений из текста);

$\equiv$  – операция сопоставления элементов различных систем на уровне знаков, обеспечивающая их тождество в функциональной, понятийной и терминологической системах.

Такая формализация позволяет задавать теоретико-множественными методами компоненты конкретных онтологий.

**Даталогическая форма.** Представление онтологии в виде графов позволяет формализовать операции над онтологиями на основе теоретико-графовых аксиом. Основными операциями, согласно [32, 34], являются: бинарные – объединения, пересечения, проекции; унарная – масштабирования онтологий.

Моделью данных функциональной<sup>1</sup> системы является помеченный (для вершин и дуг которого определены свойства  $A_f$ ) ориентированный<sup>2</sup> граф  $G(V, E) = \langle V, E \rangle$ , где  $V$  – множество вершин, а  $E$  – множество дуг. Множества вершин и дуг в совокупности соответствуют множеству элементарных фактов, представляющих соответствующие сущности, связанные типизированными отношениями. Одна и та же пара сущностей может участвовать в нескольких элементарных фактах, в графе  $G(V, E)$  для двух вершин может существовать более одной дуги, т.е.  $G(V, E)$  обладает свойством мультиграфа.

<sup>1</sup> Объектами функциональной системы могут быть понятия и знаки, что позволит иметь онтологию языка.

<sup>2</sup> Ориентированность графа онтологии определяется ориентированностью дуг и смысловой «направленностью», отражающей эволюцию смыслового образа объекта/результата. Это означает, что имя отдельной сущности или отношения в графе будет представлено в нескольких «экземплярах». Здесь понятия и отношения выступают как лингвистические переменные, конкретный смысл которых доопределяется контекстом, специфицируемым типами отношений и характером связанных сущностей. Поскольку граф представляет *целенаправленный процесс*, то для него определён порядок следования вершин, в том числе могут быть заданы исходные и конечные (целевые) вершины, что будет соответствовать понятию сеть.

На множествах  $V$  и  $E$  определяются и могут быть динамически построены:

- *Метаграф*, формально задаваемый как  $MG = \langle V, MV, E, ME \rangle$ , где  $V$  – множество вершин,  $MV$  – множество метавершин,  $E$  – множество дуг,  $ME$  – множество метадуг. Каждая метавершина соответствует метасущности ситуативного факта и представляет собой граф  $mv_i = \langle V_i, E_i \rangle$ , где  $V_i \subset V, E_i \subset E$ , а метадуга – метаотношению. Для графовых форм, отображающих семантику текстов, наличие метавершин вполне конструктивно и естественно. Метавершина соответствует сущности и выступает в качестве атомарного семантического эквивалента смысла, определяемого неатомарной конструкцией (выражением). Метаграф также обладает свойством мультиграфа, т.к. ситуативные факты могут, например, различаться только метаотношениями.
- *Гиперграф*, когда на множестве вершин  $V \cup MV$  формируется множество гиперребер  $W$ , при этом в основе правил задания гиперребра лежат множества  $A_f$  и  $Z_f$ :  
 $W = \{w_1, w_2, \dots, w_n\}, w_i = V_i \cup MV_i$ , где  $V_i \subset V, MV_i \subset MV; V_i \neq \emptyset \vee MV_i \neq \emptyset$ .

Смысловая направленность является важным фактором адекватности восприятия. Исходя из того, что схема восприятия информации предопределяется характером задачи и её когнитивным состоянием, при визуализации графа можно использовать разные варианты укладки вершин (закон композиции), которые будут определять характер «навигации» по графу. В частности, используются следующие алгоритмы укладки вершин на плоскости [35]:

- укладка вершин в соответствии с некоторой схемой действий, например, в соответствии с функциональной моделью *IDEF0*;
- построение пути между двумя указанными сущностями, обеспечивающее представление цепочки фактов;
- укладка вершин в порядке употребления сущностей в тексте;
- укладка вершин в соответствии с упорядоченностью по значимости, где в качестве значимости используется длина пути или суммарный вес вершин;
- укладка вершин методом решения гравитационной задачи  $n$ -тел.

Использование различных вариантов укладки вершин позволяет повысить уровень восприятия и глубину понимания проблемы.

Наличие в функциональной системе онтологии помимо множеств сущностей и функциональных отношений множества характеристических свойств и закона композиции позволяет группировать сущности не только «в динамике», например, по принципу соответствия синтезируемой цепочки фактов, но и «в статике», например, по принципу обладания общим свойством, по лексикографическому включению и т.п.

Сформированный таким способом граф является *рефлексивным* образом существующего решения/состояния проблемной ситуации. А поскольку рефлексия здесь – это отображение содержания текста, представленного концептуальным графом, на проблемную ситуацию субъекта, то такой граф представляет и некоторое видение будущего.

В графической форме нивелируются различия между смыслом содержания и спецификой представляющих его лингвистических конструкций и интерактивно выполняемые операции над графами позволяют «соединить» абстрактные операции и конкретные схемы. Это облегчает восприятие, понимание и оценку содержания, потому что в фокусе внимания будет компактная целостная картина.

### 3.3 Система онтологий представления знаний

Важнейшим фактором современности является перемещение информационной составляющей всех направлений в вычислительную среду, что предопределяет необходимость формализованного представления неявных и ассоциированных знаний, а также механизмов

сознания. Индивидуально-коллективные процессы деятельности и, в частности, процессы синтеза и использования знаний, можно подразделить на две взаимосвязанные сферы:

- сфера основной целевой деятельности, реализуемой в соответствии с некоторой моделью основной деятельности (ОД);
- сфера информационных коммуникаций, обеспечивающих организацию и управление деятельностью, сохранение и передачу знаний. Это соответствует понятию «информационная деятельность» (ИД), которая и сама может быть объектом ОД.

В подавляющем числе публикаций ОД представляется стеком онтологий отрасли. Однако собственно деятельность, включая организацию и управление ею, является сложным объектом, который воплощает и порождает знания, в том числе «управленческие», что влияет на характер видов и наполнения соответствующих документов, представляющих знания ОД. Деятельность, как фактор процесса построения конкретной онтологии ПрО, должна быть представлена в сопоставимой онтологической форме. Такая онтология включает семантические сети типизированных процессов и таксономию ЖЦ.

Поскольку ОД обычно ассоциируется с рациональной и/или научной деятельностью, конечной целью которой является создание некоторого нового артефакта, на верхнем уровне стека целесообразно иметь онтологию артефакта – концептуальную схему, которая определяет систему взаимосвязанных понятий, обеспечивающих согласованное проведение измерений свойств и оценку результата, а также их унифицированное представление.

Информационные коммуникации – это, во-первых, язык, основной функцией которого является фиксация знаний на носителях, обычно в виде документа. Во-вторых – это формы представления знаний, зависящие от характера работы и назначения документа. В-третьих – это информационная поддержка процессов организации и управления деятельностью (схемы типовых процессов, содержание работ и т.п.). Эти компоненты (язык, формы представления знаний, схемы процессов) являются существенными и устойчивыми составляющими, определяющими представление и использование знаний.

Основными компонентами онтологии языка являются: понятийно-знаковые системы (тезаурусы, глоссарии), словари основ и морфо-шаблоны именованных сущностей и типизированных отношений, таксономии сущностей/отношений, таксономия величин/единиц измерения.

Основными компонентами онтологии деятельности являются семантические сети типизированных процессов и таксономии ЖЦ. Такая онтология представляет собой следующую ступень развития качества познания действительности в ряду: категоризация объектов в пространстве-времени, построение схемы отношений объектов/категорий, выделение типичных ситуаций на основе схем, и, наконец, онтологии процессов, как объединение различных ситуаций в контексте деятельности. Онтологии процессов представляют отдельные типизированные «элементарные» акты, например, «получение задания», «согласование документа», «проверка результата» и т.п.

Основными компонентами онтологии форм представления являются таксономия видов документов и схемы представления содержания документов.

Приведённую типологию онтологий нужно дополнить онтологиями языка, форм представления знаний и онтологией процессов деятельности. На рисунке 4 показана *система онтологий* представления знаний. Такая совокупность онтологий обеспечит практическую возможность переноса знаний в вычислительную среду.

Для онтологий языка, деятельности и форм представления не указаны метакомпоненты этих онтологий, поскольку здесь они рассматриваются как среды, где не порождаются «себе подобные» объекты: здесь нет соотношений типа «язык-объект» - «метаязык».



Рисунок 4 — Система онтологий представления знаний

#### 4 Онтологии артефакта

Любая рациональная деятельность начинается с объявления цели деятельности и критериев её достижения (т.е. специфицируются требования к функциональным и эксплуатационным свойствам) и заканчивается проверкой соответствия достигнутых показателей заявленным. В процессе деятельности величины заявленных параметров и формулировки, их задающие, могут измениться. Могут измениться и средства проверки достижимости требований. Требования, как и любые описания, создаются человеком и используются им, и даже хорошо формализованный стиль (если это не спецификации) использует естественный язык со своими свойствами неоднозначности и вариантности, что создаёт возможность для ошибок интерпретации и использования. Другим следствием человеческого фактора является «недоговорённость» - ориентация на человеческое понимание, использующее профессиональные знания получателя: обычно указывается, например, название свойства и его значение, но каким прибором измерять или какие погрешности измерения допускаются, в тексте может и не приведено.

В контексте задачи переноса информационной составляющей деятельности в вычислительную среду и автоматизации управления можно выделить два направления: (1) задача выявления и идентификации в тексте свойств и их значений; (2) «добраивание» параметра до полного образа. Первое связано с проблемой согласованного представления свойств, непротиворечивостью значений свойств в разных частях текста и в разных частях документации. Второе - с корректностью использования единиц измерения, обозначений и т.п. На этапе эксплуатации могут возникать вопросы, далекие от форм представления и именованного того или иного свойства. Например, что будет, если крепёж для заменяемого узла будет из другой марки стали? Что произойдет, если значение некоторого параметра выйдет за предел допустимого на 10%? Безопасно ли использовать топливо с иным

химическим составом? Чтобы найти ответы на такого рода вопросы, нужно проводить информационное расследование, обращаясь к проектной и исследовательской документации. В связи с этим возникает частная задача разработки онтологии артефакта – объекта, рассматриваемого как результат целесообразной деятельности. Это поможет «восстановить полную картину» - привести в явной форме то, что представлено в документации другого этапа ЖЦ или является профессиональными знаниями человека.

Для разработки онтологии артефакта используется системный подход и семиотическая модель, что позволяет построить онтологическую модель объекта, рассматриваемого как результат рациональной деятельности, включающей не только совокупность свойств и методов их оценки, но и «сопутствующий» контекст, который в действительности часто является определяющим.

*Артефакт* – это целевой абстрактный или конкретный объект/процесс, целенаправленно создаваемый субъектом, взаимодействующим с действительностью. Артефакт - это макро-объект, обладающий измеримыми свойствами и включающий «технологии» оценки значений этих свойств.

Свойства артефакта имеют разную природу и представление в разных ПрО, методы измерения зависят от окружения, в котором измеряются величины. Исходя из этого, в онтологии должны быть представлены как фундаментальные связи величин, свойств, единиц измерения и т.д., так и связи, которые показывают соотношения этих понятий. Кроме того, необходимо учитывать семиотическую природу онтологии: помимо объектов и связей действительности, а также понятий и отношений, позволяющих моделировать эту действительность, онтология имеет лингвистическую составляющую – знаковую систему, позволяющую согласованно и гибко представлять информацию.

Концептуальная схема базовых абстракций онтологии свойств основывается на категории *Осознаваемое*, позволяющей рассматривать информационный образ (в частности, значение величины измерения) как отображение конкретного объекта на некоторую модель. Для концептуализации артефакта можно использовать нормативно-справочные материалы в области организации и управления разработками, онтологию свойств [36], а также на результаты анализа следующих проектов.

- Онтология «*Quantities, Units, Dimensions and Types*» (*QUDT*) [37], как часть информационной архитектуры *NASA*, представляет собой формализацию концепций метрологии высокого уровня. Она включает такие компоненты, как величина, единица измерения, размерность, префикс, системы величин и единиц, род величин и обеспечивает связь единиц измерения разных систем.
- *Ontology of Units of Measure* [38] предназначена для использования в научной и инженерной практике и включает следующие компоненты: величина, шкалы измерения, размерность, префиксы, единицы измерения, область применения, и род величины. Область применения определяет употребление разных единиц измерения одной величины, а род величин показывает иерархию величин.
- *The Units Ontology* [39] предназначена для описания качественных и количественных наблюдений в биологии. Онтология включает такие компоненты, как величины, единицы измерения и префиксы.
- Язык *UnitsML* [40] предназначен для кодирования научных единиц измерения и состоит из нескольких наборов элементов: единицы измерения, величины, размерность и префиксы для единиц измерения.

В качестве логико-семантической основы для построения онтологии артефакта можно использовать понятия и определения, принятые в метрологии. На рисунке 5 представлена онтологическая модель (концептуальная схема), представляющая информационный образ

артефакта как систему функционально и логически взаимосвязанных понятий, величин, объектов и процессов, построенная с использованием следующих определений.



Рисунок 5 — Онтология артефакта

*Объект* – это такой фрагмент объективной или субъективной реальности, на который направляется внимание субъекта, вовлекается им в свою деятельность и становится предметом его теоретической или практической деятельности. Предметом или ПрО является совокупность свойств объекта, которая вычленяется познавательными или практическими средствами.

*Свойство* выражает такую сторону предмета, которая обуславливает его различие или общность с другими предметами и обнаруживается в его отношении к ним [41]. Можно сказать, что свойство – это проекция объекта на предмет. Свойство имеет наименование и обозначение.

*Величина* проявляется при взаимодействии как мера, определяемая свойством и предметом. Система величин - это совокупность величин и непротиворечивых уравнений, связывающих эти величины [42].

*Мера* – это средство измерений, которое воспроизводит в процессе использования или постоянно хранит величины одного или более данных родов с приписанными им значениями. Мера имеет единицу измерения [42]. В целом, мера показывает связь качественных и количественных характеристик: «мера показывает границу, за которой изменение количества влечёт изменение качества» [43].

*Размерность* – это выражение зависимости величины от базовых величин системы величин как произведения степеней множителей, соответствующих базовым величинам, без учёта какого-либо числового множителя. Размерность величины определяется через основные единицы. *Единица измерения* – это реальная скалярная величина, определённая и принятая по соглашению, с которой любое другое количество того же рода можно сравнить, чтобы выразить отношение двух величин в виде числа [42].

*Значение величины* – это выражение размера величины в виде некоторого числа принятых единиц или чисел, баллов по соответствующей шкале измерений [44]. Исходя из определения «показатель качества продукции» [45], можно принять, что значение величины определяется посредством инструмента путём измерения показателя – количественной или качественной характеристики свойств системы, рассматриваемой применительно к определённым условиям её создания и эксплуатации или потребления.

*Инструмент измерения* – это устройство, используемое для проведения измерений, отдельно или совместно с дополнительными устройствами, для которого может быть специфицирован интервал измерений, устанавливаемый в технических условиях. *Результат измерения* - значение некоторой величины, полученное применением инструмента измерения [42].

*Результат испытаний* - значение, полученное в одном или нескольких испытаниях в зависимости от требований, регламентированных в документе на метод испытаний [46]. Получение значения основывается на критерии качества, задающем условие проверки соответствия требованиям.

*Верификация* - подтверждение на основе представления объективных свидетельств того, что установленные (функциональные) требования были выполнены [47].

*Валидация* - подтверждение на основе представления объективных свидетельств того, что требования (эксплуатационные) использования или применения выполнены [47].

*Условия испытаний* – совокупность воздействующих факторов и (или) режимов функционирования объекта при испытаниях [48].

## 5 Примеры использования нормативных документов как основы для построения онтологической схемы ЖЦ производства

Примеры построения и использования онтологического представления рассмотрены для этапов разработки технического задания (ТЗ), научно-исследовательских работ (НИР) и опытно-конструкторских работ (ОКР). Выбор этих этапов обусловлен тем, что с них рекомендуется начинать ЖЦ производственной деятельности, а результаты, полученные в ходе выполнения НИР, должны войти в ТЗ для ОКР. Для примера взяты ГОСТы серии 15 (Система разработки и постановки продукции на производство). Из текстов стандартов были выделены тройки «сущность-отношение-сущность», которые объединялись на основе общности во времени выполнения или характера работ. На следующем шаге оставшиеся после объединения тройки представлялись в виде графа (в общем случае обладающего свойствами мульти- и гиперграфа), где отдельная вершина или дуга могут рассматриваться как точка входа в информационное пространство и позволяют раскрыть содержание этапа.

Тройки выделяются автоматически, однако на рисунках 6-8 представлены обработанные графы, где вручную были выделены метавершины. Атомарные вершины графа содержат имена объектов и субъектов этапа деятельности (наименования документов, участников работ и т.п.), а характерные «временные» участки работ представлены метавершинами. Поскольку тексты стандартов представлены на естественном языке, в некоторых случаях достраивались связи и вершины, не названные явно, но следующие из контекста.

Правила формирования метавершин определяются решаемыми задачами и нормативными документами. Например, рассматривается структура вершины или отбираются однотипные вершины и отношения.

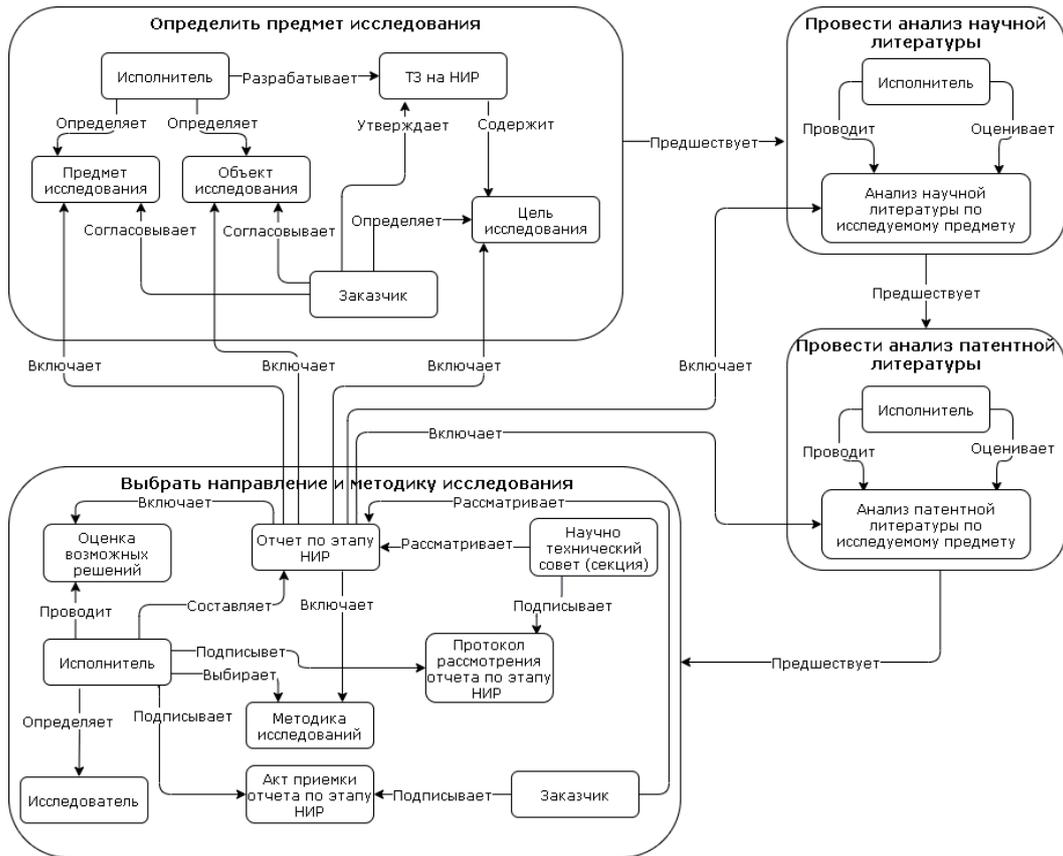


Рисунок 6 — Фрагмент метаграфа, построенный на основе текста стандарта, регламентирующего порядок выполнения НИР

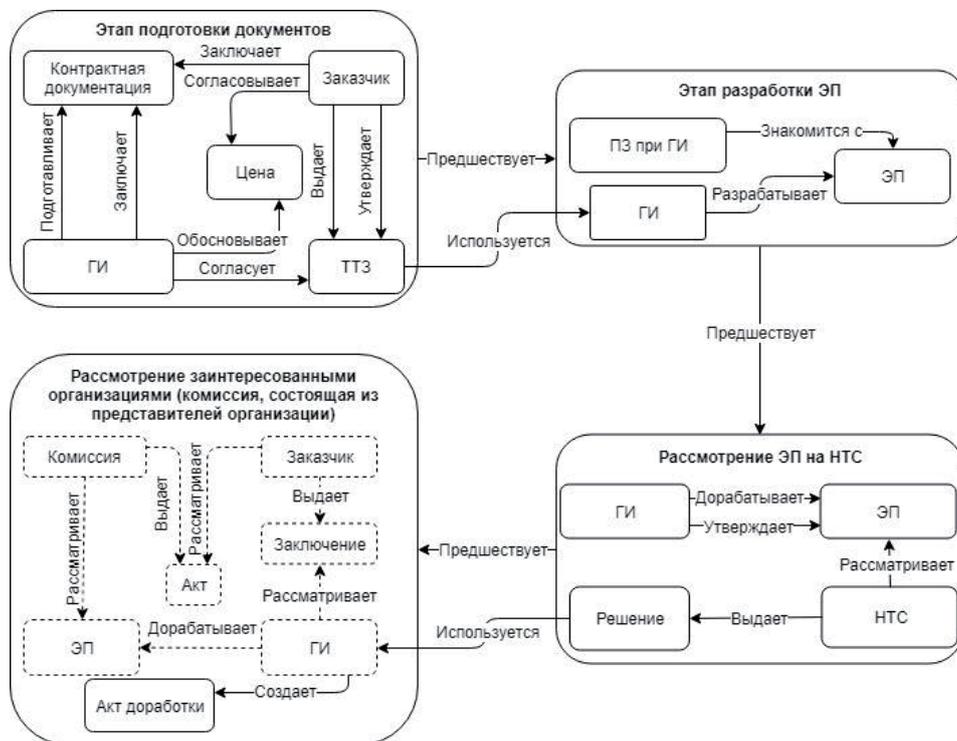


Рисунок 7 — Фрагмент метаграфа, построенного по тексту стандарта, описывающего структурную и процессную составляющие работ, выполняемых в рамках ОКР.

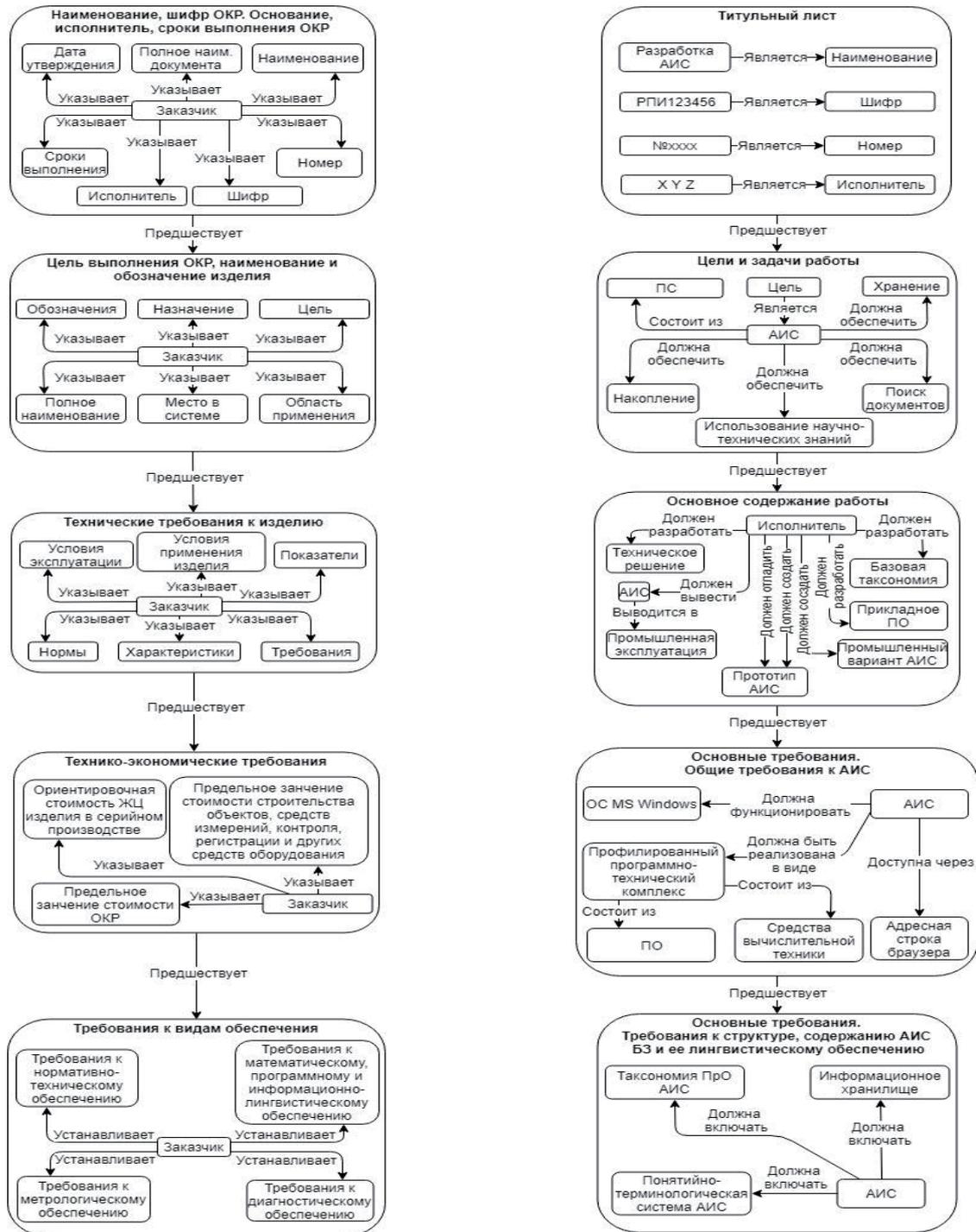


Рисунок 8 — Соотнесение подграфов «Техническое задание»: построенного на основе текста стандарта – слева, и построенного на основе текста реального технического задания – справа (здесь ПО – Программное обеспечение; ПС – Программные средства)

На рисунке 6 представлен фрагмент метаграфа, построенного на основе стандарта, метавершины которого определяют отдельные стадии НИР, а порядок переходов между стадиями определяется отношениями «Предшествует». Стадии характеризуются определённой для них деятельностью, в которой задействованы присущие этой стадии объекты. Объекты взаимодействуют согласно определённым для этой стадии операциям и нормативным документам. Каждая организационная стадия имеет характерный для неё набор документов.

Изображённый на рисунке 6 фрагмент метаграфа для описания НИР имеет четыре метавершины (*Определить предмет исследования, Провести анализ научной литературы, Провести анализ патентной литературы, Выбрать направление и методику исследования*), хотя в тексте стандарта эти виды деятельности объединены в одну стадию (*Выбор направления и методики исследования*). Подобное разделение на более мелкие стадии позволяет лучше представить порядок появления документации, определять причинно-следственные связи и историю изменений документов и объектов.

Состав работ и объектов при выполнении конкретного проекта всегда уточняется и дополняется, что фиксируется в ТЗ и другой документации проекта. В этом случае операции над онтологиями обеспечивают возможность выявления соответствий/расхождений между сравниваемыми документами формальными методами.

На рисунке 7 представлен фрагмент графа, построенного на основе текста стандарта, описывающего порядок выполнения ОКР по созданию изделий и их составных частей. Подграф метавершин включает лексику определённого этапа и может быть использован для поиска как нормативных документов, так и документов реальных НИР и ОКР и т.д.

Вершина «Тактико-техническое задание» («ТТЗ») в метавершине «Этап подготовки документов» (см. рисунок 7) может раскрываться подграфом, построенным по тексту стандарта, описывающего требования к содержанию и оформлению ТЗ (рисунок 8, слева), или же по тексту ТЗ реального проекта (рисунок 8, справа) представлен фрагмент описания реального ТЗ на разработку автоматизированной информационной системы (АИС).

Переход от вершины «ТТЗ» (рисунок 7) к графу (рисунок 8, слева), представляющему требования к ТТЗ, можно представить следующими действиями: при выборе вершины «ТТЗ» формируется соответствующий запрос, в ответ на который получаются связанные документы (нормативная и проектная документация) и построенные на их основе графы онтологий.

Чтобы выделить только те фрагменты, которые интересуют пользователя, можно воспользоваться операцией аспектной проекции (аспект задаётся классами отношений, подробнее в [1]) для сущности «ТТЗ». Выбирая вершину «ТТЗ» как исходную точку, отбираются вершины, связанные с ней отношениями типа «Являться частью/целым», а полученные вершины сами становятся новыми запросами.

Выбрав помимо интересующей вершины её окружение (т.е. контекст), становится возможным провести поиск документов по контексту. Пример выделения комбинации вершин и отношений для вершины «Главной исполнитель» («ГИ») представлен на рисунке 7 (вершины и отношения выделены пунктиром). Одним из результатов выдачи по выбранным для поиска вершинам и отношениям будет стандарт, потому что в нём содержатся конкретные предложения, которые описываются выделенными вершинами и отношениями.

Онтологический подход к представлению содержания документов позволяет также проводить и сравнение содержания документов, используя операции объединения, пересечения графов. Так, сравнение графов, изображённых на рисунке 8, показывает, что метавершина «Цель выполнения ОКР...» (рисунок 8, слева) является структурной основой для метавершины «Цели и задачи работы» (рисунок 8, справа).

Конструкция, состоящая из вершин «Цель» и «Заказчик», связанных отношением «Указывает», соотносится с вершинами «Цель» и «АИС», связанных отношением «Является» из метавершины «Цели и задачи работы», т.к. имеются понятийно тождественные вершины «Цель» и связи, принадлежащие одному классу. Вершина «Назначение» косвенно отождествляется со следующей группой троек из метавершины «Цели и задачи работы»: «АИС – должна обеспечить – Хранение», «АИС – должна обеспечить – Поиск документов», «АИС – должна обеспечить – Накопление». Вершина «Требования» из метавершины «Технические требования к изделию» (рисунок 8, слева) в нормативном документе раскрывается через ме-

тавершину «Основные требования. Технические требования к АИС» (рисунок 8, справа) графа онтологии реального ТЗ. Кроме того, на рисунке 8 справа отсутствуют регламентированные в стандарте такие составляющие, как «Характеристики», «Нормы», «Место в системе» и др., а также отсутствует часть «Технико-экономические требования».

Использование процессных онтологий позволяет узнать не только о существующих работах, этапах, порядке их выполнения, субъектах деятельности, документах, соответствующих конкретным работам, но и место определённого документа в ЖЦ проекта, а методы операций над графами позволяют проводить оценку соответствия (валидацию и верификацию) проектной документации формальным требованиям.

## 6 О координатах онтологического пространства знаний

Чтобы система управления знаниями была адекватна и практически эффективна, она должна интегрировать информационные потоки и ресурсы, относящиеся ко всем видам и этапам деятельности, начиная с НИР и заканчивая выводом из эксплуатации, от стратегического планирования до реализации и оценки результатов, от формирования образовательных программ до подготовки и переподготовки кадров. При этом рациональная деятельность состоит в установлении взаимного соответствия объектов деятельности и предметов деятельности. В результате исследования объекта создаётся его предметно-ориентированная модель, при реализации модель воплощается в экземпляре объекта.

Для полноценного сохранения и эффективного управления знаниями, основанными на адекватной идентификации как экземпляров хранения, так и структур смысловых фрагментов знания во времени и в изменяющихся условиях, необходимо иметь возможность представлять знание во всех онтологических пространствах и координатах.

Можно ввести следующие основные «пространства» жизни объекта:

- сфера основной деятельности субъекта, где по форме существования выделяются два подпространства – абстрактных объектов (концептуальные модели, теории, составляющие *предмет деятельности*) и конкретных объектов/процессов (представляющих *объект деятельности*);
- время модельное (шкала, соответствующая ЖЦ), как фактор, обуславливающий существование разных сред и представлений знаний в зависимости от вида деятельности и специализации процессов;
- сфера информационной деятельности субъекта, предопределяющая способы и формы представления объекта в виде информационных сообщений;
- «пространство» языка, как средства представления и передачи знаний в процессах сохранения, поиска, синтеза знаний;
- время астрономическое, как фактор, предполагающий изменение объектов ПрО, а также знания и условий его синтеза и использования.

Концептуальный образ (онтология экземпляра ПрО) абстрактного или конкретного объекта представлен в следующих «координатах»:

- координата «объект», задаваемая структурной таксономией (точнее, мерономией), представляющей соотношение составных частей объекта;
- координата «предмет», задаваемая функциональной таксономией (как структура модели ПрО) – структурой, представляющей теоретические и иные знания, относящиеся к этапам ЖЦ изделия;
- координата «деятельность» (как ось модельного времени), представляющая распределение работ/процессов по этапам ЖЦ и связанное с ним распределение по формам представления, задаваемая таксономией стадий и этапов ЖЦ и/или синхронизированной с

ней таксономией типов и видов документов как форм и способов описания, специфических для конкретного этапа/процесса.

В этом случае таксономии следует рассматривать как иерархические *номинальные шкалы*, градациям которых соответствуют относительно цельные семантические структуры.

Это *3D* пространство «вложено» в *2D* пространство с координатами «язык-время» (*3+2D* пространство в некотором смысле аналогично *4D* пространству, определённое в *ISO 15926* [49]), связывающее состояние «диалекта» языка, характерного для конкретной ПрО, с моментом времени формирования онтологии (рисунок 9).

В этом примере конкретное целевое знание о свойствах надёжности главного циркуляционного насоса (рисунок 9, точка с координатами {«ГЦНА», «Надёжность», «Описание»}) представлено семантической сетью графа функциональной системы онтологии, являющейся объединением графов (с распределением подграфов по «модельному времени» - соответствующим этапам ЖЦ), построенных по текстам документов разного вида. Динамика процесса (организация деятельности по созданию нового знания) отражается физическим (астрономическим) временем, если оно присутствует в виде события или свойства объекта/отношения (отношениями локативности), а также порядком следования фактов в тексте.

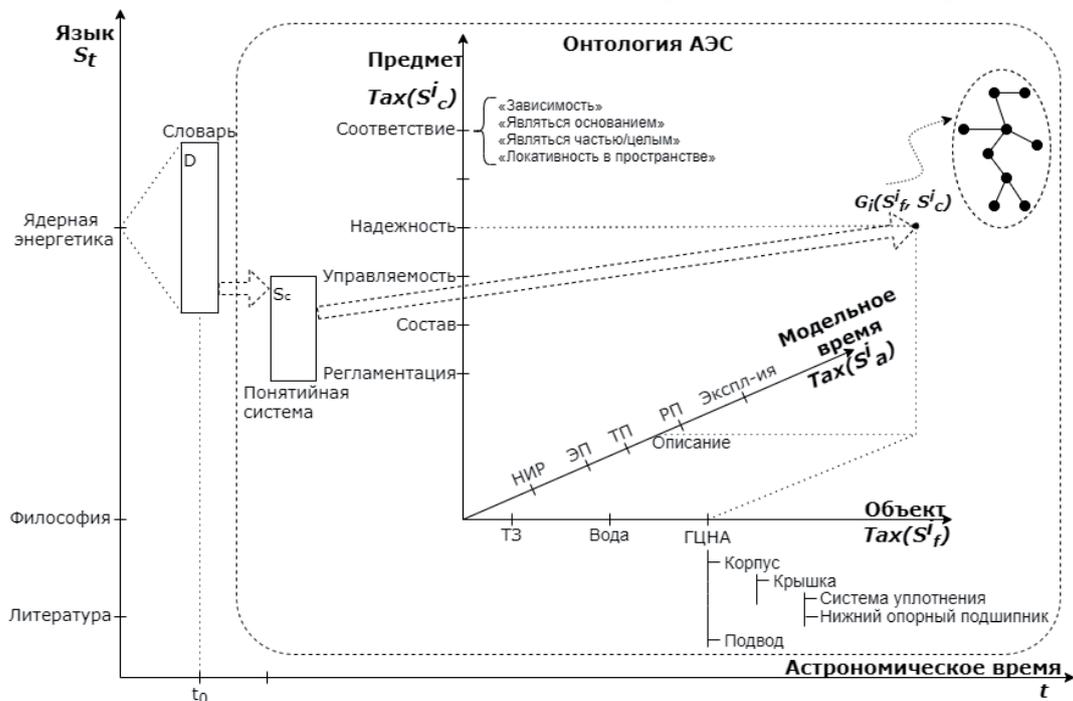


Рисунок 9 — Пространство онтологии знаний

Система онтологий, представляющих знание, является *организацией* (по А.А. Богданову [50]) – это структура частей (представляющих знание о действительности и формах представления), имеющих материальную форму и взаимодействующих во времени и пространстве. Здесь «структура частей» означает, что в процессе развития в рамках определённых условий закономерно выделяются наиболее типичные формы организации. Это означает, что имеются общие механизмы отбора структур из их возможных вариантов. Онтология – это не только статичный концептуальный образ, соответствующий некоторому состоянию ПрО, но и динамическая структура, диалектически взаимодействующая с ПрО и предопределяющая пути её развития.

Система онтологий по отношению к деятельности является семантическим ядром, фиксирующим «смыслы» ПрО – *систему* понятий и отношений, на которую «проецируются» находимые в процессе деятельности решения.

## Заключение

Понятие «онтология» многогранно. Онтология отличается от описания не только тем, что представлена на уровне концептов, но и тем, что включает в себя онтологию языка, форм представления и онтологию соответствующей ПрО. Предложенная модель системы онтологий знаний:

- включает идентифицирующую и объясняющую функции;
- учитывает фактор времени;
- позволяет формировать семантическую окрестность понятия, находить контекстно определяемый путь между семантическими компонентами.

Исходя из принципов схематизма познания, разработана система онтологий представления знаний, объединяющая язык, формы представления знаний и схемы процессов.

Определены роль и место онтологии артефакта и онтологии процессов в системе онтологий представления знаний. Разработана онтология артефакта – модель объекта, инструмента и результата рациональной деятельности, включающая набор свойств и методов их оценки и «сопутствующий» контекст. Предложенная онтология процессов, включающая иерархически связанную совокупность этапов ЖЦ деятельности, объекты и субъекты деятельности каждого этапа, исполняет роль «навигационной карты» для субъектов деятельности, позволяя им ориентироваться в сложных организационных структурах.

## Благодарности

Работа выполнена при поддержке Министерства науки и высшего образования РФ (проект государственного задания № 0723-2020-0036).

## Список источников

- [1] **Максимов, Н.В.** Средства семантического поиска, основанные на онтологических представлениях документальной информации / Н.В. Максимов, О.Л. Голицына, К.В. Монанков, А.А. Лебедев, Н.А. Баль, С.Г. Кюрчева // Научно-техническая информация. Сер. 2. – 2019. – №7. – С.8-19.
- [2] **Maksimov, N.** Toward a Building an Ontology of Artefact / N. Maksimov, A. Lebedev // In: Samsonovich A.V., Gudwin R.R., Simões A.S. (eds) Brain-Inspired Cognitive Architectures for Artificial Intelligence: BICA\*AI 2020. BICA 2020. Advances in Intelligent Systems and Computing. - Springer, Cham, 2021. vol.1310. P.225-232. [https://doi.org/10.1007/978-3-030-65596-9\\_29](https://doi.org/10.1007/978-3-030-65596-9_29).
- [3] **Павлов, П.Ю.** Онтологическая поддержка технологической подготовки производства трубопроводов летательного аппарата / П.Ю. Павлов, П.И. Соснин // Известия Самарского научного центра Российской академии наук. – 2017. – Т. 19, №1. – С.187-194.
- [4] **Скобелев, П.О.** Онтологии деятельности для ситуационного управления предприятиями в реальном времени / П.О. Скобелев // Онтология проектирования. – 2012. – №1 (3). – С.6-38.
- [5] **Белов, М.В.** Структура методологии комплексной деятельности / М.В. Белов, Д.А. Новиков // Онтология проектирования. – 2017. – Т.7, №4(26). – С.366-387. DOI: 10.18287/2223-9537-2017-7-4-366-387.
- [6] **Максимов, Н.В.** Методологические основы онтологического моделирования документальной информации / Н.В. Максимов // Научно-техническая информация. Сер. 2. – 2018. – № 3. – С.6-22.
- [7] **Гуревич, И.М.** Законы информатики – основа строения и познания сложных систем / И.М. Гуревич. ☐ М.: ТОРУС ПРЕСС, 2007. 400 с.
- [8] **Огнев, А.И.** Учение Канта о схематизме понятий чистого рассудка / А.И. Огнев // Кантовский сборник. ☐ 2012. ☐ №2(40). – С.79-86.
- [9] **Пиаже, Ж.** Теория, эксперименты, дискуссии / Ж. Пиаже. ☐ М.: Гардарики, 2001. 624 с.
- [10] **Сазонов, Б.В.** Техника схематизации и схемы в ММК: к проблеме онтологического статуса схем / Б.В. Сазонов. // Третья конференция по схематизации на тему «Разворачивание и жизнь методологических схем» 2009. Тексты к конференции. <https://www.fondgp.ru/old/lib/conferences/2009/texts.html>.
- [11] **Морозов, Ф.М.** Схемы как средство описания деятельности (эпистемологический анализ) / Ф.М. Морозов. – М.: ИФ РАН, 2005. 182 с.

- [12] **Розин, М.В.** Введение в схемологию. Схемы в философии, культуре, науке, проектировании / М.В. Розин. М.: Книжный дом «ЛИБРОКОМ», 2011. 256 с.
- [13] **Лотман, Ю.М.** Культура и взрыв / Ю.М. Лотман. М.: Гнозис; Издат. группа «Прогресс», 1992. 272 с.
- [14] **Налимов, В.В.** Спонтанность сознания. Вероятностная теория смыслов и смысловая архитектура личности. / В.В. Налимов. – М: Прометей, 1989. 288 с.
- [15] **Тарасов, В.Б.** Онтологии жизненного цикла сложной технической системы / В.Б. Тарасов, А.В. Федотова, Н.В. Черепанов // Открытые семантические технологии проектирования интеллектуальных систем. Матер. IV междунар. науч.-техн. конф. (OSTIS-2014, Минск, БГУИР, 20–22 февраля 2014 г.). – Минск, БГУИР, 2014. – С.471-482.
- [16] **Гришин, М.В.** Онтологическая поддержка конструкторской деятельности в условиях технологической подготовки производства на основе концептуальных экспериментов / М.В. Гришин, А.В. Лебедев, П.И. Соснин // Известия Самарского научного центра Российской академии наук. – 2016. – Т.18, №4(3). – С.451-458.
- [17] **Ермолаев, К.А.** Использование онтологии для управления знаниями предприятия / К. А. Ермолаев // Исследования по информатике. – 2007. – вып. 12. – С.65–78.
- [18] **Мельник, Е.Н.** Онтологические модели для систем управления электроснабжением олимпийских объектов в Сочи / Е.Н. Мельник, А.Ю. Бадалов, Б.Я. Шведин, Д.Б. Гвоздев, Л.В. Бузаев // Онтология проектирования. – 2014. – №1 (11). – С.6-23.
- [19] ГОСТ Р 57193-2016. Процессы жизненного цикла систем / Системная и программная инженерия. – Москва: Стандартинформ, 2016.
- [20] ГОСТ Р 57102-2016/ISO/IEC TR 24748-2:2011. Управление жизненным циклом. Часть 2. Руководство по применению ИСО/МЭК 15288 / Информационные технологии (ИТ). Системная и программная инженерия. – Москва: Стандартинформ, 2016.
- [21] ГОСТ Р ИСО/МЭК 12207-2010. Процессы жизненного цикла программных средств / Информационная технология. Системная и программная инженерия. – Москва: Стандартинформ, 2011.
- [22] **Голенков, В.В.** Проектирование предприятий рецептурного производства на основе онтологий / В.В. Голенков, В.В. Таберко, Д.С. Иванюк // Онтология проектирования. – 2017. – Т.7, №2(24). – С.123-144. DOI: 10.18287/2223-9537-2017-7-2-123-144.
- [23] **Gruninger, M.** Ontology of the Process Specification Language / Michael Gruuninger. 20 p. - [https://tsapps.nist.gov/publication/get\\_pdf.cfm?pub\\_id=822123](https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=822123).
- [24] **Pedrinaci, C.** A Core Ontology for Business Process Analysis / C. Pedrinaci, J. Domingue, A.K. Alves de Medeiros // In: Bechhofer S., Hauswirth M., Hoffmann J., Koubarakis M. (eds). The Semantic Web: Research and Applications. ESWC 2008. Lecture Notes in Computer Science. – Springer, Berlin, Heidelberg, 2008. Vol.5021. – P.49–64. [https://doi.org/10.1007/978-3-540-68234-9\\_7](https://doi.org/10.1007/978-3-540-68234-9_7).
- [25] **De Roure D.** The Semantic Grid: a Future e-Science Infrastructure / D. De Roure, N. R. Jennings, N. R. Shadbolt // In: Berman F, Fox G, Hey AJG (eds). Grid Computing: Making the Global Infrastructure a Reality. – Wiley, New York, 2003. – P.437-470.
- [26] **Загоруйко, Н. Г.** Система ontogrid для автоматизации процессов / Н. Г. Загоруйко, В. Д. Гусев, А. В. Завертайлов, С. П. Ковалев, А. М. Налетов, Н. В. Саломатина // Автометрия. – 2005. – Т.41, №5. – С.13-25.
- [27] **Schaffert, S.** Semantic Wiki for Collaborative Knowledge Formation / S. Schaffert, A. Gruber, R. Westenthaler // In: Reich, S., et al. Semantic Content Engineering. Proc. der Semantics 2005. – Trauner, Linz, 2005. – P.188–202.
- [28] Theory and History of Ontology. <https://www.ontology.co>.
- [29] **Витяев, Е.Е.** Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов / Е.Е. Витяев. – Новосибирск: Новосибир. гос. ун-т, 2006. 293 с.
- [30] **Guarino, N.** Formal ontology, conceptual analysis and knowledge representation / N. Guarino // International Journal of Human-Computer Studies. – 1995. – Vol.43, Is.5-6. – P.625-640.
- [31] **Gangemi, A.** ONIONS: An Ontological Methodology for Taxonomic Knowledge Integration / A. Gangemi, G. Steve, F. Giacomelli// In: Workshop on Ontological Engineering. ECAI-96. Budapest, Hungary. – Wiley, Chichester/London/New York, 1996. – P.29-40.
- [32] **Голицына, О.Л.** Онтологический подход к идентификации информации в задачах документального поиска / О.Л. Голицына, Н.В. Максимов, О.В. Окропишина, В.И. Строгонов // Научно-техническая информация. Сер.2. – 2012, № 5. – С.1-9.
- [33] **Урманцев, Ю.А.** Общая теория систем в доступном изложении / Ю.А. Урманцев. Ижевск: НИЦ «Регулярная и хаотическая динамика», 2014. 408 с.
- [34] **Голицына, О.Л.** Онтологический подход к идентификации информации в задачах документального поиска: практическое применение / О.Л. Голицына, Н.В. Максимов, О.В. Окропишина, В.И. Строгонов // Научно-техническая информация. Сер. 2. – 2013, №3. – С.1-8.

- [35] **Максимов, Н. В.** Документальная информационно-аналитическая система xIRBIS (редакция 6.0): программа для ЭВМ. / Н.В. Максимов, О.Л. Голицына, К.В. Монанков, А.С. Гаврилкина // Свидетельство о гос. регистрации №2020661683 от 29.09.2020.
- [36] **Maksimov, N.** Ontology of Properties and its Methods of Use: Properties and Unit extraction from texts. / N. Maksimov, A. Gavrilkina, V. Kuzmina, E. Borodina // Procedia Computer Science. – 2020. –169. – P.70–75.
- [37] QUDT- Quantities, Units, Dimensions and Types. <http://www.qudt.org>
- [38] **Rijgersberg, H.** Ontology of Units of Measure and Related Concepts. / H. Rijgersberg, M. van Assem, J. Top // Semantic Web. – 2013. - 4(1). – P.3-13.
- [39] **Gkoutos, G.V.** The Units Ontology: a tool for integrating units of measurement in science / G.V. Gkoutos, P. N. Schofield, R. Hoehndorf // Database. – 2012. – bas033.
- [40] **Celebi, I.** Improving Interoperability by Incorporating UnitsML Into Markup Languages / I. Celebi, R.A. Dragoset, K.J. Olsen, R. Schaefer, G.W. Kramer // Journal of Research of the National Institute of Standards and Technology. – 2010. – P.15-22.
- [41] **Ильичев, Л.Ф.** Философский энциклопедический словарь / Л.Ф. Ильичев, П.Н. Федосеев, С.М. Ковалев, В.Г. Панов. – М.: Советская энциклопедия, 1983. 840 с.
- [42] International vocabulary of metrology – Basic and general concepts and associated terms (VIM). – [https://www.bipm.org/utis/common/documents/jcgm/JCGM\\_200\\_2008.pdf](https://www.bipm.org/utis/common/documents/jcgm/JCGM_200_2008.pdf)
- [43] **Гегель, Г.В.Ф.** Наука логики. В 3-х т. Т.1 Мысль / Г.В.Ф. Гегель. – М.: Мысль, 1970. 501 с.
- [44] РМГ 29-2013. Метрология. Основные термины и определения / Государственная система обеспечения единства измерений. – Москва: Стандартинформ, 2014.
- [45] ГОСТ 15467-79. Основные понятия. Термины и определения / Управление качеством продукции. – Москва: Стандартинформ, 2009.
- [46] ГОСТ 33701-2015. Определение и применение показателей точности методов испытаний нефтепродуктов. – Москва: Стандартинформ, 2016.
- [47] ГОСТ Р ИСО 9000-2015. Основные положения и словарь / Системы менеджмента качества. – Москва: Стандартинформ, 2015.
- [48] ГОСТ 16504-81. Испытания и контроль качества продукции. Основные термины и определения / Система государственных испытаний продукции. – Москва: Стандартинформ, 2011.
- [49] ГОСТ Р ИСО 15926-2-2010. Интеграция данных жизненного цикла для перерабатывающих предприятий, включая нефтяные и газовые производственные предприятия. Часть 2. Модель данных / Системы промышленной автоматизации и интеграция. – Москва: Стандартинформ, 2013.
- [50] **Богданов, А.А.** Тектология: Всеобщая организационная наука / А.А. Богданов. – Москва: Финансы, 2003. 496 с.

## Сведения об авторах



**Максимов Николай Вениаминович**, 1952 г. рождения. В 1975 г. окончил Московский инженерно-физический институт по специальности «Прикладная математика», д.т.н. (2002 г.). Профессор кафедры финансового мониторинга НИЯУ МИФИ. Область научных интересов: моделирование и разработка документальных информационно-поисковых систем и баз данных, лингвистическое обеспечение документальных информационно-поисковых систем и систем управления знаниями; человеко-машинные информационные системы, интерфейсы на основе когнитивных и поведенческих моделей. Автор более 120 научных работ. ORCID: 0000-0002-8191-1521. [nv-maks@yandex.ru](mailto:nv-maks@yandex.ru)

**Лебедев Александр Анатольевич**, 1986 г. рождения. В 2009 г. окончил НИЯУ МИФИ по специальности «Физика конденсированного состояния вещества и фотоника». В 2015 г. окончил магистратуру в НИЯУ МИФИ по специальности «Прикладная математика и информатика». Инженер кафедры финансового мониторинга НИЯУ МИФИ. Область научных интересов: моделирование информационных взаимодействий в документальных базах данных. ResearcherID (WoS): M-3125-2015, ORCID: 0000-0002-3780-8092. [lebedevalex@live.ru](mailto:lebedevalex@live.ru)



Поступила в редакцию 15.03.2021, после рецензирования 20.04.2021. Принята к публикации 25.05.2021.

## Ontological system "knowledge-activity"

N.V. Maksimov, A.A. Lebedev

National Research Nuclear University MEPhI

### Abstract

The paper considers an approach to construction of human rational activity ontological description. Such a description can serve as an additional source for entering the information space, expands the class of possible indexes for describing documents and allows to implement contextual search. The purpose and features of the use of ontologies in describing the subject area, both from a theoretical and an applied point of view, are considered. Criteria for identifying types of ontologies are indicated. Based on the principles of knowledge schematism, a knowledge representation ontologies system has been developed, which unites language, forms of knowledge representation and process schemes. It is shown that precisely such kind of ontologies system makes possible the practical use of ontologies in computing environments. For a comprehensive description of production (rational) activity, ontologies of artifacts and processes are presented, developed on the basis of an analysis of state and international standards. The ontology of an artifact is a description of an object - a means and a result of purposeful activity, an ontology of processes is a formalized description of the life cycle of an activity. With the help of the ontology of the artifact, it becomes possible to represent the logical structure of activity, and with the help of the ontology of processes, the temporal structure, which jointly determine the two sides of the description of complex activity. The proposed ontology of processes includes a hierarchically related set of stages of the life cycle of activity, objects and subjects of activity of each stage and plays the role of a "navigation map" for subjects of activity, allowing them to navigate in complex organizational structures.

**Key words:** *ontology, knowledge, activity, ontology of artefact, ontology of processes.*

**Citation:** *Maksimov NV, Lebedev AA. Ontological system "knowledge-activity" [In Russian]. Ontology of designing. 2021; 11(2): 185-211. DOI: 10.18287/2223-9537-2021-11-2-185-211.*

**Acknowledgment:** This work was supported by the Ministry of science and higher education of Russian Federation (project of state assignment № 0723-2020-0036).

### List of figures

- Figure 1 - The relationship of various aspects of activities in terms of the nature of the work, purpose and time of completion
- Figure 2 - Ontologies systematization according to the level of knowledge they represent (scope), scope of applicability (acceptance), means of presentation (expressivity)
- Figure 3 - Types of ontologies by type of activity
- Figure 4 - Ontologies system of knowledge representation
- Figure 5 - The ontology of artefact
- Figure 6 - Fragment of metagraph, built on basis of text of standard that regulates the procedure for performing research work
- Figure 7 - Fragment of metagraph, built according to text of standard, describing structural and process components carried out within the framework of Development work
- Figure 8 - Interrelation of subgraphs "Statement of work": built on the basis of standard text - on the left, and built on the basis of text of the real statement of work - on the right
- Figure 9 - Knowledge ontology space

### References

- [1] *Maksimov NV, Golitsyna OL, Monankov KV, Lebedev AA, Bal NA, Kyurcheva SG.* Semantic search tools based on ontological representations of documentary information [In Russian]. *Automatic Documentation and Mathematical Linguistics*, 2019; 2(7): 8-19.
- [2] *Maksimov N, Lebedev A.* Toward a Building an Ontology of Artefact. *Brain-Inspired Cognitive Architectures for Artificial Intelligence: BICA \* AI 2020. BICA 2020. Advances in Intelligent Systems and Computing.* Cham: Springer; 2021: 225-232.
- [3] *Pavlov PYu, Sosnin PI.* Ontological support of technological preparation of aircraft pipeline production [In Russian]. *Bulletin of the Samara Scientific Center of the Russian Academy of Sciences*, 2017; 19(1): 187-194.

- [4] **Skobelev PO.** Activity ontology for situational management of enterprises in real time [In Russian]. *Ontology of designing*, 2012; 1(3): 6-38.
- [5] **Belov MV, Novikov DA.** Structure of methodology of complex activity [In Russian]. *Ontology of designing*, 2017; 7; 4(26): 366-387.
- [6] **Maksimov NV.** Methodological foundations of ontological modeling of documentary information [In Russian]. *Automatic Documentation and Mathematical Linguistics*, 2018; 3: 6-22.
- [7] **Gurevich IM.** The laws of informatics are the basis for complex systems structure and knowledge [In Russian]. Moscow: TORUS PRESS; 2007. 400 p.
- [8] **Ognev AI.** Kant's doctrine of the pure reason concepts schematism [In Russian]. *Kantian collection*, 2012; 2(40): 79-86.
- [9] **Piaget J.** Theory, experiments, discussions [In Russian]. Moscow: Gardariki; 2001. 624 p.
- [10] **Sazonov BV.** Technique of schematization and schemes in MMC: to the problem of schemes ontological status [In Russian]. The third conference on schematization on the topic "Development and life of methodological schemes" '2009. Texts for the conference. - <https://www.fondgp.ru/old/lib/conferences/2009/texts.html>
- [11] **Morozov FM.** Schemes as a means of describing activities (epistemological analysis) [In Russian]. Moscow: Institute of Philosophy of RAS; 2005. 182 p.
- [12] **Rozin MV.** Introduction to schemology. Schemes in philosophy, culture, science, design [In Russian]. Moscow: Book House "LIBROKOM"; 2011. 256 p.
- [13] **Lotman YM.** Culture and explosion [In Russian]. Moscow: Gnosis, Publis. group "Progress"; 1992. 272 p.
- [14] **Nalimov VV.** Spontaneity of consciousness. Probabilistic theory of senses and semantic architectonics of personality [In Russian]. Moscow: Prometheus; 1989. 288 p.
- [15] **Tarasov VB, Fedotova AV, Cherepanov NV.** Ontologies of the life cycle of a complex technical system [In Russian]. Proc. of IV international scientific and technical conf. (OSTIS-2014, Minsk, BSUIR, February 20-22, 2014). Minsk: BSUIR; 2014: 471-482.
- [16] **Grishin MV, Lebedev AV, Sosnin PI.** Ontological support of design activity in conditions of technological preparation of production based on conceptual experiments [In Russian]. *Bulletin of Samara Scientific Center of RAS*, 2016; 18; 4(3): 451-458.
- [17] **Ermolaev KA.** The use of ontology for enterprise knowledge management [In Russian]. *Research in informatics*, 2007; 12: 65-78.
- [18] **Melnik EN, Badalov AY, Shvedin BY, Gvozdev DB, Buzaev LV.** Ontological model for control systems of power supply of Olympic facilities in Sochi [In Russian]. *Ontology of designing*, 2014; 1(11): 6-23.
- [19] GOST R 57193-2016. System and software engineering. Systems life cycle processes [In Russian]. Moscow: Standardinform, 2016.
- [20] GOST R 57102-2016 / ISO / IEC TR 24748-2: 2011. Information technology (IT). System and software engineering. Life cycle management. Part 2. Guidance on the application of ISO / IEC 15288 [In Russian]. Moscow: Standardinform, 2016.
- [21] GOST R ISO / IEC 12207-2010. Information technology. System and software engineering. Software life cycle processes [In Russian]. Moscow: Standardinform, 2011.
- [22] **Golenkov VV, Taberko VV, Ivanyuk DS.** Ontology-based design of prescription production enterprises [In Russian]. *Ontology of designing*, 2017; 7; 2(24): 123-144. DOI: 10.18287/2223-9537-2017-7-2-123-144.
- [23] **Gruninger M.** Ontology of the Process Specification Language. [https://tsapps.nist.gov/publication/get\\_pdf.cfm?pub\\_id=822123](https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=822123).
- [24] **Pedrinaci C, Domingue J, Alves de Medeiros AK.** A Core Ontology for Business Process Analysis. In: Bechhofer S., Hauswirth M., Hoffmann J., Koubarakis M. (eds). *The Semantic Web: Research and Applications. ESWC 2008. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer; 2008: 49-64.
- [25] **De Roure D, Jennings NR, Shadbolt NR.** The Semantic Grid: a Future e-Science infrastructure. In: Berman F, Fox G, Hey AJG (eds). *Grid Computing: Making the Global Infrastructure a Reality*. New York: Wiley; 2003: 437-470.
- [26] **Zagoruiko NG, Gusev VD, Zavertaylov AV, Kovalev SP, Naletov AM, N. Salomatina NV.** The ontogrid system for process automation [In Russian]. *Autometry*, 2005; 41; 5: 13-25.
- [27] **Schaffert S.** Semantic Wiki for Collaborative Knowledge Formation. In: Reich, S., et al. *Semantic Content Engineering. Proc. der Semantics 2005*. Linz: Trauner, Linz; 2005: 188-202.
- [28] *Theory and History of Ontology*. - <https://www.ontology.co>
- [29] **Vityaev EE.** Extracting knowledge from data. *Computer cognition. Models of cognitive processes* [In Russian]. Novosibirsk: Novosibirsk State university; 2006. 293 p.
- [30] **Guarino N.** Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*, 1995; 43; 5-6: 625-640.
- [31] **Gangemi A, Steve G, Giacomelli F.** ONIONS: An Ontological Methodology for Taxonomic Knowledge Integration. In: *Workshop on Ontological Engineering. ECAI-96*. Budapest, Hungary. London/NY: Wiley; 1996: 29-40.

- [32] **Golitsyna OL, Maksimov NV, Okropishina OV, Strogonov VI.** The ontological approach to the identification of information in tasks of document retrieval [In Russian]. *Automatic Documentation and Mathematical Linguistics*, 2012; 5: 1-9.
- [33] **Urmantsev YuA.** General theory of systems in an accessible presentation [In Russian]. Izhevsk: Research Center "Regular and Chaotic Dynamics"; 2014. 408 p.
- [34] **Golitsyna OL, Maksimov NV, Okropishina OV, Strogonov VI.** The ontological approach to the identification of information in tasks of documentary search: practical application retrieval [In Russian]. *Automatic Documentation and Mathematical Linguistics*, 2013; 3: 1-8.
- [35] **Maksimov NV, Golitsyna OL, Monankov KV, Gavrilkina AS.** Documentary information and analytical system xIRBIS (rev.6.0): computer program [In Russian]. Certificate of state. registr. No. 2020661683 dated 09/29/2020.
- [36] **Maksimov N, Gavrilkina A, Kuzmina V, Borodina E.** Ontology of Properties and its Methods of Use: Properties and Unit extraction from texts. *Procedia Computer Science*, 2020; 169: 70-75.
- [37] QUDT- Quantities, Units, Dimensions and Types. - <http://www.qudt.org>
- [38] **Rijgersberg H, van Assem M, Top J.** Ontology of Units of Measure and Related Concepts. *Semantic Web*, 2013; 4(1): 3-13.
- [39] **Gkoutos GV, Schofield PN, Hoehndorf R.** The Units Ontology: a tool for integrating units of measurement in science. *Database*, 2012; bas033.
- [40] **Celebi I, Dragoset RA, Olsen KJ, Schaefer R, Kramer GW.** Improving Interoperability by Incorporating UnitsML Into Markup Languages. *Journal of Research of the National Institute of Standards and Technology*, 2010; 15-22.
- [41] **Ilyichev LF, Fedoseev PN, Kovalev SM, Panov VG.** *Philosophical Encyclopedic Dictionary* [In Russian]. Moscow: Soviet Encyclopedia; 1983. 840 p.
- [42] International vocabulary of metrology - Basic and general concepts and associated terms (VIM). - [https://www.bipm.org/utis/common/documents/jcgm/JCGM\\_200\\_2008.pdf](https://www.bipm.org/utis/common/documents/jcgm/JCGM_200_2008.pdf)
- [43] **Hegel GWF.** *Science of logic*. In 3 volumes. Vol. 1 Thought [In Russian]. Moscow: Idea; 1970. 501 p.
- [44] RMG 29-2013. State system for ensuring the uniformity of measurements. *Metrology. Basic terms and definitions* [In Russian]. Moscow: Standardinform, 2014.
- [45] GOST 15467-79. Product quality management. Basic concepts. Terms and definitions [In Russian]. Moscow: Standardinform, 2009.
- [46] GOST 33701-2015. Determination and application of accuracy indicators for testing methods for petroleum products [In Russian]. Moscow: Standardinform, 2016.
- [47] GOST R ISO 9000-2015. Quality management systems. Fundamentals and vocabulary [In Russian]. Moscow: Standardinform, 2015.
- [48] GOST 16504-81. System of state product testing. Product testing and quality control. Basic terms and definitions [In Russian]. Moscow: Standardinform, 2011.
- [49] GOST R ISO 15926-2-2010. Industrial Automation Systems and Integration. Integration of lifecycle data for refineries, including oil and gas production plants. Part 2. Data Model [In Russian]. Moscow: Standardinform, 2013.
- [50] **Bogdanov AA.** *Tectology: General organizational science* [In Russian]. Moscow: Finance; 2003. 496 p.

## About the authors

**Nikolai Veniaminovich Maksimov** (b. 1952) graduated from the Moscow Engineering Physics Institute in 1975, D. Sc. Eng. (2002). Professor of Financial monitoring Department of National Research Nuclear University MEPhI. Research interests: modeling and development of documentary information search systems and databases; linguistic support of documentary information search systems and knowledge management systems; human-machine information systems, interfaces based on cognitive and behavioral models. Author of more than 120 scientific papers, textbooks and tutorials. ORCID: 0000-0002-8191-1521. [nv-maks@yandex.ru](mailto:nv-maks@yandex.ru)

**Lebedev Alexander Anatolyevich** (b. 1986) graduated from the Moscow Engineering Physics Institute in 2009 with a degree in Condensed Matter Physics and Photonics. Graduated from the National Research Nuclear University MEPhI in 2015 with a degree in Applied Mathematics and Informatics. Engineer of Financial monitoring Department of National Research Nuclear University MEPhI. Research interests: modeling of information interactions in documentary databases. ResearcherID (WoS): M-3125-2015, ORCID: 0000-0002-3780-8092. [lebedevalex@live.ru](mailto:lebedevalex@live.ru)

*Received March 15, 2021. Revised April 20, 2021. Accepted May 25, 2021.*

## Подход к созданию онтологий на основе электронных таблиц с произвольной структурой

А.В. Видия, Н.О. Дородных, А.Ю. Юрин

*Институт динамики систем и теории управления имени В.М. Матросова СО РАН, Иркутск, Россия*

### Аннотация

Использование семантических технологий, в том числе онтологий, является широко распространенной практикой при создании современных интеллектуальных систем. Электронные таблицы являются одним из наиболее доступных и распространенных способов представления и хранения информации, который характеризуется большим разнообразием и разнородностью компоновок, стилей и содержания, оставаясь ценным источником структурированных предметных знаний. В работе предлагается автоматизировать процесс формирования онтологий на основе анализа и преобразования электронных таблиц, обладающих произвольной структурой. Представлен подход, обеспечивающий восстановление семантики табличных данных, концептуализацию и формализацию табличного содержания в форме онтологии. Приведены основные этапы подхода и описание разработанных программных средств. Данные средства использовались при решении практической задачи формирования онтологии для диагностирования и оценки технического состояния нефтехимического оборудования. В качестве исходных данных использовались электронные таблицы, извлечённые из отчётов по экспертизе промышленной безопасности нефтехимических комплексов. По результатам апробации сделан вывод о целесообразности использования предлагаемого подхода при прототипировании предметных онтологий.

**Ключевые слова:** *приобретение знаний, семантические технологии, онтологический инжиниринг, онтология, электронная таблица, экспертиза промышленной безопасности.*

**Цитирование:** *Видия, А.В.* Подход к созданию онтологий на основе электронных таблиц с произвольной структурой / *А.В. Видия, Н.О. Дородных, А.Ю. Юрин* // *Онтология проектирования.* – 2021. – Т.11, №2(40). – С.212-226. – DOI: 10.18287/2223-9537-2021-11-2-212-226.

### Введение

Семантические технологии позволяют объединить разнородную информацию из различных источников и обеспечить её представление с помощью формальных стандартизованных средств моделирования знаний. Центральным элементом семантических технологий является онтология, которая может быть использована для эффективного представления знаний и перевода знаний в форму, пригодную для интерпретации компьютерами и людьми [1]. В основном онтология используются экспертами предметной области (ПрО), системными аналитиками и инженерами по знаниям на этапах концептуализации и формализации знаний. Построение онтологий остаётся сложной, творческой и трудоёмкой задачей. Многие годы ведутся исследования по повышению эффективности создания онтологий, начиная от разработки крупных методологий и проектов, таких как: *KACTUS* [2], *METHONTOLOGY* [3] *On-To-Knowledge* [4], *NeOn* [5], и заканчивая различными программными инструментами (например, *Protégé*, *Fluent Editor*, *OntoStudio*, *ONTOedit*, *WebOnto* и др.). Данные инструменты предоставляют широкий спектр возможности по моделированию онтологий и поддерживают основные стандарты консорциума *W3C*.

Автоматизация формирования онтологий на основе различных информационных источников, помимо экспертов (например, баз данных, текстов, концептуальных моделей), является

ся перспективным направлением, так как позволяет использовать большие объёмы накопленной информации, аккумулированной в различных организациях и представленной в разных форматах. В качестве такого источника могут выступать электронные таблицы (ЭТ), которые являются одним из наиболее доступных и распространённых способов представления и хранения информации. ЭТ характеризуются большим разнообразием и разнородностью компоновок, стилей и содержания, оставаясь ценным источником предметных знаний.

В работе предлагается подход для автоматизированного создания онтологий в формате *OWL2 DL* [6] на основе анализа и преобразования данных, извлечённых из ЭТ, представленных в формате *CSV*. Особенностью подхода является использование канонической (реляционной) формы для представления ЭТ, обладающих произвольной компоновкой, что обеспечивает унификацию входных данных. Подход использовался при формировании онтологии для интеллектуальной системы диагностирования и оценки технического состояния нефтехимического оборудования в рамках задачи экспертизы промышленной безопасности (ЭПБ).

## 1 Состояние вопроса

В настоящее время накоплен большой объём ЭТ, представленных в различных форматах: *HTML*, *EXCEL*, *CSV* и др. Современные оценки показывают, что их количество исчисляется, как минимум, сотнями миллионов [7], из которых можно извлечь миллиарды ценных фактов. Это делает ЭТ привлекательным источником знаний для различных приложений, в том числе для создания вопросно-ответных и рекомендательных систем, конструирования баз знаний, автодополнения ЭТ и др. В связи с этим два последних десятилетия активно развиваются методологические основы извлечения и преобразования данных из ЭТ.

Особый интерес представляют подходы и программные средства извлечения связанных данных и онтологий из ЭТ. В частности, существует ряд программных инструментов, позволяющих извлекать *RDF*-триплеты из ЭТ (например, *RDF123* [8], *csv2rdf4lod* [9], *Datalift* [10], *Spread2RDF* [11]), а также формировать *OWL*-онтологии (например, *Owlifier* [12], *Populous* [13], *Any2OWL* [14], *Excel2OWL* [15]). Некоторые из существующих решений включают:

- специальные предметно-ориентированные языки для описания преобразования ЭТ в наборы связанных данных или онтологии, например: *XLWrap* [16], *Mapping Master* [17], *RML* [18] и *PEARL* (плагин *Sheet2RDF* для фреймворка «*Semantic Turkey*») [19];
- языки и средства на основе расширения языка запросов *SPARQL*, в частности, *XSPARQL* [20], *SPARQL-Generate* [21] и *Tarql* [22];
- предметно-ориентированные языки, основанные на расширении различных языков программирования высокого уровня, например, *Ruby* [11].

Следует также отметить, что в данной области существует стандарт *W3C*, определяющий подход *CSVW* [23] для генерации *RDF*-документов на основе табличных данных в формате *CSV*. Также консорциум публикует список существующих конвертеров табличных данных в *RDF*-формат на сайте [24].

Несмотря на значительный прогресс в области преобразования табличных данных, существующие программные решения обладают следующими недостатками:

- Использование определённых моделей исходных ЭТ со смешанными физическими и логическими схемами. Данный факт ограничивает использование этих инструментов для обработки произвольных таблиц, представленных, например, в различных статистических отчётах. Хотя некоторые решения, основанные на применении предметно-ориентированных языков (например, *Mapping Master*, *SPARQL-Generate* и др.), предлагают достаточно гибкий механизм преобразования произвольных табличных данных, но они сложны в понимании и применении (пользователю необходимо знать синтаксис дан-

ных языков и правила построения соответствий между элементами двух форматов), а также зачастую отсутствуют средства визуальной поддержки преобразования и т.д.

- Слабая поддержка формирования полных агрегированных онтологий на основе анализа и преобразования множества ЭТ, принадлежащих к одной ПрО, т.е., как правило, средства поддерживают единичную трансформацию таблицы в онтологию (или некоторый её фрагмент) и не предполагают объединение полученных онтологий (фрагментов) в пределах обработки одного набора таблиц.
  - Практически все подходы ориентированы на программистов и специалистов в области семантических технологий и не предполагают использования непрограммирующими пользователями (например, экспертами ПрО).
  - Направленность средств на генерацию наборов связанных данных в формате *RDF* и слабая поддержка формата *OWL*. Как правило, в работах не уточняется, какую конкретную версию *RDF* и *OWL* поддерживает средство.
  - Отсутствие проверки качества полученных онтологий и наборов связанных данных.
- В статье предлагается подход к быстрому прототипированию онтологий на основе анализа и преобразования данных, извлечённых из ЭТ, обладающих произвольной структурой.

## 2 Предлагаемый подход

### 2.1 Постановка задачи

Постановку задачи можно формализовать следующим образом: необходимо определить оператор  $T$  преобразования произвольных ЭТ:

$$T: AS^{CSV} \rightarrow Ont^{OWL}, \quad (1)$$

где  $AS^{CSV}$  – исходная произвольная ЭТ в формате *CSV*;  $Ont^{OWL}$  – целевая онтология в формате *OWL2 DL*.

Оператор преобразования (1) может быть представлен в виде следующих этапов:

$$T = \langle T_{AS-CS}, T_{CS-OM}, T_{OM-Ont} \rangle,$$

$$T_{AS-CS}: AS^{CSV} \rightarrow CS^{CSV}, T_{CS-OM}: CS^{CSV} \rightarrow OM, T_{OM-Ont}: OM \rightarrow Ont^{OWL}, \quad (2)$$

где  $CS^{CSV}$  – исходная ЭТ, представленная в реляционной форме;  $OM$  – онтологическая модель, представляющая собой описание ПрО на терминологическом (*TBox*) и аксиоматическом (*ABox*) уровнях;  $T_{AS-CS}$  – набор правил преобразования исходной произвольной ЭТ в формате *CSV* в реляционную форму;  $T_{CS-OM}$  – набор правил преобразования реляционной ЭТ в онтологическую модель;  $T_{OM-Ont}$  – набор правил преобразования онтологической модели в код онтологии в формате *OWL2 DL*.

### 2.2 Описание этапов предлагаемого подхода

Для автоматизированного формирования онтологий на основе анализа и преобразования произвольных ЭТ предлагается подход, который может быть представлен в виде последовательности действий (см. рисунок 1).

**Этап 1.** Анализ и трансформация ЭТ в каноническую форму.

Задача этапа - перейти от таблиц с произвольной структурой к однотипному реляционному представлению. Данное преобразование ( $T_{AS-CS}$ ) включает фазы: распознавание, ролевой (функциональный) и структурный анализ [25]. Для этого используется следующая структура таблицы:

$$CS^{CSV} = \{D, RH, CH\},$$

где  $D$  – блок данных, который описывает конкретные значения данных (записи), принадлежащие к одному и тому же типу данных (например, числовые, текстовые и т.д.);  $RH$  – набор заголовков строк, т.е.  $RH = \{rh^1, \dots, rh^n\}$ , где  $rh^i$  –  $i$ -ячейка заголовка строки;  $CH$  – набор заголовков столбцов, т.е.  $CH = \{ch^1, \dots, ch^m\}$ , где  $ch^j$  –  $j$ -ячейка заголовка столбца. Значения в ячейках блоков заголовков могут быть разделены символом «|», с помощью которого осуществляется представление иерархических отношений между заголовками (разделение заголовков на подзаголовки). Эта структура основана на представлении таблиц, предложенном в работе [26] и адаптированном для системы *TabbyXL* [27], которая используется на данном этапе.

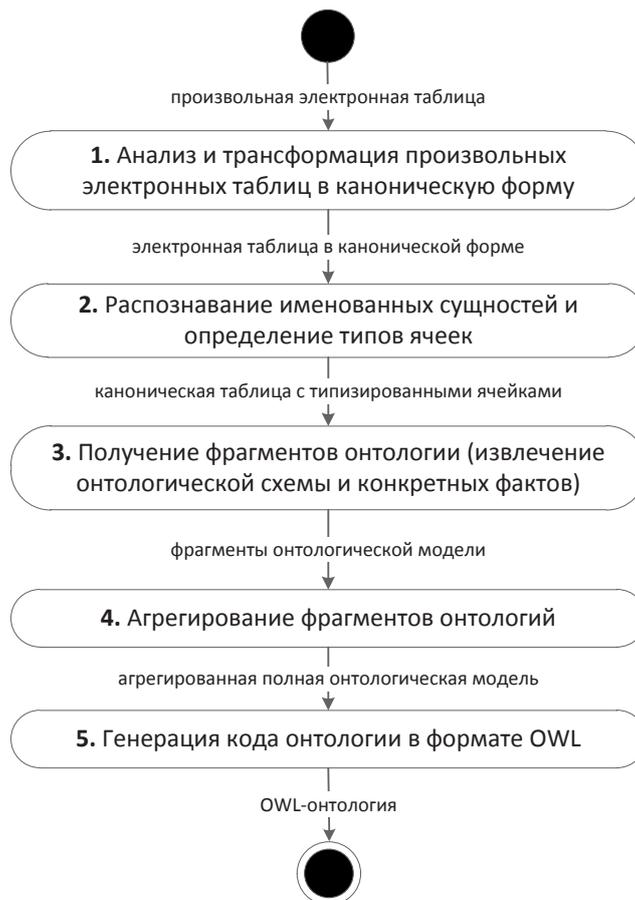


Рисунок 1 – Основные этапы предлагаемого подхода

Для реализации преобразования  $T_{AS-CS}$  используется предметно-ориентированный язык *Cells Rule Language (CRL)* [25]. При этом набор правил может быть реализован для конкретной задачи с учётом требований к исходным и целевым данным. Был сформирован набор *CRL*-правил для двух выделенных форм таблиц, описанных в [28].

На данном этапе осуществляется подготовка данных канонической ЭТ для дальнейшей обработки, включающая:

- исправление «битых» символов Юникода;
- удаление различных «мусорных» символьных значений, кроме букв и цифр;
- расшифровка акронимов;
- удаление множественных пробелов;
- идентификация и удаление единиц измерения и т.п.

**Эман 2.** Распознавание именованных сущностей и определение типов ячеек.

На данном этапе осуществляется процедура извлечения и распознавания именованных сущностей, содержащихся в ячейках канонической ЭТ. Для этого используется библиотека для обработки естественного языка – *Stanford CoreNLP* и, в частности, *Java*-реализация распознавателя *Stanford Named-Entity Recognizer (Stanford NER)* [29]. *Stanford NER* отмечает слова в тексте, которые являются названиями объектов, такие как имена людей и компаний или наименования городов и стран. *Stanford NER* определяет множество классов именованных сущностей. В работе использовалось 8 классов (типов): *Location, Country, City, Person, Organization, Number, Percent, Date*. Эти классы присваиваются каждой ячейке в канонической таблице, характеризуя данные, которые она содержит. Используется неопределённый класс – *None*, который присваивается ячейке, не отнесённой ни к одному из определённых классов. Определённые типизированные ячейки подразделяются на два вида: ячейки с именованными сущностями (*named-entity cells*) и ячейки с литеральными значениями (*literal cells*). Такая типизация ячеек позволяет отнести табличные данные к разным онтологическим уровням (уровень свойств классов и конкретных экземпляров) на последующем этапе преобразования.

**Эман 3.** Получение фрагментов онтологии.

Задача этого этапа – получить онтологические фрагменты в виде набора классов, их отношений (объектных свойств – *object properties*) и свойств-значений (*datatype properties*), а также конкретных экземпляров (фактов – *instances*), которые описывают определённую ПрО, на основе анализа и трансформации данных канонических ЭТ. Получение фрагментов онтологии осуществляется в два последовательных шага.

На первом шаге из канонической ЭТ извлекается онтологическая схема (терминологический уровень онтологии – *TBox*) на основе заголовков (*RH* и *CH*). Анализ осуществляется построчно. В подходе используются следующие основные эвристические правила преобразования канонических ЭТ (*TCS-OM*).

**Правило 1:** ЕСЛИ задан только *CH* И он содержит одно значение в  $ch^i$ , ТО  $ch_1^i$  преобразуется в класс.

**Правило 2:** ЕСЛИ задан только *CH* И он содержит два или более значений с разделителем («|») в  $ch^i$ , т.е.  $ch^i = \{ch_1^i, \dots, ch_m^i\}$ , ТО  $ch_1^i$  преобразуется в класс со свойствами-значений, соответствующими следующим значениям в множестве  $ch^i$ .

**Правило 3:** ЕСЛИ *CH* содержит одно значение в  $ch^i$  И оно соответствует только одному *RH*, содержащему одно значение в  $rh^i$ , ТО  $ch_1^i$  преобразуется в класс со свойством-значением  $rh_1^i$ .

**Правило 4:** ЕСЛИ *CH* содержит два или более значений с разделителем («|»), т.е.  $ch^i = \{ch_1^i, \dots, ch_m^i\}$  И они соответствуют только одному *RH*, содержащему одно значение в  $rh^i$ , ТО  $ch_1^i$  преобразуется в класс со свойством-значением из  $rh_1^i$  и с дополнительными свойствами-значениями, которые соответствуют следующим значениям в множестве  $ch^i$ .

**Правило 5:** ЕСЛИ *CH* И *RH* содержат два или более значений с разделителем («|»), т.е.  $ch^i = \{ch_1^i, \dots, ch_m^i\}$  И  $rh^i = \{rh_1^i, \dots, rh_n^i\}$ , ТО  $ch_1^i$  И  $rh_1^i$  преобразуются в соответствующие классы и указывается связь между ними.

**Правило 6:** ЕСЛИ *CH* содержит одно значение в  $ch^i$  И соответствует двум и более значениям *RH* с разделителем («|»), т.е.  $rh^i = \{rh_1^i, \dots, rh_n^i\}$ , ТО  $ch_1^i$  преобразуется в класс со свойством-значением  $rh_1^i$ , а все последующие значения из множества  $rh^i$  преобразуются в соответствующие классы и указывается связь между ними.

Подобные правила разработаны и для ситуации, где  $CH$  и  $RH$  меняются местами, т.е. структура классов формируется исходя из меток в  $RH$ . При этом все полученные иерархические связи заголовков интерпретируются как объектные свойства (отношения между классами). По умолчанию конкретные значения для свойств-значений устанавливаются на основе записей из  $D$ , которые определены на этапе 3 как литеральные.

На втором шаге из канонической ЭТ извлекаются конкретные экземпляры (аксиоматический уровень онтологии –  $ABox$ ) на основе  $D$ . При этом учитываются только те значения ячеек, которые на этапе 3 были классифицированы как ячейки, содержащие именованные сущности.

Основным результатом этого этапа являются фрагменты онтологической модели. Эти фрагменты необходимо агрегировать, включая операции по уточнению названий классов, их свойств и отношений, а также их возможное слияние и разделение.

**Этап 4.** Агрегация фрагментов онтологии.

Задача этого этапа - объединение полученных фрагментов онтологии в единую агрегированную  $OM$ . Данная модель предназначена для унифицированного представления и хранения знаний, извлечённых из различных информационных источников. Модель позволяет абстрагироваться от особенностей описания знаний на различных языках и их диалектах, используемых при реализации онтологий (например,  $OWL$ ,  $RDFS$ ). Для автоматического агрегирования фрагментов  $OM$  используются следующие основные эвристические правила.

- Классы с одинаковыми именами объединяются, формируя общий набор свойств-значений, объектных свойств и экземпляров.
- Дублирующие классы с одинаковыми именами и структурой свойств удаляются.
- Классы с похожими именами объединяются. Полученные фрагменты  $OM$  могут описывать одни и те же объекты или процессы. Предлагается использовать метод сравнения строк, основанный на расстоянии Левенштейна, чтобы определить сходство между двумя именами классов. Если расстояние Левенштейна лежит в диапазоне от 0 до 3, то классы считаются подобными. Этого может быть недостаточно, поэтому учитывается структура классов (названия свойств должны частично совпадать).
- Создание новых объектных свойств (отношений между классами), если существуют одноимённые классы и свойства-значения. При этом создаётся новый класс с именем свойства-значения, а одноименное свойство-значение удаляется.
- Повторяющиеся объектные свойства между классами удаляются.
- Дублирующие свойства-значения удаляются.
- Повторяющиеся экземпляры удаляются.

**Этап 5.** Генерация  $OWL$ -кода онтологии.

Задача данного этапа - генерация кода онтологии в формате  $OWL2 DL$  на основе полученной  $OM$ . Сгенерированный  $OWL$ -код онтологии может быть модифицирован и дополнен с помощью различных редакторов онтологического моделирования, например, *Protégé*.

## 2.3 Программная реализация

Первый этап подхода реализован с использованием *TabbyXL* - консольного *Java*-приложения [27], которое обрабатывает файлы ЭТ в формате *MS Excel (XLSX)* или *CSV*. Каждый файл может содержать одну или несколько ЭТ с произвольной структурой. *TabbyXL* использует *CRL*-правила для преобразования данных, извлекаемых из таблиц, в каноническую форму. Преобразованные данные сохраняются в отдельных *CSV*-файлах.

Остальные этапы подхода реализованы в форме программного модуля расширения – *PKBD.Onto* [30] для системы прототипирования экспертных систем *Personal Knowledge Base Designer (PKBD)* [31]. *PKBD* реализован в форме настольного приложения, ориентированно-

го на непрограммирующих пользователей. Основная цель *PKBD* – это создание прототипов баз знаний, использующих формализм логических правил и онтологий. *PKBD* обладает модульной архитектурой, которая обеспечивает динамическое подключение модулей поддержки различных языков представления знаний и интеграции с инструментами концептуального и онтологического моделирования при импорте и экспорте понятий и отношений. *PKBD.Onto* представляет собой динамическую библиотеку (*Dynamic Link Library*) подключаемую через унифицированный интерфейс *API PKBD*.

### 3 Применение

Разработанный подход применялся в пилотном проекте Иркутского научно-исследовательского и конструкторского института химического и нефтяного машиностроения (ИркутскНИИХиммаш) при создании программного обеспечения для поддержки решения задач диагностирования и оценки технического состояния нефтехимического оборудования [32, 33]. Решаемые задачи являлись частью процедуры ЭПБ, которая заключается в подтверждении соответствия технического объекта требованиям промышленной безопасности. Анализ данной процедуры показал, что реализация этапов «разработка программы ЭПБ», «анализ и интерпретация результатов», «принятие решений по ремонту» требует обработки больших объёмов плохо формализованной информации. При этом эффективность данной обработки может быть повышена с помощью специализированных интеллектуальных систем, которые позволяют:

- интерпретировать условия и параметры функционирования технических систем;
- обосновать программу технической диагностики;
- интерпретировать параметры диагностики технического состояния технических систем.

Для решения этих задач была построена онтология на основе концептуальных моделей, разработанных экспертами-предметниками, а также ЭТ, которые использовались в качестве дополнительного источника информации. Создание данной онтологии на основе анализа и преобразования ЭТ с использованием предлагаемого подхода и средств содержит следующие этапы.

**Этап 1.** Анализ и трансформация ЭТ в каноническую форму.

Исходные ЭТ были извлечены из шести отчётов по ЭПБ. Для этого использовались отчёты АО ИркутскНИИХиммаш, в частности, разделы, содержащие информацию о карте исходных данных по технической диагностике, результатах диагностики, расчёте прочности и остаточного ресурса. В результате анализа было извлечено 216 ЭТ, 173 из которых обладали уникальными структурой и содержанием и состояли из 5817 ячеек. Было определено два основных типа структур отобранных ЭТ.

В таблице 1 приведён пример фрагмента исходной ЭТ, извлечённой из отчёта по ЭПБ. Таблица описывает толщину стенок для основных конструктивных элементов.

В таблице 2 приведён пример фрагмента исходной ЭТ, который описывает исходные данные и результаты расчёта показателей для основных конструктивных элементов, где:  $P_p$  – давление расчётное;  $D$  – диаметр внутренний;  $T$  – температура (расчётная);  $[G]$  – допустимое напряжение материала при расчётной температуре;  $S_f$  – толщина стенки (фактическая);  $S_p$  – толщина стенки (расчётная);  $[P]$  – допустимое внутреннее избыточное давление.

*CRL*-правила были сформированы для обработки выделенных типов структур ЭТ. Таким путём были отобраны 161 ЭТ из 173 и преобразованы в каноническую форму с помощью *TabbyXL*. Оценки точности, полноты и *F*-меры составили: 0.99, 0.87 и 0.93 соответственно. *TabbyXL* показал высокую эффективность при обработке ЭТ, извлечённых из отчётов ЭПБ. «Мусорные» символы и единицы измерения были удалены из ячеек канонической ЭТ. В таб-

лицах 3 и 4 показаны фрагменты соответствующих ЭТ в каноническом виде, полученные из исходных таблиц (см. таблицы 1 и 2). Блок *D* представлен столбцом с заголовком «*DATA*», блок *CH* представлен столбцом с заголовком «*Column Heading*», а блок *RH* отсутствует.

Таблица 1 – Пример фрагмента исходной ЭТ «Толщина стенок для основных конструктивных элементов»

Конструктивный элемент	Минимальная толщина стенки по толщинометрии <i>S<sub>min</sub></i> , мм	Погрешность измерения, мм	Фактическая толщина стенки <i>S<sub>ф</sub></i> , мм
Обечайка корпуса межтрубного пространства	23,7	+0,1	23,6
Обечайка штуцера №1 ввода трубного пучка межтрубного пространства	31,7	+0,1	31,6
Обечайка штуцера №2 ввода трубного пучка межтрубного пространства	31,7	+0,1	31,6
Обечайка штуцера №3 ввода трубного пучка межтрубного пространства	31,7	+0,1	31,6
Обечайка распределительной камеры №1 трубного пространства	10,0	+0,1	9,9
Обечайка распределительной камеры №2 трубного пространства	10,0	+0,1	9,9
Обечайка распределительной камеры №3 трубного пространства	10,0	+0,1	9,9

Таблица 2 – Пример фрагмента исходной ЭТ «Исходные данные и результаты расчета»

Конструктивный элемент	<i>P<sub>p</sub></i> , МПа	<i>D</i> , мм	<i>T</i> , С	<i>[G]</i> , МПа	<i>S<sub>ф</sub></i> , мм	<i>S<sub>p</sub></i> , мм	<i>[P]</i> , МПа
Обечайка корпуса межтрубного пространства	1.25	2600,0	195	142,3	23,6	12,75	2,21
Обечайка штуцера №1 ввода трубного пучка межтрубного пространства	1.25	700,0	195	148,6	31,6	3,3	11,2
Обечайка штуцера №2 ввода трубного пучка межтрубного пространства	1.25	700,0	195	148,6	31,6	3,3	11,2
Обечайка штуцера №3 ввода трубного пучка межтрубного пространства	1.25	700,0	195	148,6	31,6	3,3	11,2
Обечайка распределительной камеры №1 трубного пространства	1.2	700,0	270	140,6	9,9	3,3	3,18
Обечайка распределительной камеры №2 трубного пространства	1.2	700,0	270	140,6	9,9	3,3	3,18
Обечайка распределительной камеры №3 трубного пространства	1.2	700,0	270	140,6	9,9	3,3	3,18

**Этап 2.** Распознавание именованных сущностей и определение типов ячеек.

В таблицах 3b и 4b приведены фрагменты канонических ЭТ с распознанными именованными сущностями. Все ячейки данных таблиц со значением *NONE* интерпретированы как ячейки с именованными сущностями, а *NUMBER* - как ячейки с литеральными значениями.

**Этап 3.** Получение фрагментов онтологии.

Для формирования фрагментов онтологической схемы на основе канонических таблиц 3a и 4a были применены *Правила 1* и *2*. Следует отметить, что экземпляры были извлечены из ячеек, содержащих только именованные сущности. Список полученных экземпляров из фрагментов таблиц:

- обечайка Корпуса Межтрубного Пространства;
- обечайка Штуцера 1 Ввода Трубного Пучка Межтрубного Пространства;
- обечайка Штуцера 2 Ввода Трубного Пучка Межтрубного Пространства;
- обечайка Штуцера 3 Ввода Трубного Пучка Межтрубного Пространства;
- обечайка Распределительной Камеры 1 Трубного Пространства;
- обечайка Распределительной Камеры 2 Трубного Пространства;
- обечайка Распределительной Камеры 3 Трубного Пространства.

Таблица 3 – Фрагменты ЭТ: а) в каноническом виде, полученной из таблицы 1; б) с распознанными сущностями

а)		б)	
DATA	Column Heading	DATA	Column Heading
обечайка корпуса межтрубного пространства	конструктивный элемент	NONE	NONE
23,7	минимальная толщина стенки по толщимоетрии s <sub>min</sub>	NUMBER	NONE
+0,1	погрешность измерения	NUMBER	NONE
23,6	фактическая толщина стенки s <sub>ф</sub>	NUMBER	NONE
обечайка штуцера №1 ввода трубного пучка межтрубного пространства	конструктивный элемент	NONE	NONE
31,7	минимальная толщина стенки по толщимоетрии s <sub>min</sub>	NUMBER	NONE
+0,1	погрешность измерения	NUMBER	NONE
31,6	фактическая толщина стенки s <sub>ф</sub>	NUMBER	NONE
обечайка штуцера №2 ввода трубного пучка межтрубного пространства	конструктивный элемент	NONE	NONE
31,7	минимальная толщина стенки по толщимоетрии s <sub>min</sub>	NUMBER	NONE
+0,1	погрешность измерения	NUMBER	NONE
31,6	фактическая толщина стенки s <sub>ф</sub>	NUMBER	NONE
обечайка штуцера №3 ввода трубного пучка межтрубного пространства	конструктивный элемент	NONE	NONE
31,7	минимальная толщина стенки по толщимоетрии s <sub>min</sub>	NUMBER	NONE
+0,1	погрешность измерения	NUMBER	NONE
31,6	фактическая толщина стенки s <sub>ф</sub>	NUMBER	NONE
обечайка распределительной камеры №1 трубного пространства	конструктивный элемент	NONE	NONE
10,0	минимальная толщина стенки по толщимоетрии s <sub>min</sub>	NUMBER	NONE
+0,1	погрешность измерения	NUMBER	NONE
9,9	фактическая толщина стенки s <sub>ф</sub>	NUMBER	NONE
обечайка распределительной камеры №2 трубного пространства	конструктивный элемент	NONE	NONE
10,0	минимальная толщина стенки по толщимоетрии s <sub>min</sub>	NUMBER	NONE
+0,1	погрешность измерения	NUMBER	NONE

Таблица 4 – Фрагменты ЭТ: а) в каноническом виде, полученной из таблицы 2; б) с распознанными сущностями

а)		б)	
DATA	Column Heading	DATA	Column Heading
обечайка корпуса межтрубного пространства	конструктивный элемент	NONE	NONE
1.25	pp	NUMBER	NONE
2600,0	d	NUMBER	NONE
195	t	NUMBER	NONE
142,3	g	NUMBER	NONE
23,6	s <sub>ф</sub>	NUMBER	NONE
12,75	sp	NUMBER	NONE
2,21	p	NUMBER	NONE
обечайка штуцера №1 ввода трубного пучка межтрубного пространства	конструктивный элемент	NONE	NONE
1.25	pp	NUMBER	NONE
700,0	d	NUMBER	NONE
195	t	NUMBER	NONE
148,6	g	NUMBER	NONE
31,6	s <sub>ф</sub>	NUMBER	NONE
3,3	sp	NUMBER	NONE
11,2	p	NUMBER	NONE
обечайка штуцера №2 ввода трубного пучка межтрубного пространства	конструктивный элемент	NONE	NONE
1.25	pp	NUMBER	NONE
700,0	d	NUMBER	NONE
195	t	NUMBER	NONE
148,6	g	NUMBER	NONE
31,6	s <sub>ф</sub>	NUMBER	NONE
3,3	sp	NUMBER	NONE

Остальные значения в столбце «DATA» были интерпретированы как конкретные значения для свойств-значений в классе «КонструктивныйЭлемент». Для коротких названий свойств-значений (до 3 символов) в начале названия добавляется приставка «имеет». Полученные фрагменты онтологии представлены на рисунке 2.



Рисунок 2 – Примеры полученных фрагментов онтологии

**Этап 4.** Агрегация фрагментов онтологии.

Пример агрегации полученных отдельных фрагментов онтологии в единую *ОМ* представлен на рисунке 3. Следует отметить, что при объединении классов «*КонструктивныйЭлемент*» также произошло объединение их экземпляров.



Рисунок 3 – Пример агрегированной онтологической модели на основе полученных фрагментов онтологии

**Этап 5.** Генерация OWL-кода онтологии.

На основе полученной *ОМ* (см. рисунок 3) была произведена генерация кода онтологии в формате *OWL2 DL*. В результате преобразования полученных канонических таблиц с помощью *PKBD* выделено: 87 классов, 32 объектных свойства (отношений между классами), 429 свойств-значений.

**Заключение**

В статье описан подход для автоматизированного формирования онтологий, как на уровне структуры классов (*ТВох*), так и на уровне конкретных экземпляров (*АВох*). В качестве исходных данных использованы ЭТ, обладающие произвольной компоновкой и приведённые к канонической форме. Предлагаемый подход реализован в форме плагина для инструментального средства *PKBD*. Полученные *OWL*-коды являются синтаксически корректными, при этом содержательную оценку результатов должен выполнять конечный пользователь (эксперт).

Применение предлагаемого подхода осуществлено в пилотном проекте для АО ИркутскНИИхиммаш при решении задач диагностирования и оценки технического состояния нефтехимического оборудования и технологических комплексов в рамках ЭПБ. Полученные результаты показывают перспективность использования предлагаемого подхода для поддержки прототипирования онтологий.

**Благодарности**

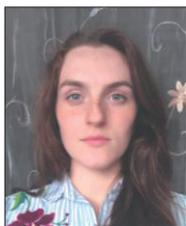
Работа выполнена при финансовой поддержке Совета по грантам Президента России (проект МК-1647.2020.9).

## СПИСОК ИСТОЧНИКОВ

- [1] **Гаврилова, Т.А.** Инженерия знаний. Модели и методы / Т.А. Гаврилова, Д.В. Кудрявцев, Д.И. Муромцев. – СПб.: Лань, 2016. – 324 с.
- [2] **Schreiber, G.** The KACTUS View on the ‘O’ Word / G. Schreiber, B.J. Wielinga, W.N.H. Jansweijer // In: Proceedings of IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing, 1995. - P.159-168.
- [3] **Lopez, M.F.** Building a chemical ontology using METHONTOLOGY and the ontology design environment / M.F. Lopez, A. Gomez-Perez, J.P. Sierra, A.P. Sierra // IEEE Intelligent Systems. - 1999. - Vol. 14(1). - P.37-46.
- [4] **Staab, S.** Knowledge processes and ontologies / S. Staab, R. Studer, H.-P. Schnurr, Y. Sure // IEEE Intelligent Systems. - 2001. - Vol. 16(1). - P.26-34.
- [5] **Suárez-Figueroa, M.C.** Ontology engineering in a networked world / M.C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, A. Gangemi // Springer-Verlag Berlin Heidelberg. - 2012.
- [6] **Grau, B.C.** OWL 2: The next step for OWL / B.C. Grau, I. Horrocks, B. Motik, B. Parsia, P. Patel-Schneider, U. Sattler // Web Semantics: Science, Services and Agents on the World Wide Web. - 2008. - Vol. 6(4). - P.309-322.
- [7] Web Data Commons. - <http://webdatacommons.org>.
- [8] **Han, L.** RDF123: From spreadsheets to RDF / L. Han, T. Finin, C. Parr, J. Sachs, A. Joshi // In: Proceedings of the 7th International Semantic Web Conference (ISWC). Lecture Notes in Computer Science, vol. 5318. - Springer-Verlag Berlin Heidelberg, 2008. - P.451-466.
- [9] **Lebo, T.** Converting governmental datasets into Linked Data / T. Lebo, G. Williams // In: Proceedings of the 6th International Conference on Semantic Systems, 2010. - P.1-3.
- [10] **Scharffe, F.** Enabling Linked Data publication with the Datalift platform / F. Scharffe, G. Atemezing, R. Troncy, F. Gandon, S. Villata, B. Bucher, F. Hamdi, L. Bihanic, G. Képéklían, F. Cotton, J. Euzenat, Z. Fan, P.Y. Vandebussche, B. Vatant // In: Proceedings of the AAAI workshop on semantic cities. In 26th Conference on Artificial Intelligence, W10: Semantic Cities. - 2012. - P.25-30.
- [11] Spread2RDF. - <https://github.com/marcelotto/spread2rdf>.
- [12] **Bowers, S.** Owlifier: Creating OWL-DL ontologies from simple spreadsheet-based knowledge descriptions / S. Bowers, J.S. Madin, M.P. Schildhauer // Ecological Informatics. - 2010. - Vol. 5(1). - P.19-25.
- [13] **Jupp, S.** Populous: A tool for building OWL ontologies from templates / S. Jupp, M. Horridge, L. Iannone, J. Klein, S. Owen, J. Schanstra, K. Wolstencroft, R. Stevens // BMC Bioinformatics. - 2011. - Vol. 13. - P.1-12.
- [14] **Zhang, X.** Ontology based data conversion from spreadsheet to OWL / X. Zhang, R. Di, X. Feng // In: Proceedings of the 2012 Seventh China Grid Annual Conference. - 2012. - P.76-79.
- [15] **Tahar, K.** An Approach to support collaborative ontology construction / K. Tahar, M. Schaaf, F. Jahn, C. Kücherer, B. Paech, H. Herre, A. Winter // Studies in health technology and informatics. - 2016 - Vol. 228. - P.369-373.
- [16] **Langegger, A.** XLWrap - Querying and Integrating Arbitrary Spreadsheets with SPARQL / A. Langegger, W. Wob // In: Proceedings of the 8th International Semantic Web Conference (ISWC). Lecture Notes in Computer Science, vol. 5823. - Springer-Verlag Berlin Heidelberg, 2009. - P.359-374.
- [17] **O’Connor, M.J.** Mapping Master: A flexible approach for mapping spreadsheets to OWL / M.J. O’Connor, C. Halaschek-Wiener, M.A. Musen // In: Proceedings of the 9th International Semantic Web Conference (ISWC). Lecture Notes in Computer Science, vol. 6497. - Springer-Verlag Berlin Heidelberg, 2010. - P.194-208.
- [18] **Dimou, A.** RML: A generic language for integrated RDF mappings of heterogeneous data / A. Dimou, M.V. Sande, P. Colpaert, R. Verborgh, E. Mannens, R.V.D. Walle // In: Proceedings of the Workshop on Linked Data on the Web, Co-located with the 23rd International World Wide Web Conference. - 2014. - P.1-5.
- [19] **Fiorelli, M.** Sheet2RDF: A flexible and dynamic spreadsheet import & lifting framework for RDF / M. Fiorelli, T. Lorenzetti, M.T. Paziienza, A. Stellato, A. Turbati // In: Proceedings of the 28th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Lecture Notes in Computer Science, vol. 9101. - Springer-Verlag Berlin Heidelberg, 2015. - P.131-140.
- [20] **Bischof, S.** Mapping between RDF and XML with XSPARQL / S. Bischof, S. Decker, T. Krennwallner, N. Lopes, A. Polleres // Journal on Data Semantics. - 2012. - Vol. 1(3). - P.147-185.
- [21] **Lefrançois, M.** A SPARQL extension for generating RDF from heterogeneous formats / M. Lefrançois, A. Zimmermann, N. Bakerally // In: Proceedings of the 14th International Conference, European Semantic Web Conference (ESWC). Lecture Notes in Computer Science, vol. 10249. - Springer-Verlag Berlin Heidelberg, 2017. - P.35-50.
- [22] Tarql: SPARQL for Tables. - <http://tarql.github.io/>.
- [23] Generating RDF from Tabular Data on the Web. - <https://www.w3.org/TR/csv2rdf/>.
- [24] W3C ConverterToRdf. - <https://www.w3.org/wiki/ConverterToRdf>.

- [25] **Shigarov, A.O.** Rule-based spreadsheet data transformation from arbitrary to relational tables / A.O. Shigarov, A.A. Mikhailov // Information Systems. - 2017. - Vol. 71. - P.123-136.
- [26] **Tijerino, Y.A.** Towards ontology generation from tables / Y.A. Tijerino, D.W. Embley, D.W. Lonsdale, Y. Ding, G. Nagy // World Wide Web: Internet and Web Information Systems. - 2005. - Vol. 8(8). - P.261-285.
- [27] **Shigarov, A.O.** TabbyXL: Software platform for rule-based spreadsheet data extraction and transformation / A.O. Shigarov, V.V. Khristyuk, A.A. Mikhailov // SoftwareX. - 2019. - Vol. 10. - 100270.
- [28] **Dorodnykh, N.O.** Conceptual model engineering for industrial safety inspection based on spreadsheet data analysis / N.O. Dorodnykh, A.Yu. Yurin, A.O. Shigarov // In: Proceedings of the 6th International Conference on Modeling and Development of Intelligent Systems (MDIS 2019). Communications in Computer and Information Science, vol. 1126. - Springer-Verlag Berlin Heidelberg, 2020. - P.51-65.
- [29] Stanford Named Entity Recognizer. - <https://nlp.stanford.edu/software/CRF-NER.html>.
- [30] **Дородных, Н.О.** Разработка схем онтологий на основе преобразования электронных таблиц / Н.О. Дородных, А.В. Видия, А.Ю. Юрин // Программные продукты и системы. - 2021. - № 1. - С.124-131.
- [31] **Yurin, A.Yu.** Personal knowledge base designer: Software for expert systems prototyping / A.Yu. Yurin, N.O. Dorodnykh // SoftwareX. - 2020. - Vol. 11. - 100411.
- [32] **Берман, А.Ф.** Информационно-аналитическая поддержка экспертизы промышленной безопасности объектов химии, нефтехимии и нефтепереработки / А.Ф. Берман, К.А. Кузнецов, О.А. Николайчук, А.И. Павлов, А.Ю. Юрин // Химическое и нефтегазовое машиностроение. - 2018. - № 8. - С.30-36.
- [33] **Грищенко, М.А.** Разработка интеллектуальных диагностических систем на основе онтологий / М.А. Грищенко, Н.О. Дородных, С.А. Коршунов, А.Ю. Юрин // Онтология проектирования. - 2018. - Т. 8. - № 2(28). - С.265-284. DOI: 10.18287/2223-9537-2018-8-2-265-284.

## Сведения об авторах



**Видия Анастасия Владимировна**, 1996 г. рождения. Обучается в Институте информационных технологий и анализа данных Иркутского национального исследовательского технического университета (ИрНИТУ) по направлению Информационные системы и технологии. Программист Института динамики систем и теории управления имени В.М. Матросова СО РАН (ИДСТУ СО РАН). [Vidiya\\_av@mail.ru](mailto:Vidiya_av@mail.ru).

**Дородных Никита Олегович**, 1990 г. рождения. Окончил ИрНИТУ (2012), к.т.н. (2018). Старший научный сотрудник ИДСТУ СО РАН. В списке научных трудов около 70 работ в области автоматизации создания интеллектуальных систем и баз знаний, получения знаний на основе преобразования концептуальных моделей и электронных таблиц. ORCID: 0000-0001-7794-4462; Author ID (RSCI): 979843; Author ID (Scopus): 57202323578; Researcher ID (WoS): E-8870-2014. [tualatin32@mail.ru](mailto:tualatin32@mail.ru).



**Юрин Александр Юрьевич**, 1980 г. рождения. Окончил ИрНИТУ (2002), к.т.н. (2005). Заведующий лабораторией Информационных технологий исследования природной и техногенной безопасности ИДСТУ СО РАН, доцент Института информационных технологий и анализа данных ИрНИТУ. Член Российской ассоциации искусственного интеллекта и Ассоциации вычислительной техники. Член редколлегии международного научного журнала «Computer, Communication & Collaboration». В списке научных трудов более 100 работ в области разработки систем поддержки принятия решений, экспертных систем и баз знаний, использования прецедентного подхода и семантических технологий

при проектировании интеллектуальных диагностических систем. ORCID: 0000-0001-9089-5730; Author ID (RSCI): 174845; Author ID (Scopus): 16311168300; Researcher ID (WoS): A-4355-2014. [iskander@icc.ru](mailto:iskander@icc.ru).

Поступила в редакцию 15.05.2021, после рецензирования 16.06.21 Принята к публикации 21.06.2021.

## An approach to creating freeform spreadsheet ontology

A.V. Vidia, N.O. Dorodnykh, A.Yu. Yurin

*Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences (ISDCT SB RAS), Irkutsk, Russia*

### Abstract

The use of semantic technologies including ontologies is a widespread practice in modern intelligent system engineering. Spreadsheets are one of the most accessible and common ways of representing and storing information which are characterized by a wide variety and heterogeneity of layouts, styles and content while remaining a valuable source of domain knowledge. The paper proposes to automate the process of ontology engineering based on the analysis and transformation of spreadsheets with an arbitrary layout. For this purpose a new approach is presented that provides the restoration of the semantics of tabular data, conceptualization, and formalization of tabular content in the form of ontology. The main stages of the proposed approach and a description of the software are presented. The developed software was used to solve the practical problem of ontology engineering for diagnosing and assessing the technical condition of petrochemical equipment. Spreadsheets extracted from reports on industrial safety inspection of petrochemical complexes were used as the initial data. Based on the results of approbation, it was concluded that it is advisable to use the proposed approach when prototyping subject ontologies.

**Key words:** *knowledge acquisition, semantic web, ontology engineering, ontology, spreadsheet, table transformation, industrial safety inspection.*

**Citation:** *Vidia AV, Dorodnykh NO, Yurin AYu. An approach to creating freeform spreadsheet ontology [In Russian]. Ontology of designing. 2021; 11(2): 212-226. DOI: 10.18287/2223-9537-2021-11-2-212-226.*

**Acknowledgment:** The reported study was supported by the Council for Grants of the President of Russia (grant No. MK-1647.2020.9).

### List of figures and tables

- Figure 1 - The main stages of the proposed approach
- Figure 2 - Examples of obtained ontology fragments
- Figure 3 - An example of an aggregated ontological model based on obtained ontology fragments
- Table 1 - Sample snippet of the "Wall Thickness for Major Features" source spreadsheet
- Table 2 - Sample snippet of the "Initial data and calculation results" source spreadsheet
- Table 3 - Fragments: a) a spreadsheet in a canonicalized form obtained from a source spreadsheet (see Table 1),  
b) a spreadsheet with recognized named entities
- Table 4 - Fragments: a) a spreadsheet in canonicalized form obtained from a source spreadsheet (see Table 2),  
b) a spreadsheet with recognized named entities

### References

- [1] *Gavrilova TA, Kudryavtsev DV, Muromtsev DI* Knowledge Engineering. Models and methods [In Russian]. SPb.: Lan; 2016. 324 p.
- [2] *Schreiber G, Wielinga BJ, Jansweijer WNH* The KACTUS View on the 'O' Word. Proc. of IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing. 1995: 159-168.
- [3] *Lopez MF, Gomez-Perez A, Sierra JP, Sierra AP* Building a chemical ontology using METHONTOLOGY and the ontology design environment. IEEE Intelligent Systems. 1999; 14(1): 37-46.
- [4] *Staab S, Studer R, Schnurr H-P, Sure Y* Knowledge processes and ontologies. IEEE Intelligent Systems. 2001; 16(1): 26-34.
- [5] *Suárez-Figueroa MC., Gómez-Pérez A, Motta E, Gangemi A* Ontology engineering in a networked world. Springer-Verlag Berlin Heidelberg. 2012.
- [6] *Grau BC, Horrocks I, Motik B, Parsia B, Patel-Schneider P, Sattler U* OWL 2: The next step for OWL. Web Semantics: Science, Services and Agents on the World Wide Web. 2008; 6(4): 309-322.

- [7] Web Data Commons. Source: <http://webdatacommons.org>.
- [8] **Han L, Finin T, Parr C, Sachs J, Joshi A** RDF123: From spreadsheets to RDF. Proc. of the 7th International Semantic Web Conference (ISWC). Lecture Notes in Computer Science, vol. 5318. Springer-Verlag Berlin Heidelberg, 2008: 451-466.
- [9] **Lebo T, Williams G** Converting governmental datasets into Linked Data. Proc. of the 6th International Conference on Semantic Systems. 2010: 1-3.
- [10] **Scharffe F, Atemezing G, Troncy R, Gandon F, Villata S, Bucher B, Hamdi F, Bihanic L, Képéklian G, Cotton F, Euzenat J, Fan Z, Vandebussche PY, Vatant B** Enabling Linked Data publication with the Datalift platform. Proc. of the AAAI workshop on semantic cities. In 26th Conference on Artificial Intelligence, W10: Semantic Cities. 2012: 25-30.
- [11] Spread2RDF. Source: <https://github.com/marcelotto/spread2rdf>.
- [12] **Bowers S, Madin JS, Schildhauer MP** Owlifier: Creating OWL-DL ontologies from simple spreadsheet-based knowledge descriptions. Ecological Informatics. 2010; 5(1): 19-25.
- [13] **Jupp S, Horridge M, Iannone L, Klein J, Owen S, Schanstra J, Wolstencroft K, Stevens R** Populous: A tool for building OWL ontologies from templates. BMC Bioinformatics. 2011; 13: 1-12.
- [14] **Zhang X, Di R, Feng X**. Ontology based data conversion from spreadsheet to OWL. Proc. of the 2012 Seventh China Grid Annual Conference. 2012: 76-79.
- [15] **Tahar K, Schaaf M, Jahn F, Kücherer C, Paech B, Herre H, Winter A** An Approach to support collaborative ontology construction. Studies in health technology and informatics. 2016; 228: 369-373.
- [16] **Langegger A, Woß W** XLWrap - Querying and Integrating Arbitrary Spread-sheets with SPARQL. Proc. of the 8th International Semantic Web Conference (ISWC). Lecture Notes in Computer Science, vol. 5823. Springer-Verlag Berlin Heidelberg, 2009: 359-374.
- [17] **O'Connor MJ, Halaschek-Wiener C, Musen MA** Mapping Master: A flexible approach for mapping spreadsheets to OWL. Proc. of the 9th International Semantic Web Conference (ISWC). Lecture Notes in Computer Science, vol. 6497. Springer-Verlag Berlin Heidelberg, 2010: 194-208.
- [18] **Dimou A, Sande MV, Colpaert P, Verborgh R, Mannens E, Walle RVD** RML: A generic language for integrated RDF mappings of heterogeneous data. Proc. of the Workshop on Linked Data on the Web, Co-located with the 23rd International World Wide Web Conference. 2014: 1-5.
- [19] **Fiorelli M, Lorenzetti T, Paziienza MT, Stellato A, Turbati A** Sheet2RDF: A flexible and dynamic spreadsheet import & lifting framework for RDF. Proc. of the 28th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Lecture Notes in Computer Science, vol. 9101. Springer-Verlag Berlin Heidelberg, 2015: 131-140.
- [20] **Bischof S, Decker S, Krennwallner T, Lopes N, Polleres A** Mapping between RDF and XML with XSPARQL. Journal on Data Semantics. 2012; 1(3): 147-185.
- [21] **Lefrançois M, Zimmermann A, Bakerally N** A SPARQL extension for generating RDF from heterogeneous formats. Proc. of the 14th International Conference, European Semantic Web Conference (ESWC). Lecture Notes in Computer Science, vol. 10249. Springer-Verlag Berlin Heidelberg, 2017: 35-50.
- [22] Tarql: SPARQL for Tables. Source: <http://tarql.github.io/>.
- [23] Generating RDF from Tabular Data on the Web. Source: <https://www.w3.org/TR/csv2rdf/>.
- [24] W3C ConverterToRdf. Source: <https://www.w3.org/wiki/ConverterToRdf>.
- [25] **Shigarov AO, Mikhailov AA** Rule-based spreadsheet data transformation from arbitrary to relational tables. Information Systems. 2017; 71: 123-136.
- [26] **Tijerino YA, Embley DW, Lonsdale DW, Ding Y, Nagy G** Towards ontology generation from tables. World Wide Web: Internet and Web Information Systems. 2005; 8(8): 261-285.
- [27] **Shigarov AO, Khristyuk VV, Mikhailov AA** TabbyXL: Software platform for rule-based spreadsheet data extraction and transformation. SoftwareX. 2019; 10: 100270.
- [28] **Dorodnykh NO, Yurin AY, Shigarov AO** Conceptual model engineering for industrial safety inspection based on spreadsheet data analysis. Proc. of the 6th International Conference on Modelling and Development of Intelligent Systems (MDIS 2019). Communications in Computer and Information Science, vol. 1126. Springer-Verlag Berlin Heidelberg, 2020: 51-65.
- [29] Stanford Named Entity Recognizer. Source: <https://nlp.stanford.edu/software/CRF-NER.html>.
- [30] **Dorodnykh NO, Vidia AV, Yurin AY** Developing ontology schemas based on spreadsheet transformation. Software & Systems. 2021; 1: 124-131.
- [31] **Yurin AY, Dorodnykh NO** Personal knowledge base designer: Software for expert systems prototyping. SoftwareX. 2020; 11: 100411.
- [32] **Berman AF, Kuznetsov KA, Nikolaychuk OA, Pavlov AI, Yurin AY** Information and analytical support for the examination of industrial safety of chemical, petrochemical and oil refining facilities. Chemical and Petroleum Engineering. 2018; 8: 30-36.

- [33] **Grishchenko MA, Dorodnykh NO, Korshunov SA, Yurin AY** Ontology-based development of diagnostic intelligent systems [In Russian]. *Ontology of Designing*. 2018; 2(28): 265-284. DOI: 10.18287/2223-9537-2018-8-2-265-284.
- 

### About the authors

**Anastasia Vladimirovna Vidia** (b. 1996) a student of the Institute of information technologies and data analysis at the Irkutsk National Research Technical University (INRTU) in the direction of Information systems and technologies. She is a Programmer of Matrosov Institute for System Dynamics and Control Theory of SB RAS (ISDCT SB RAS). *Vidiya\_av@mail.ru*.

**Nikita Olegovich Dorodnykh** (b. 1990) graduated from INRTU in 2012, PhD (2018). He is a senior associate researcher at ISDCT SB RAS. Co-author of about 70 publications in the field of computer-aided development of intelligent systems and knowledge bases, knowledge acquisition based on the transformation of conceptual models and tables. ORCID: 0000-0001-7794-4462; Author ID (RSCI): 979843; Author ID (Scopus): 57202323578; Researcher ID (WoS): E-8870-2014. *tualatin32@mail.ru*.

**Alexander Yurievich Yurin** (b.1980) graduated from the INRTU in 2002, PhD (2005). He is the Head of the "Information and telecommunication technologies for investigation of natural and technogenic safety" laboratory at ISDCT SB RAS and associate professor of the Institute of information technologies and data analysis of INRTU. He is a member of the Russian Association of Artificial Intelligence (RAAI) and Association for Computing Machinery (ACM). He is a member of the Editorial Board of the international scientific journal "Computer, Communication & Collaboration". The list of scientific works includes more than 100 scientific papers in the field of development of decision support systems, expert systems and knowledge bases, application of the case-based reasoning and semantic technologies in the design of diagnostic intelligent systems, maintenance of reliability and safety of complex technical systems. ORCID: 0000-0001-9089-5730; Author ID (RSCI): 174845; Author ID (Scopus): 16311168300; Researcher ID (WoS): A-4355-2014. *iskander@icc.ru*.

---

*Received May 15, 2021. Revised June 16, 2021. Accepted June 21, 2021.*

---

## МЕТОДЫ И ТЕХНОЛОГИИ ПРИНЯТИЯ РЕШЕНИЙ

УДК 519.711.3

DOI:10.18287/2223-9537-2021-11-2-227-238

## Оптимизация при ограничении числа проектных переменных

В.П. Офицеров<sup>1</sup>, С.В. Смирнов<sup>2</sup><sup>1</sup> *Московский авиационный институт (национальный исследовательский университет), Москва, Россия*<sup>2</sup> *Самарский федеральный исследовательский центр РАН,  
Институт проблем управления сложными системами РАН, Самара, Россия*

## Аннотация

Изложен новый подход к постановке и решению оптимизационных задач линейного и нелинейного типа. От классической задачи линейного программирования оптимального распределения ограниченных ресурсов между заданными процессами, рассматриваемая постановка задачи отличается необходимостью выбора ограниченного числа процессов из некоторого конечного множества и распределения ресурсов по этим процессам. Целью является получение оптимального значения целевой функции по отношению к другим вариантам выбора числа процессов из этого же множества и распределению ресурсов между ними. Целевая функция может быть как линейной, так и нелинейной. Нелинейная функция должна обладать определёнными свойствами для корректной работы предложенного алгоритма поиска оптимального решения. Описываемый метод основан на развитии идей динамического программирования Беллмана. Приводятся доказательства оптимальности получаемых решений. В статье даётся оценка вычислительной сложности алгоритма и сравнение с классическими методами решения рассматриваемых задач. Охарактеризованы типы прикладных задач, решаемых с использованием предложенного метода. Компьютерные реализации описанного алгоритма могут использоваться в автоматизированных системах поддержки принятия решений.

**Ключевые слова:** оптимизация, динамическое программирование, линейное и нелинейное программирование, поддержка принятия решений.

**Цитирование:** Офицеров, В.П. Оптимизация при ограничении числа проектных переменных / В.П. Офицеров, С.В. Смирнов // Онтология проектирования. – 2021. – Т.11, №2(40). – С.227-238. – DOI:10.18287/2223-9537-2021-11-2-227-238.

## Введение

Для производственных, транспортных, вычислительных и других систем важно эффективное функционирование подсистем планирования и принятия решений. Обычно на вход подобных подсистем поступают сведения о множестве ресурсов, таких как люди, материалы, оборудование, устройства, капитал, время, и множестве операций (поручений, требований), которые с помощью данных ресурсов должны быть выполнены; их выходом служит распределение ресурсов по операциям. Под улучшением характеристик рассматриваемых подсистем понимают повышение эффективности использования ресурсов, а именно уменьшение потребности в ресурсах без соответствующего изменения в объёме, стоимости и прибыли или увеличение прибыли при сохранении прежней потребности в ресурсах [1-3].

В статье представлен алгоритм решения одного класса задач комбинаторной оптимизации, возникающих при решении дискретных задач линейного и нелинейного программирования с концептуально важным дополнительным ограничением на число используемых в решении переменных [4-10]. Алгоритм является глубокой модификацией метода динамиче-

ского программирования [11, 12]. В отличие от классического метода определения функции Беллмана, в предложенном алгоритме на первом этапе находятся экстремальные значения целевой функции при использовании только одной переменной, определяемой в процессе поиска экстремума на каждом допустимом значении ресурсов. На последующих этапах используется рекурсия поиска экстремальных значений целевой функции путём пошагового увеличения числа переменных, используемых в решении.

Большинство задач комбинаторной оптимизации являются NP-трудными [8, 13, 14], что обуславливает сложность построения эффективных алгоритмов их решения. Для многих задач хорошо известные классические методы решения (динамическое программирование, методы ветвей и границ, «жадные» алгоритмы) оказываются неэффективными, что обуславливает необходимость разработки новых методов.

## 1 Постановка и решение задачи линейного программирования

Большое число задач можно сформулировать в терминах линейного программирования [4-6, 15]. Многие из них приобретают новое прикладное значение, если рассматривать задачу отыскания экстремума линейной функции  $n$  переменных при использовании в решении  $d^*$  переменных, где  $d^* \leq n$ . Математически задача формулируется так: найти

$$\max (c_1x_1 + \dots + c_nx_n) \quad (1)$$

при ограничениях

$$\begin{aligned} a_{11}x_1 + \dots + a_{1n}x_n &\leq b^*_1 \\ &\dots \\ a_{m1}x_1 + \dots + a_{mn}x_n &\leq b^*_m \\ a_{ji} > 0, i = 1, \dots, n, j = 1, \dots, m, \end{aligned} \quad (2)$$

$$\sum_{i=1}^n \tilde{x}_i \leq d^*, \quad (3)$$

где

$$x_i \geq 0, d^* \leq n, \tilde{x}_i = \begin{cases} 1, x_i > 0, \\ 0, x_i = 0. \end{cases}$$

От традиционной постановки задачи линейного программирования [4, 5] рассматриваемую задачу отличает наличие ограничений (3), которые определяют задачу выбора не более  $d^*$  переменных (процессов), между которыми необходимо разделить ограниченные ресурсы, определяемые ограничениями (2) при условии нахождения максимума целевой функции (1).

Одним из примеров такого класса задач является выбор  $d^*$  типов изделий для производства. Пусть для создания некоторого типа изделий требуется  $m$  видов ресурсов и по условиям производства можно выпускать  $d^*$  типов изделий, но при этом есть возможность выбора типов из  $n$  возможных ( $n \geq d^*$ ). Расход ресурса номера  $j$  при единичном выпуске изделия типа  $i$  обозначен через  $a_{ji}$ . Через  $c_i$  обозначен доход от продажи единицы  $i$ -го типа продукции, а через  $x_i$  – количество продукции  $i$ -го типа. Зависимость расходов и доходов от количества выпускаемых изделий  $i$ -го типа принята линейной, а производству доступно  $b_j$  единиц  $j$ -го ресурса. В этом случае задача выбора оптимальной номенклатуры типов изделий и объёмов их выпуска имеет вид (1)–(3), если требуется максимизировать доход.

При  $d^* = n$  задача сводится к обычной задаче линейного программирования [4-6].

Требование целочисленности решения не несёт принципиальных изменений предлагаемого подхода, считается, что переменные  $x_i$  могут принимать только целые значения, когда это необходимо.

Используя приём динамического программирования [13, 14], рассматривается множество задач типа (1)–(3) при всевозможных значениях  $b_j$  и  $d$  ( $0 \leq b_j \leq b^*_j, j = 1, \dots, m; 1 \leq d \leq d^*$ ).

Тогда при  $d = 1$  решается задача отыскания максимума (1) по одной переменной при всех  $0 \leq b_j \leq b_j^*, j = 1, \dots, m$ , что можно записать так:

$$f_1(b_1, \dots, b_m) = \max_{1 \leq i \leq n} c_i x_i \quad (4)$$

при ограничениях для каждого выбираемого значения  $x_i$  и всевозможных значений  $b_j$

$$x_i \geq 0, 0 \leq b_j \leq b_j^*, j = 1, \dots, m. \quad (5)$$

На практике каждое  $b_j$ , как правило, принимает конечное множество значений, образующее некоторую  $\varepsilon_j$ -сеть на  $[0, b_j^*]$ .

Найденные значения максимумов при использовании в решении одной переменной из  $n$  при всевозможных допустимых значениях выделяемых ресурсов  $\Delta b_1, \dots, \Delta b_m$  обозначены

$$f_1(\Delta b_1, \dots, \Delta b_m), \quad (6)$$

где  $0 \leq \Delta b_j \leq b_j, 0 \leq b_j \leq b_j^*, j = 1, \dots, m$ . Здесь  $\Delta b_j, j = 1, \dots, m$  – это всевозможные значения ресурсов, на которых найдены максимумы в соответствии с (4) и (5). Необходимо заметить, что для каждого набора  $\Delta b_j, j = 1, \dots, m$  максимальное значение  $f_1(\Delta b_1, \dots, \Delta b_m)$  может давать любая из  $n$  возможных переменных.

Для получения максимума в (1) при использовании в решении не более двух переменных, т.е. при  $d = 2$ , необходимо найти максимальные значения от одной переменной для всех возможных допустимых значений ресурса и просуммировать их с максимальными значениями целевой функции также от одной переменной, найденных на остающихся после выделения первой переменной ресурсах. Доказательство этого утверждения может быть получено следующим образом. В рамках динамического подхода рекуррентное выражение для поиска максимума при использовании двух переменных запишется так:

$$f_2(b_1, \dots, b_m) = \max [c_i x_i + f_1(b_1 - \Delta b_1, \dots, b_m - \Delta b_m)], 1 \leq i \leq n, \quad (7)$$

при следующих ограничениях для каждого возможного значения  $x_i$  и значений  $b_j$ :

$$a_{ji} x_i \leq \Delta b_j, 0 \leq \Delta b_j \leq b_j, j = 1, \dots, m. \quad (8)$$

В (7)  $f_1(b_1 - \Delta b_1, \dots, b_m - \Delta b_m)$  – найденное на первом шаге алгоритма максимальное значение функции от одной переменной на конкретном значении выделенных ресурсов, которое складывается со значением  $c_i x_i$ , где  $i$  выбирается из множества  $\{1, \dots, n\}$  и определяется также на конкретном (в момент выбора) наборе значений ресурсов  $\Delta b_j, j = 1, \dots, m$ .

Для получения максимума двух слагаемых, одно из которых имеет фиксированное значение при выбранных на текущий момент значениях  $\Delta b_j, j = 1, \dots, m$ , нужно выбрать переменную с номером  $i$ , которая даст максимальное значение  $c_i x_i$  на выделенном ресурсе  $\Delta b_j, j = 1, \dots, m$ . На первом шаге алгоритма такое значение уже было определено и записано в (6). Вариант с возможным совпадением номеров переменных даёт значения  $f_1(b_1 - \Delta b_1, \dots, b_m - \Delta b_m)$  и  $f_1(\Delta b_1, \dots, \Delta b_m)$ . Следовательно, выражения (7) и (8) можно записать так:

$$f_2(b_1, \dots, b_m) = \max [f_1(\Delta b_1, \dots, \Delta b_m) + f_1(b_1 - \Delta b_1, \dots, b_m - \Delta b_m)], \quad (9)$$

при выполнении ограничений:

$$0 \leq \Delta b_j \leq b_j, 0 \leq b_j \leq b_j^*, j = 1, \dots, m.$$

Максимальное значение целевой функции не более чем от двух переменных на ресурсах  $b_1, \dots, b_m$  получено перебором сумм максимальных значений двух функций, каждая из которых использует только одну из  $n$  возможных переменных, определённых на первом шаге алгоритма для всевозможных значений ресурсов.

Если вернуться к исходным обозначениям, выражение (9) после вычисления  $f_2(b_1, \dots, b_m)$  можно записать так:

$$f_2(b_1, \dots, b_m) = \max (c_k x_k + c_l x_l),$$

где максимальное значение суммы получено после выполнения полного перебора вариантов возможных распределений ресурсов между двумя переменными. При этом на каждом варианте распределения ресурса выбирались две переменные, дающие максимальное значение целевой функции на выделенных им ресурсах. Найденные номера переменных, их значения и соответствующие им ресурсы обозначены так: номера -  $k$  и  $l$ , значения переменных -  $x_k$  и  $x_l$ , значение целевой функции -  $c_k x_k + c_l x_l$ , ограничения -  $a_{jk} x_k \leq \Delta b_j$ ,  $a_{jl} x_l \leq b_j - \Delta b_j$ ,  $j = 1, \dots, m$ .

Можно предположить, что существуют переменные с номерами  $y$  и  $z$  со значениями  $x_y$  и  $x_z$ , которые на тех же выделенных ресурсах дают значение больше значения  $f_2(b_1, \dots, b_m)$ . Это означает, что либо на ресурсе  $\Delta b_j$ , либо на ресурсе  $b_j - \Delta b_j$  ( $j = 1, \dots, m$ ) выполняется хотя бы одно из неравенств

$$\begin{aligned} c_y x_y &> c_k x_k, \\ c_z x_z &> c_l x_l, \end{aligned}$$

что невозможно в силу определения значений  $c_k x_k$  и  $c_l x_l$  как максимальных на этих ресурсах.

Аналогичные рассуждения можно провести для любого возможного распределения ресурсов между двумя любыми переменными. Это означает, что если в исходной постановке задачи не добавляется условие (3) выбора переменных, участвующих в решении задачи, т.е. если предложенным методом решается классическая задача линейного программирования с двумя переменными в целевой функции, то будет найдено глобальное оптимальное решение.

Таким образом, доказано, что вычисленное значение  $f_2(b_1, \dots, b_m)$  является глобальным максимумом на всех допустимых значениях  $b_1, \dots, b_m$ .

Продолжая процесс, можно получить рекуррентное уравнение относительно числа используемых в решении переменных:

$$f_d(b_1, \dots, b_m) = \max [f_1(\Delta b_1, \dots, \Delta b_m) + f_{d-1}(b_1 - \Delta b_1, \dots, b_m - \Delta b_m)], \quad (10)$$

для всех  $0 \leq \Delta b_j \leq b_j$ ,  $0 \leq b_j \leq b_j^*$ ,  $j = 1, \dots, m$ ,  $d = 2, \dots, d^*$ .

Используя метод математической индукции и свойства линейных операций, можно доказать, что для линейной целевой функции и линейных ограничений значение  $f_d(b_1, \dots, b_m)$  является глобальным максимумом для всех допустимых  $b_1, \dots, b_m$ .

При использовании в решении только одной переменной ( $d = 1$ ) значения  $f_1(b_1, \dots, b_m)$  являются максимальными для допустимых значений  $b_1, \dots, b_m$  (в том числе и для значений  $\Delta b_1, \dots, \Delta b_m$ , используемых в рекурсивном выражении (10)). Вычисленное значение  $f_2(b_1, \dots, b_m)$  является глобальным максимумом на всех допустимых значениях  $b_1, \dots, b_m$ . Можно предположить, что выражение (10) даёт максимальное значение при использовании в решении не более  $(d - 1)$  переменных. Оно даёт максимальное значение и при использовании в решении не более  $d$  переменных. Действительно, в выражении (10) значения  $f_{d-1}$  по допущению являются максимальными на всех допустимых значениях ресурсов, а значения  $f_1(\Delta b_1, \dots, \Delta b_m)$  являются максимальными в силу алгоритма их вычисления. Пусть значение максимума  $f_d(b_1, \dots, b_m)$  получено суммированием

$$S_1(\Delta b_1, \dots, \Delta b_m) + f_{d-1}(b_1 - \Delta b_1, \dots, b_m - \Delta b_m),$$

где  $S_1(\Delta b_1, \dots, \Delta b_m)$  - некоторое значение целевой функции при использовании в решении какой-то одной переменной из  $n$  возможных. По определению значение  $f_1(\Delta b_1, \dots, \Delta b_m)$  является максимальным на ресурсе  $\Delta b_1, \dots, \Delta b_m$ . Это означает, что  $S_1(\Delta b_1, \dots, \Delta b_m) = f_1(\Delta b_1, \dots, \Delta b_m)$ , иначе значение максимума  $f_d(b_1, \dots, b_m)$  не было бы получено. Таким образом, доказано, что вычисленное значение  $f_d(b_1, \dots, b_m)$  является глобальным максимумом на всех допустимых значениях  $b_1, \dots, b_m$ .

Если на любом шаге  $d$  значение  $f_d(b_1, \dots, b_m)$  получается при повторном использовании хотя бы одной из  $n$  переменных (т.е., некоторая переменная с номером  $i$  входит в оба слагаемых соотношения (10)), то

$$f_d(b_1, \dots, b_m) = f_{d-1}(b_1, \dots, b_m). \quad (11)$$

Действительно, пусть

$$f_d(b_1, \dots, b_m) = \max [f_1^d(\Delta b_1^d, \dots, \Delta b_m^d) + f_{d-1}(b_1 - \Delta b_1, \dots, b_m - \Delta b_m)], \quad 0 \leq \Delta b_j \leq b_j,$$

а число использованных переменных для получения  $f_d$  не увеличилось. Это означает, что некоторая переменная  $x_i$ , которой соответствует  $f_1^d(\Delta b_1^d, \dots, \Delta b_m^d)$ , была использована на одном из предыдущих шагов при получении  $f_{d-1}(b_1 - \Delta b_1, \dots, b_m - \Delta b_m)$ . Согласно (4)-(10), значение целевой функции формируется таким образом, что будет справедливо выражение

$$f_d(b_1, \dots, b_m) = f_1^d(\Delta b_1^d, \dots, \Delta b_m^d) + f_1^k(\Delta b_1^k, \dots, \Delta b_m^k) + f_{d-2}(b_1 - \Delta b_1^d - \Delta b_1^k, \dots, b_m - \Delta b_m^d - \Delta b_m^k),$$

где  $f_1^d(\Delta b_1^d, \dots, \Delta b_m^d)$  и  $f_1^k(\Delta b_1^k, \dots, \Delta b_m^k)$  по предположению, получены при использовании одной и той же переменной  $i$  со значениями  $x_i^d$  и  $x_i^k$ .

В силу линейности функций и ограничений можно показать, что для  $x_i^d$  и  $x_i^k$  справедливо равенство

$$f_1^d(\Delta b_1^d, \dots, \Delta b_m^d) + f_1^k(\Delta b_1^k, \dots, \Delta b_m^k) = f_1(\Delta b_1^d + \Delta b_1^k, \dots, \Delta b_m^d + \Delta b_m^k).$$

Действительно, это соответствует свойству линейных операций  $c_i x_i^d + c_i x_i^k = c_i x_i$ , где  $x_i = x_i^d + x_i^k$ , и преобразованию ограничений для этой переменной

$$a_{ji} x_i^d \leq \Delta b_j^d, \quad a_{ji} x_i^k \leq \Delta b_j^k, \quad j = 1, \dots, m,$$

в ограничения

$$a_{ji} x_i^d + a_{ji} x_i^k \leq \Delta b_j^d + \Delta b_j^k, \quad j = 1, \dots, m,$$

или, что то же самое, в ограничения

$$a_{ji} x_i \leq \Delta b_j^d + \Delta b_j^k, \quad j = 1, \dots, m,$$

где значения  $\Delta b_j^d + \Delta b_j^k$  для переменной с номером  $i$  рассмотрены на первом шаге алгоритма, и соответствующее значение  $c_i x_i$  использовано для формирования значения  $f_{d-1}(b_1, \dots, b_m)$  на шаге  $(d-1)$ . Отсюда следует, что найденное повторное включение переменной  $i$  в целевую функцию можно записать так

$$\begin{aligned} c_i(x_i^d + c_i x_i^k) + f_{d-2}(b_1 - \Delta b_1^d - \Delta b_1^k, \dots, b_m - \Delta b_m^d - \Delta b_m^k) = \\ = c_i x_i + f_{d-2}(b_1 - \Delta b_1^i, \dots, b_m - \Delta b_m^i) = f_{d-1}(b_1, \dots, b_m). \end{aligned}$$

Здесь  $\Delta b_1^i = \Delta b_1^d + \Delta b_1^k, \dots, \Delta b_m^i = \Delta b_m^d + \Delta b_m^k$ ;  $c_i x_i = c_i(x_i^d + x_i^k)$  - линейные ограничения для  $x_i^d$  и  $x_i^k$  складываются и дают ограничение для  $x_i$ .

Таким образом, показано отсутствие нарушения структуры целевой функции при попытке повторного включения в решение уже использованной переменной. Поэтому если на шаге  $d$  значение  $f_d(b_1, \dots, b_m)$  увеличилось, то оно получено при использовании  $d$  переменных, так как в противном случае это увеличение произошло бы на одном из предыдущих шагов (имеется в виду минимальное число используемых переменных).

Для определения фактических номеров, числа и значений переменных, на которых достигается  $f_d(b_1, \dots, b_m)$ , можно использовать соотношение (10), двигаясь по шагам в обратном направлении и определяя:

- на шаге  $d$  - значение  $f_1^d(\Delta b_1^d, \dots, \Delta b_m^d)$ , равное  $f_1(\Delta b_1, \dots, \Delta b_m)$  в (10) при получении  $f_d(b_1, \dots, b_m)$ , и соответствующие номер и значение переменной;
- на шаге  $(d-1)$  - значение  $f_1^{d-1}(\Delta b_1^{d-1}, \dots, \Delta b_m^{d-1})$ , дающее  $f_{d-1}(b_1 - \Delta b_1^d, \dots, b_m - \Delta b_m^d)$  и соответствующие номер и значение переменной.

Продолжая этот пошаговый процесс, можно получить конкретные номера переменных и их значения, определяющие максимальное значение целевой функции.

В классическом представлении функция Беллмана определяется последовательно по пронумерованным переменным [11]. Когда ресурс выделяется переменной  $i$ , то предполагается, что оставшийся ресурс оптимально распределён по уже рассмотренным  $(i-1)$  упорядоченным по номерам переменным. Задача будет решена, когда будут получены результаты

для последней переменной в списке. В результате ограничение на число используемых в решении переменных не будет учтено. Чтобы его учесть при использовании классического рекуррентного уравнения Беллмана [13] придётся решить  $C_n^{d^*}$  задач, последовательно перебирая варианты наборов  $d^*$  переменных из  $n$  возможных, и выбрать из этих решений наилучшее. Такой же подход (с перебором вариантов) можно использовать и при применении других традиционных методов решения задач линейного программирования с дополнительным ограничением на число используемых переменных.

Для иллюстрации работы предложенного метода рассмотрена задача:

$$\text{найти } \mathbf{max} (x_1 + 4x_2 + 6x_3 + 9x_4)$$

при ограничениях

$$\begin{aligned} x_1 + 3x_2 + 5x_3 + 7x_4 &\leq 10, \\ x_i &\in \{0, 1, 2, \dots\}, i = 1, \dots, 4; \sum_{i=1}^4 \tilde{x}_i \leq 2, \tilde{x}_i = \begin{cases} 1, & x_i > 0, \\ 0, & x_i = 0. \end{cases} \end{aligned}$$

Данные, получаемые в результате вычислений по (4)-(10) для возможных значений  $b_1 = 1, 2, \dots, 10$ , объединены в таблице 1. В примере используется только один тип ресурса, который обозначен как  $b_1$ .

На первом шаге получены значения

$$f_1(b_1) = \mathbf{max}_{1 \leq i \leq 4} c_i x_i$$

для всех  $b_1$  при выполнении ограничений для каждого возможного значения ресурса. Результаты вычислений  $f_1$  показаны во второй строке таблицы. Каждый результат в отдельном столбце соответствует:

- значению ограничения на ресурс, показанному в первой строке таблицы;
- номеру  $i$  выбранной по алгоритму вычисления  $f_1(b_1)$  переменной, показанному в третьей строке;
- значению переменной  $x_i$ , вычисленному при условии выполнения ограничения на ресурс, показанному в четвёртой строке.

На втором шаге алгоритма для каждого возможного значения ресурса вычисляется значение целевой функции от двух переменных по формуле (10), при этом используются результаты вычислений первого шага:

$$f_2(b_1) = \mathbf{max} [f_1(\Delta b_1) + f_1(b_1 - \Delta b_1)], 0 \leq \Delta b_1 \leq b_1,$$

при  $a_i x_i \leq \Delta b_1; i = 1, 2, \dots, 4; b_1 \leq b_1^*, b_1^* = 10; b_1 \in \{1, 2, \dots, 10\}$ .

Для ресурса, принимающего значения от 1 до 3, значения целевой функции получаются при использовании в решении только одной переменной и совпадают с результатами первого шага. Эти результаты показаны в пятой строке таблицы 1, обозначенной как « $f_2 = f_1^2 + f_1^1$ ».

При значении ресурса  $b_1 = 4$  после просмотра возможных значений  $f_1(\Delta b_1)$  при  $\Delta b_1 \in \{1, 2, 3, 4\}$  получено максимальное значение  $f_2(4) = f_1(1) + f_1(4 - 1) = 1 + 4 = 5$  при  $x_1 = 1, x_2 = 1$ . Эти результаты показаны в строках 5-8 таблицы 1 и столбце со значением  $b_1 = 4$ . В строке, обозначенной « $f_1^2 + f_1^1$ », показаны значения  $f_1(\Delta b_1) + f_1(b_1 - \Delta b_1)$ , на которых получен максимум; в строке, обозначенной « $i, j$ », показаны номера переменных, формирующих оптимальное значение, а в строке 8, обозначенной как « $x_i, x_j$ », значения соответствующих переменных.

При значении ресурса  $b_1 = 5$  на втором шаге получено оптимальное значение при двух вариантах использования ресурса. Первым получается значение целевой функции равное 6 при  $f_1(\Delta b_1 = 0) = 0$  и  $f_1(5 - 0) = 6$  при  $x_3 = 1$ . Этот же результат получается при  $f_1(\Delta b_1 = 2) = 2$  и  $f_1(5 - 2) = 4$  при  $x_1 = 2, x_2 = 1$ :

$$f_2(5) = f_1(0) + f_1(5 - 0) = 6 = f_1(2) + f_1(5 - 2) = 2 + 4 = 6.$$

Так как в задаче не требуется обязательного использования двух переменных в решении, то в таблице записан первый результат. Продолжая процесс для значения ресурса  $b_1 = 10$ , значение целевой функции получается равное 13, когда  $f_2(10) = f_1(1) + f_1(9) = 13$  при  $x_1 = 1, x_2 = 3$ .

В данном примере вычисление значений  $f_2(b_1)$  для всех возможных значений  $b_1 \in \{1, 2, \dots, 10\}$  было избыточным. Достаточно было сразу рассмотреть вариант с  $b_1 = 10$ , так как третьего шага в случае поиска оптимального решения при использовании трёх переменных нет. Если бы в задаче было условие использования в решении не менее трёх переменных из четырёх возможных, то в рекуррентных выражениях для поиска  $f_3(b_1)$  следовало бы использовать  $f_2(b_1)$  для всех возможных значений  $b_1$ , хотя  $f_3(b_1)$  опять можно было вычислять только для  $b_1 = 10$  и  $\Delta b_1 \in \{1, 2, \dots, 10\}$ . Такое правило можно использовать и в общем случае на последнем шаге алгоритма.

Из таблицы 1 видно, что при использовании в решении одной переменной максимум целевой функции равен 12 при  $x_2 = 3$ , а максимум при использовании в решении двух переменных равен 13 при  $x_1 = 1, x_2 = 3$ .

Таблица 1 – Данные, получаемые при решении примера:  
 возможные значения ресурса  $b_1$ ,  
 соответствующие им значения целевой функции от одной используемой переменной  $f_1$ ,  
 значения номеров используемой переменной  $i$ , значения переменной  $x_i$ ,  
 значения целевой функции от не более двух используемых переменных  $f_2 = f_1^2 + f_1^1$ ,  
 значения  $f_1^2 + f_1^1$  при получении  $f_2$ ,  
 $i, j$  - значения номеров используемых переменных  $i, j$  для, соответственно,  $f_1^2$  и  $f_1^1$ ,  
 значения переменных  $x_i, x_j$  в  $f_2 = f_1^2 + f_1^1$

1	$b_1$	1	2	3	4	5	6	7	8	9	10
2	$f_1$	1	2	4	4	6	8	9	9	12	12
3	$i$	1	1	2	2	3	2	4	4	2	2
4	$x_i$	1	2	1	1	1	2	1	1	3	3
5	$f_2 = f_1^2 + f_1^1$	1	2	4	5	6	8	9	10	12	13
6	$f_1^2 + f_1^1$	0+1 1	0+2 2	0+4 4	1+4 4	0+6 6	0+8 8	0+9 9	1+9 9	0+12 12	1+12
7	$i, j$		1		1		2		1		2
8	$x_i, x_j$	0	1	0	2	0	1	1	0	1	3

При использовании рекуррентного соотношения (10) для определения последовательности  $\{f_d(b_1, \dots, b_m)\}$  в общем случае считается, что  $b_j$  может принимать значения  $0, \Delta_j, 2\Delta_j, \dots, K_j\Delta_j = b_j^*, j = 1, \dots, m$ . Тогда в памяти вычислительной среды необходимо хранить минимум  $\prod_{j=1}^m (K_j + 1)$  значений  $f_{d-1}(b_1, \dots, b_m)$  для последующего получения  $f_d(b_1, \dots, b_m)$ .

Объём вычислений в общем случае можно оценить величиной  $O(n + d^* \times K^m)$ , где  $n$  - общее число переменных,  $d^*$  - ограничение на число переменных в решении,  $K$ - максимальное количество точек дискретизации среди всех ограничений,  $m$  - количество ограничений. Вычисления просты и используют результаты предыдущих шагов. При использовании для решения классического рекуррентного уравнения Беллмана объём вычислений оценивается как  $O(C_n^{d^*} \times d^* \times K^m)$ . При  $d^* = n$  получаются оценки  $O(n + n \times K^m)$  и  $O(n \times K^m)$  соответственно. Таким образом, вычисления на основе уравнения Беллмана [11, 12] немного эффективнее, если в решении можно использовать все описанные в задаче переменные.

При решении задачи с большим числом ограничений  $m$  и больших числах  $K_j$  может потребоваться большой объём памяти, либо нахождение решения может быть достаточно долгим. В то же время число переменных  $n$  и  $d^*$  мало влияют на эффективность рассматриваемого алгоритма, поэтому возможности метода весьма широки при решении дискретных задач линейного программирования с большим числом переменных и с малым числом ограничений. Кроме того, количество точек дискретизации во многих задачах может выбираться кратно наибольшему общему делителю коэффициентов в каждом конкретном ограничении  $j$ , что может существенно уменьшить величины  $K_j, j = 1, \dots, m$ .

## 2 Нелинейный целевой функционал

Пусть вместо (1) рассматривается задача вида: найти

$$\mathbf{max} [c_1(x_1) + \dots + c_n(x_n)] \quad (12)$$

при ограничениях (2–3).

Используя идею динамического программирования, можно рассматривать множество задач типа (1)–(3), при всевозможных значениях  $b_j$  и  $d$  ( $0 \leq b_j \leq b_j^*, j = 1, \dots, m; 1 \leq d \leq d^*$ ), но при этом имея ввиду, что функции  $c_i(x_i)$  нелинейные.

При  $d = 1$  решается задача отыскания максимума (12) при использовании одной из  $n$  переменных при всех  $0 \leq b_j \leq b_j^*, j = 1, \dots, m$ :

$$f_1(b_1, \dots, b_m) = \mathbf{max}_{1 \leq i \leq n} c_i(x_i) \quad (13)$$

при ограничениях (2)–(3).

Далее используется рекуррентное соотношение (10).

По определению значение  $f_1(b_1, \dots, b_m)$  является максимумом целевой функции при распределении ресурсов  $b_1, \dots, b_m$  для одной из  $n$  переменных.

Пусть хотя бы одна функция  $c_i(x_i)$  удовлетворяет следующему условию для значений аргумента  $x_i$  равных  $x$  и  $y$ :

$$c_i(x) + c_i(y) > c_i(x + y). \quad (14)$$

Тогда при поиске максимума по соотношениям (13) и (10) может возникнуть ситуация, когда вычисленное по (10) значение  $f_d(b_1, \dots, b_m)$  будет получено за счёт «постепенной» выдачи ресурсов  $i$ -му процессу. Однако, при выполнении условия (14) фактическое значение  $f_d(b_1, \dots, b_m)$  не будет максимальным значением на выделенных ресурсах. Чтобы избежать этого, следует ограничиться задачами, в которых для всех  $i \in \{1, \dots, n\}$  выполняется условие:

$$c_i(x) + c_i(y) \leq c_i(x + y). \quad (15)$$

При  $d = 1$  значение  $f_d(b_1, \dots, b_m)$  максимально по определению. Пусть при условии (15) выражение (10) даёт максимальное значение функционала при  $d = k$ . Можно показать, что при  $d = k + 1$  соотношение (10) даст фактическое максимальное значение для всех  $0 \leq b_j \leq b_j^*, j = 1, \dots, m$ .

Линейность ограничений и условие (15) исключают «повторное» распределение ресурса  $i$ -му процессу. Следовательно, в (10) при  $d = k + 1$  достигается искомый максимум, а из метода математической индукции следует, что максимум достигается при любом  $d \leq n$ .

В качестве примера рассмотрен поиск максимума целевой функции вида

$$\begin{aligned} & (x_1)^2 + 2(x_2)^4 + 3(x_3)^2 \\ & \text{при } x_1 + 2x_2 + 4x_3 \leq 5, \\ & x_i \in \{0, 1, 2, \dots\}, i = 1, 2, 3; \sum_{i=1}^3 \check{x}_i \leq 2, \check{x}_i = \begin{cases} 1, x_i > 0, \\ 0, x_i = 0. \end{cases} \end{aligned}$$

Следуя (13) и (10), вычисляются  $f_1(1) = 1, f_1(2) = 4, f_1(3) = 9, f_1(4) = 32, f_1(5) = 32$ .

Далее, так как условие (15) выполняется для всех  $i$ , получаются согласно (10)  $f_2(1) = 1, f_2(2) = 4, f_2(3) = 9, f_2(4) = 32, f_2(5) = 33$ .

Важно отметить, что, например, все функции вида

$$a_0 + a_1x + a_2(x_2)^2 + \dots + a_n(x_n)^n,$$

где  $a_i \geq 0, x \geq 0, n > 0$  удовлетворяют условию (15), т.е. класс задач, к которым применим разработанный метод, достаточно широк.

## Заключение

В статье изложен новый подход к постановке и решению оптимизационных задач линейного и нелинейного типов. От классической задачи линейного программирования оптимального распределения ограниченных ресурсов между заданными процессами рассматриваемая постановка задачи отличается необходимостью выбора ограниченного числа процессов из некоторого конечного множества. Ресурсы распределяются по этим процессам так, что достигается оптимальное значение целевой функции по отношению к другим вариантам состава проектных переменных с таким же числом процессов из этого же множества. Предложенный метод решения основан на развитии идей динамического программирования. Приведены доказательства оптимальности получаемых решений.

Обоснована оценка вычислительной сложности решения поставленных задач. Приведено сравнение описываемого подхода с классическими методами решения рассматриваемых задач. Показано, что классические методы можно применять путём сведения исходной постановки задачи к  $C_n^{d^*}$  задачам без дополнительных ограничений на число используемых в решении процессов, где  $d^*$  - число процессов в решении, выбираемых из  $n$  возможных. После решения  $C_n^{d^*}$  задач необходимо выбрать из них оптимальное.

Рассмотренные в статье математические постановки задач и их решения имеют ряд актуальных практических приложений. Например, можно сформулировать задачу для оптимального размещения ограниченной номенклатуры грузов на ограниченных складских площадях [2], или задачу выбора ограниченного числа типов ракет-носителей из возможного набора типов для оптимального выполнения программ космических исследований [16] и другие подобные [17].

При решении практических задач приходится сталкиваться с многокритериальной оптимизацией. Если многокритериальный функционал качества каким-либо способом приводится к одному критерию [18-21], то затем можно применять предложенный метод решения.

## Список источников

- [1] *Singiresu, S.R.* Engineering optimization: theory and practice / S.R. Singiresu - New York: Wiley, 2009. - 813 p.
- [2] *Офицеров, В.П.* Моделирование и оптимизация снабженческой деятельности для крупных компаний / В.П. Офицеров, С.В. Смирнов // Проблемы управления и моделирования в сложных системах: Труды V международной конф. (17-21 июня 2003 г., Самара, Россия). – Самара: СамНЦ РАН, 2003. – С. 197–205.
- [3] *Стрекаловский, А.С.* Линейные и квадратично-линейные задачи двухуровневой оптимизации / А.С. Стрекаловский, А.В. Орлов. — Новосибирск: Изд-во СО РАН, 2019. – 261 с.
- [4] *Гасс, С.* Линейное программирование / С. Гасс. — М.: Гос. изд-во физико-математической литературы, 2015. — 304 с.
- [5] *Юдин, Д.Б.* Задачи и методы линейного программирования: Конечные методы / Д.Б. Юдин, Е.Г. Гольштейн. – М.: URSS, 2010. – 264 с.
- [6] *Гераськин, М.И.* Линейное программирование / М.И. Гераськин, Л.С. Клентак. – Самара: Изд-во СГАУ, 2014. – 104 с.

- [7] **Кормен, Т.** Алгоритмы: построение и анализ / Т. Кормен, Ч. Лейзерсон, Р. Ривест, К. Штайн. М.: ООО «И.Д. Вильямс», 2013.- 1328 с.
- [8] **Martelo, S.** Knapsack problems / S. Martelo, P. Toth. — London: Wiley, 1990. — 306 p.
- [9] **Kellerer, H.** Knapsack Problems / H. Kellerer, U. Pferschy, D. Pisinger — Springer Science+Business Media, 2004. — 548 p.
- [10] **Bretthauer, K.M.** The nonlinear knapsack problem – algorithms and applications / K.M. Bretthauer, B. Shetty // European Journal of Operational Research — 2002. — Vol. 138, Iss. 3. — P. 459–472.
- [11] **Беллман, Р.** Прикладные задачи динамического программирования / Р. Беллман, С. Дрейфус. – М.: Наука, 1965. – 460 с.
- [12] **Bellman, R.E.** Dynamic Programming / R.E. Bellman. - Princeton, NJ: Princeton University Press, 2010. – 392 p.
- [13] **Пападимитриу, Х.** Комбинаторная оптимизация: Алгоритмы и сложность / Х. Пападимитриу, К. Стайглиц. - М.: Мир, 1985.- 510 с.
- [14] **Гэри, М.** Вычислительные машины и трудно решаемые задачи / М. Гэри, Д. Джонсон. — М.: Мир, 1982. – 416 с.
- [15] **Струченков, В.И.** Методы оптимизации в прикладных задачах / В.И. Струченков. – М.: СОЛОН–Пресс, 2009. – 320 с.
- [16] **Офицеров, В.П.** Об оптимальном выполнении программы космических исследований заданным числом типов ракет-носителей / В.П. Офицеров // АН СССР. Космические исследования. – 1980. – Т. 18, № 4. – С. 550–555.
- [17] **Офицеров, В.П.** Об одном подходе к автоматизации и информатизации процесса составления программ обучения / В.П. Офицеров, М.В. Офицеров, О.А. Бочарова // Вестник РУДН. Серия «Информатизация образования». – 2012. – № 4. – С. 105–113.
- [18] **Пиявский, С.А.** Метод универсальных коэффициентов при принятии многокритериальных решений / С.А. Пиявский // Онтология проектирования. – 2018. – Т. 8, №3(29). – С. 449-468. – DOI: 10.18287/2223-9537-2018-8-3-449-468.
- [19] **Пиявский, С.А.** Формулы для вычисления универсальных коэффициентов при принятии многокритериальных решений / С.А. Пиявский // Онтология проектирования. – 2019. – Т. 9, №2(32). – С. 282-298 – DOI: 10.18287/2223-9537-2019-9-2-282-298.
- [20] **Ерасов, И.В.** Об одном кардинальном алгоритме обработки экспертной информации на основе метода анализа иерархий / И.В. Ерасов, В.П. Офицеров // Информатизация образования и науки. – 2013. – № 4 (20). – С. 153–161.
- [21] **Саати, Т.** Принятие решений при зависимостях и обратных связях: Аналитические сети / Т. Саати – М.: Изд-во ЛКИ, 2008. – 360 с.

## Сведения об авторах



**Офицеров Владимир Петрович** 1953 г. рождения. Окончил Куйбышевский авиационный институт им. С.П. Королёва в 1975 г., к.т.н. (1982). Доцент кафедры «Вычислительные машины, системы и сети» Московского авиационного института (национального исследовательского университета). В списке научных трудов более 100 работ в области прикладной математики, информационных технологий, прогнозирования развития сложных социально-технических систем. Author ID (РИНЦ): 618798. [ofitserovvp@mai.ru](mailto:ofitserovvp@mai.ru)

**Смирнов Сергей Викторович**, 1952 г. рождения. Окончил Куйбышевский авиационный институт им. С.П. Королёва в 1975 г., д.т.н. (2002). Главный научный сотрудник Института проблем управления сложными системами Самарского федерального исследовательского центра РАН. Член Российской ассоциации искусственного интеллекта, Международной ассоциации по онтологиям и их приложениям (IAOA). В списке научных трудов более 170 работ в области прикладной математики, компьютерного моделирования сложных систем, создания интеллектуальных систем поддержки принятия решений. ORCID: 0000-0002-3332-5261; Author ID (РИНЦ): 17628; Author ID (Scopus): 57205017287; Researcher ID (WoS): AAB-2013-2021. [smirnov@iccs.ru](mailto:smirnov@iccs.ru)



Поступила в редакцию 15.10.2020, после рецензирования 05.05.21. Принята к публикации 05.06.2021.

## Optimization while limiting the number of design variables

V.P. Ofitserov<sup>1</sup>, S.V. Smirnov<sup>2</sup>

<sup>1</sup> *Moscow aviation institute (National Research University), Moscow, Russia*

<sup>2</sup> *Samara Federal Research Scientific Center RAS, Institute for the Control of Complex Systems RAS, Samara, Russia*

### Abstract

A new approach to the formulation and solution of optimization problems of linear and nonlinear type is stated in this article. The problem statement under consideration differs from the classical linear programming problem of the optimal distribution of limited resources between given processes by the need to choose a limited number of processes from a certain finite set and allocate resources over these processes. The goal is to obtain the optimal value of the objective function in relation to other options for choosing the number of processes from the same set and the distribution of resources between them. The objective function can be either linear or non-linear. A nonlinear function must have certain properties for the correct operation of the proposed algorithm for finding the optimal solution. The described method is based on the development of Bellman's ideas of dynamic programming. The proofs of the optimality of the obtained solutions are provided. The article gives an estimate of the computational complexity of the algorithm and a comparison with classical methods for solving the problems under consideration. The types of applied problems solved using the proposed method are characterized. Computer implementations of the described algorithm can be used in automated decision support systems.

**Keywords:** *optimization, dynamic programming, linear and non-linear programming, decision support.*

**Citation:** *Ofitserov VP, Smirnov SV. Optimization while limiting the number of design variables [In Russian]. Ontology of designing. 2021; 11(2): 227-238. DOI: 10.18287/2223-9537-11-2-227-238.*

### List of tables

Table 1 - The data obtained by solving the example:

- possible values of the resource  $b_1$ ,
- the corresponding values of the objective function from one used variable  $f_1$ ,
- values of numbers of the used variable  $i$ ,
- the values of the variable  $x_i$ ,
- the values of the objective function from no more than two used variables  $f_2 = f_1^2 + f_1^1$ ,
- values  $f_1^2 + f_1^1$  when obtaining  $f_2$ ,
- $i, j$  - value of numbers of used variables  $i$  for, respectively  $f_1^2$  and  $f_1^1$ ,
- the values of the variables  $x_i, x_j$  in  $f_2 = f_1^2 + f_1^1$

### References

- [1] *Singiresu SR*. Engineering optimization: theory and practice. New York: Wiley, 2009. 813 p.
- [2] *Ofitserov VP, Smirnov SV*. Modeling and optimization of supplying activity for the large companies [In Russian]. Proc. of V Int. conf. "Complex systems: Control and Modeling Problems" (June 17-21, 2003, Samara, Russia). Samara: Samara Scientific Center RAS, 2003: 197–205.
- [3] *Strekalovsky AS, Orlov AV*. Linear and quadratic-linear problems of two-level optimization [In Russian]. Novosibirsk: Publishing house Siberian Branch of RAS, 2019: 261 p.
- [4] *Gus S*. Linear programming [In Russian]. Moscow: State publishing house of physical and mathematical literature, 2015: 304 p.
- [5] *Yudin DB, Golstein EG*. Tasks and methods of linear programming: Final methods [In Russian]. Moscow: URSS, 2010: 264 p.
- [6] *Geraskin MI, Klentak LS*. Linear programming [In Russian] Samara: SGAU publishing house, 2014: 104 p.
- [7] *Kormen T, Leyzerson Ch, Rivest R, Stein K*. Algorithms: creation and analysis [In Russian]. Moscow: LLC "I.D. Williams", 2013:- 1328 p.
- [8] *Martelo S, Toth P*. Knapsack problems. London: Wiley, 1990: 306 p.
- [9] *Kellerer H, Pferschy U, Pisinger D*. Knapsack Problems. Springer Science+Business Media, 2004: 548 p.

- [10] **Brethauer KM, Shetty B.** The nonlinear knapsack problem - algorithms and applications. *European Journal of Operational Research*. 2002; 138(3): 459-472.
  - [11] **Bellman R, Dreyfus S.** Applied problems of dynamic programming [In Russian]. Moscow: Publishing house "Nauka", 1965: - 460 p.
  - [12] **Bellman RE.** Dynamic Programming. Princeton, NJ: Princeton University Press, 2010: 392 p.
  - [13] **Papadimitriu H, Stayglits K.** Combinatory optimization: Algorithms and complexity [In Russian]. Moscow: Publishing house "Mir", 1985: 510 p.
  - [14] **Gary M, Johnson D.** Computers and hardly solvable tasks [In Russian]. Moscow: Publ. house "Mir", 1982: 416 p.
  - [15] **Struchenkov VI.** Optimization methods in applied tasks [In Russian]. Moscow: SOLON PRESS, 2009: 320 p.
  - [16] **Ofitserov VP.** About optimum program execution of space researches the set number of types of carrier rockets [In Russian]. *Space researches*. 1980; 18(4): 550–555.
  - [17] **Ofitserov VP, Ofitserov MV, Bocharova OA.** About one approach to automation and informatization of process of drawing up programs of training [In Russian]. *RUDN Bulletin. Education Informatiz. series*. 2012; 4: 105–113.
  - [18] **Piyavsky SA.** Method of universal coefficients for the multi-criterial decision making [In Russian]. *Ontology of designing*. 2018; 8(3): 449-468. - DOI: 10.18287/2223-9537-2018-8-3-449-468.
  - [19] **Piyavsky SA.** Forms for calculation of universal coefficients when adopting multiple critical decisions [In Russian]. *Ontology of designing*. 2019; 9(2): 282-298. - DOI: 10.18287/2223-9537-2019-9-2-282-298.
  - [20] **Erasov IV, Ofitserov VP.** About one cardinal algorithm of processing of expert information on the basis of a method of the analysis of hierarchies [In Russian]. *Informatization of science and education*. 2013; 4: 153–161.
  - [21] **Saati T.** Decision-making at dependences and feedback: Analytical networks [In Russian]. Moscow: Publishing house LKI, 2008: 360 p.
- 

## About the authors

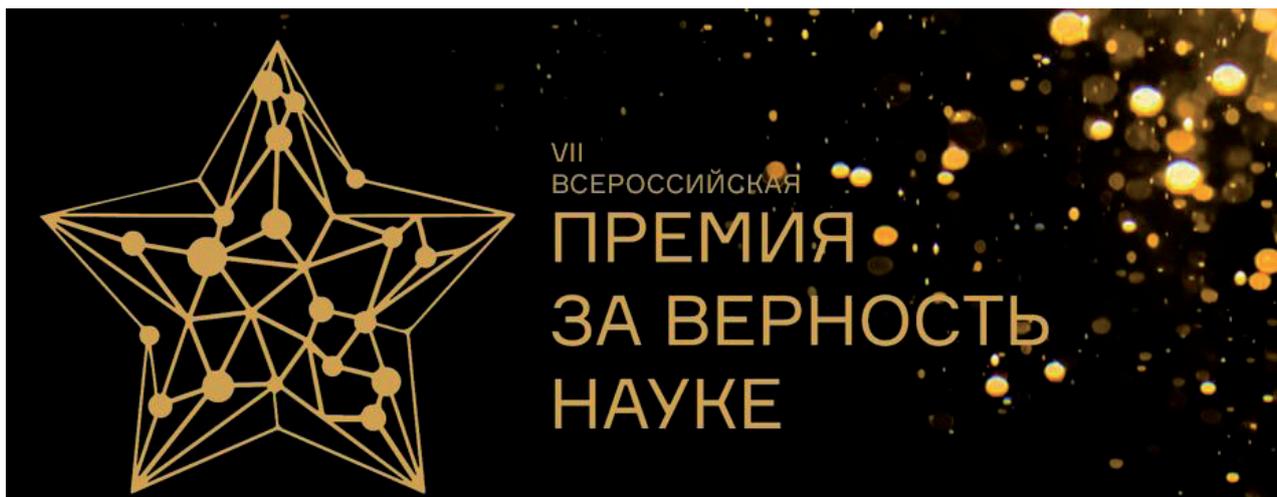
**Vladimir Petrovich Ofitserov** (b. 1953) graduated from the Korolyov aviation Institute (Kuibyshev-city) in 1975, PhD (1982). He is an Associate professor of the department "Computers, systems and networks" of the Moscow aviation Institute (National Research University). He is a co-author of about 100 scientific articles and abstracts in applied mathematics, information technology, forecasting the development of complex socio-technical systems. Author ID (RSCI): 618798. [ofitserovvp@mai.ru](mailto:ofitserovvp@mai.ru)

**Sergey Victorovich Smirnov** (b. 1952) graduated from the Korolyov aviation Institute (Kuibyshev-city) in 1975, D. Sc. Eng. (2002). Chief Researcher at Institute for the Control of Complex Systems of Russian Academy of Sciences, professor at Povolzhskiy State University of Telecommunication and Informatics. He is a member of Russian Association of Artificial Intelligence and International Association for Ontology and its Applications. He is a co-author of more than 170 works in the field of applied mathematics, computer modeling of complex systems, creation of intelligent decision support systems. ORCID: 0000-0002-3332-5261; Author ID (RSCI): 17628; Author ID (Scopus): 57205017287; Researcher ID (WoS): AAB-2013-2021. [smirnov@iccs.ru](mailto:smirnov@iccs.ru)

---

*Received October 15, 2020. Revised May 05, 2021. Accepted June 05, 2021.*

---



*«Премия «За верность науке» — яркий пример того, как можно объединить усилия университетов, научного сообщества, медиаотрасли и экспертов для популяризации достижений наших ученых. В Год науки и технологий эта премия имеет особый статус. В ней примут участие те, кто каждый день доступно и понятно рассказывают о достижениях российских ученых, объясняют, как устроены процессы в нашем мире и Вселенной, знакомя зрителя, читателя и слушателя с новейшими технологическими достижениями страны. Уверен, благодаря премии мы узнаем о многих новых и интересных проектах и ближе познакомимся с популяризаторами науки», —* отмечает Министр науки и высшего образования **Валерий Фальков**.

Редакция научного журнала «Онтология проектирования», заручившись поддержкой своих читателей и членов редколлегии, подала заявку на VII Всероссийскую премию «За верность науке» в номинации *лучшее периодическое издание о науке*. Оставаясь приверженцем открытости и гласности, редакция публикует свою заявку и ждёт откликов на неё.

## ЗАЯВКА УЧАСТНИКА

### **Название проекта**

Верность науки - поиск истины и просвещение

### **Авторская группа проекта**

Боргест Н.М., Козлов Д.М., Смирнов С.В.

### **Название организации, которая подает проект**

Исполнительная редакция научного журнала «Онтология проектирования»  
(Соучредителями СМИ, зарегистрированного Роскомнадзором 16.06.2017, Свидетельство ПИ № ФС 77-70157, являются: ИПУСС РАН - СамНЦ РАН, Самарский университет, ООО «Новая техника»)

### **Номинация**

Лучшее периодическое издание о науке

### **Сайт номинанта (ссылка на проект)**

[http://agora.guru.ru/display.php?conf=scientific\\_journal](http://agora.guru.ru/display.php?conf=scientific_journal)

<https://www.ontology-of-designing.ru/>

## ОПИСАНИЕ НОМИНАНТА

### **История проекта**

«Онтологи и проектанты всех стран и предметных областей, присоединяйтесь!» - под таким лозунгом проект стартовал в 2011 году с регистрации научного журнала «Онтология проектирования» издательством «Новая техника».

Онтология проектирования, как направление в искусственном интеллекте, исследует и стремится формализовать знания субъектов проектирования о процессе проектирования артефактов, включая разработку тезаурусов предметной области, баз данных и знаний, процедур учёта проектной неопределённости и оптимизации.

Этими вопросами активно занимаются университеты, академические институты, проектные и ИТ-компании.

В 2015 году Самарский национальный исследовательский университет имени академика С.П. Королева (Самарский университет) и Институт проблем управления сложными системами РАН (ИПУСС РАН – СамНЦ РАН), сотрудники которых активно участвуют в работе журнала, стали соучредителями периодического издания о науке.

**Цель журнала** - в рамках научной дискуссии попытаться понять и объяснить суть процесса сотворения, развития, самоорганизации, моделирования и проектирования будущего через познание процесса создания артефактов, эволюции и отбора.

### **Задачи журнала**

- объединить усилия отечественных и зарубежных учёных, сконцентрировав внимание на актуальных проблемах методологии проектирования и предложив площадку для научных дискуссий,
- предложить подходы, технологии, методы и алгоритмы для решения практических задач в различных областях и сферах проектной деятельности,
- оказать содействие молодым ученым в понимании их ответственности за развитие современных технологий, проектирующих и создающих в итоге будущее цивилизации.

### **Описание**

*Поиск истины в науке и просвещение* – это то, что члены исполнительной редакции взяли за основу своей деятельности. Содержание журнала распространяется по международной лицензии CC-BY 4.0, которая позволяет всем неограниченно использовать, распространять, копировать публикации и создавать новые материалы.

Онтология объединяет гуманитариев и инженеров в создании информационных систем и способствует выстраиванию «моста» между гуманитарными и техническими знаниями.

Важным в журнале является сопутствующее информационное окружение, которое готовит исполнительная редакция в каждом номере. Помимо традиционных для научных журналов публикаций, редакционные статьи направлены на актуальные вопросы современной науки, на анализ достижений гигантов, на плечах которых развиваются современные исследования, на обзор публикуемых статей. Переводы ежегодных Онтологических саммитов Международной ассоциации по прикладным онтологиям (The International Association for Ontology and its Applications, IAOA), анонсы родственных научных конференций, подборки аннотаций рекомендуемых изданий по тематике журнала формируют актуальное информационное поле для исследователей, отвечая главному тезису – поиску истины и просвещению. Значимым в реализации этого тезиса является регулярное проведение одноимённого журналу научного семинара.

## Результаты

С 2012 года журнал включён в РИНЦ, в 2015 году - в Перечень ВАК, а в 2020 году - в список журналов, входящих в базу данных RSCI на платформе Web of Science.

Помимо сайта журнала, полный текст номеров журнала можно найти в базах данных EBSCO Publishing на платформе EBSCOhost и научной электронной библиотеки КиберЛенинка. Журнал размещён в коллекции «Издания по естественным наукам» на платформе EastView.

Выпущено 40 номеров журнала, проведено 20 заседаний научного семинара.

## Информация об авторской группе

Авторская группа – это ядро исполнительской редакции научного журнала «Онтология проектирования», который демонстрирует связь университетской и академической, фундаментальной и прикладной науки, объединяя усилия университетов, научного сообщества, медиаотрасли и экспертов для популяризации достижений в науке.

Работая совместно со дня основания научного журнала, связывающего гуманитарные и технические знания, ключевым для редакции была подготовка редакционных статей. Эти статьи ориентированы на вовлечение в научный поиск молодых людей, в них показаны связь и опора на ушедших гигантов, в них - тренды в науке и оценка публикуемых научных результатов.

*Боргест Николай Михайлович* – к.т.н., профессор Самарского университета, выпускающий редактор журнала «Онтология проектирования», автор одноимённого учебного курса для магистрантов и аспирантов; с.н.с. ИПУСС РАН - СамНЦ РАН; директор издательства «Новая техника»; член ИАОА и Российской ассоциации по искусственному интеллекту (РАИИ), автор более 200 работ по автоматизации проектирования и искусственному интеллекту.

*Козлов Дмитрий Михайлович* – к.т.н., доцент Самарского университета, редактор журнала «Онтология проектирования», Почётный работник высшего профессионального образования России. В списке научных трудов более 100 работ.

*Смирнов Сергей Викторович* – д.т.н., г.н.с. ИПУСС РАН - СамНЦ РАН, профессор Поволжского государственного университета телекоммуникаций и информатики, зам. главного редактора журнала «Онтология проектирования». Член РАИИ, ИАОА. В списке научных трудов более 200 работ в области прикладной математики, компьютерного моделирования, создания интеллектуальных систем поддержки принятия решений.

ФИО заявителя	Боргест Николай Михайлович
Место работы	Самарский национальный исследовательский университет имени академика С.П. Королева (Самарский университет)
Должность	профессор
E-mailадрес	<a href="mailto:borgest@yandex.ru">borgest@yandex.ru</a>

*Dumspiro, spero!*

*Ontologists and designers of all countries and subject areas, join us!*



## ВКЛАД РЕДАКЦИИ ЖУРНАЛА В «ПОИСК ИСТИНЫ И ПРОСВЕЩЕНИЕ»

Публичные мероприятия, организуемые редакцией журнала, в формате научного семинара (2011-2021 гг.)

Дата	Темы научных семинаров	Докладчики
29 июня 2021 года	Онтологический подход к обучению автоматизированному проектированию в машиностроении	<i>Шараухова А.Г.</i> , аспирант Самарского государственного технического университета (Самара)
25 мая 2021 года	Формализация смысла: понятия, принципы, модели, примеры	<i>Редозубов А.Д.</i> , сотрудник Фонда им. академика Натальи Бехтеревой (Санкт-Петербург)
20 апреля 2021 года	Понятность онтологической модели как характеристика её качества	<i>Микони С.В.</i> , д.т.н., профессор, в.н.с. СПИИРАН (Санкт-Петербург)
4 февраля 2021 года	Онтология научного направления: формирование, развитие, примеры	<i>Боргест Н.М.</i> , к.т.н., профессор Самарского университета (Самара)
14 февраля 2020 года	Роль конструктора при проектировании авиационной техники с целевой себестоимостью	<i>Орлов Е.В.</i> , главный специалист управления технико-экономического анализа АО «ТРАНСМАШХОЛДИНГ» (Москва)
14 ноября 2019 года	О научных конференциях ПУМСС (Самара), ЗОНТ (Новосибирск), КИИ (Ульяновск), национальной стратегии развития ИИ Онтология инжиниринга качества и системы менеджмента	<i>Смирнов С.В.</i> , д.т.н., г.н.с. ИПУСС РАН – СамНЦ РАН (Самара)  <i>Дмитриев А.Я.</i> , к.т.н., доцент Самарского университета, заместитель директора Научно-консультационного центра экспертизы (Самара)
27 марта 2019 года	Разработка концепции и проект скоростного амфибийного транспортного средств	<i>Назаров Д.В.</i> , к.т.н., доцент Самарского университета (Самара)
9 ноября 2018 года	<b>Технологии принятия многокритериальных решений:</b> Точные формулы для вычисления универсальных коэффициентов при принятии многокритериальных решений. Современные технологии оптимизации. Примеры использования применительно к сложным техническим системам.	<i>Пиявский С.А.</i> , д.т.н., профессор Поволжского государственного университета телекоммуникаций и информатики (Самара) <i>Егоров И.Н.</i> , д.т.н., профессор, заместитель генерального конструктора по системному анализу НПО «Сатурн» (Москва)
12 апреля 2018 года	Онтология проектирования науки: будущее научных журналов, роль университетов	<i>Боргест Н.М.</i> , к.т.н., профессор Самарского университета (Самара)
8 декабря 2017 года	Онтология проектирования атмосферных псевдоспутников (АПС)	<i>Болдырев А.В., Камалиева Р.Н., Кириллов А.В., Лукьянов О.Е., Назаров А.А., Назаров Д.В.</i> - сотрудники Самарского университета (Самара)
1 ноября 2016 года	Онтогносеологический потенциал проблемы конструктивной условности Как «нумеризовать» понятие «важнее»  Холонический подход к управлению интеллектуальным развитием обучаемых	<i>Огнев А.Н.</i> , к.филос.н., доцент Самарского университета (Самара) <i>Пиявский С.А.</i> , д.т.н., профессор Самарского государственного технического университета (Самара) <i>Самойлов Е.А.</i> , д.пед.н., профессор Самарского государственного социально-педагогического университета (Самара)
30 июня 2016 года	<b>Пути развития онтологии проектирования:</b> Презентация юбилейного 20-го номера журнала «Онтология проектирования». Пятилетние итоги издания Проблемы подготовки статей в журнал «Онтология проектирования»: опыт редактирования Коммюнике онтологического саммита 2016	<i>Боргест Н.М.</i> , к.т.н., профессор Самарского университета, выпускающий редактор журнала «Онтология проектирования» (Самара) <i>Козлов Д.М.</i> , к.т.н., доцент Самарского университета, редактор журнала (Самара) <i>Коровин М.Д.</i> , аспирант Самарского университета (Самара)

	Онтология проектирования – взгляд философа  Композиционный метод проектирования	<i>Нестеров А.Ю.</i> , д.филос.н., профессор, заведующий кафедрой философии Самарского университета (Самара) <i>Малахов С.А.</i> , д.архитектуры, заведующий кафедрой инновационного проектирования Самарского государственного архитектурно-строительного университета (Самара)
23 апреля 2015 года	Конструктор онтологий: функционал, примеры решенных задач	Брифинг НПК «Разумные решения» (Самара)
3 июля 2014 года	Управление сложностью	<i>Ржевский Г.</i> , почетный профессор Открытого университета Милтона Кейнса (Букенгем)
18 февраля 2014 года	Наука о сознании: нерешённые проблемы	<i>Агафонов А.Ю.</i> , д.психол.н., профессор, заведующий кафедрой общей психологии Самарского государственного университета (Самара)
22 марта 2013 года	Мультиагентные технологии для управления ресурсами в реальном времени	<i>Скобелев П.О.</i> , д.т.н., главный конструктор НПК «Разумные решения»
27 сентября 2012 года	Университет будущего - каков он?	Мозговой штурм, дискуссия
27 октября 2011 года	Неокибернетика: алгоритмы, математика эволюции и технологии будущего	<i>Крылов С.М.</i> , д.т.н., профессор Самарского государственного технического университета (Самара)
26 мая 2011 года	<b>Концепция робота-конструктора самолета:</b> Онтология проектирования – «новое» научное направление. Роботизация исследовательской и проектной деятельности. Тезаурус – основа интеллекта систем проектирования.	<i>Боргест Н.М.</i> , к.т.н., доцент Самарского государственного аэрокосмического университета (Самара) <i>Шустова Д.В.</i> , аспирант СГАУ (Самара)
6 мая 2011 года	Представление знаний в инженерном деле	<i>Смирнов С.В.</i> , д.т.н., директор ИПУСС РАН (Самара)



## ВКЛАД РЕДАКЦИИ ЖУРНАЛА В «ПОИСК ИСТИНЫ И ПРОСВЕЩЕНИЕ»

Редакционные статьи в номерах журнала  
(2011-2021 гг.)

Том	№	Темы редакционных статей
11	2(40)	Оппозиция знания и незнания
	1(39)	Бесконечность... В ожидании «чёрных лебедей»
10	4(38)	Искусственного много – интеллекта мало
	3(37)	«Всего самого светлого»
	2(36)	Онтология пандемии: реальная и мнимая
	1(35)	Онтология конституции: проблемы проектирования
9	4(34)	«Где исчезают виртуальные миры?»
	3(33)	У науки женское лицо
	2(32)	Измерять неизмеримое
	1(31)	Выход за границы Индустрии
8	4(30)	30-й юбилейный выпуск
	3(29)	20 лет «ПУМСС»: краткие итоги
	2(28)	Модный тренд: цифровая экономика
	1(27)	Вперёд, в будущее!
7	4(26)	Дедлайн в жизненном цикле
	3(25)	«Физиолика»
	2(24)	Гештальт: генезис, поиск, проектирование
	1(23)	Муза Урания нам в помощь
6	4(22)	«Мера лучше всего»
	3(21)	«Всё своё ношу с собой»
	2(20)	20-й юбилейный выпуск журнала
	1(19)	Изучая настоящее, проектируем будущее
5	4(18)	Ум как первопричина
	3(17)	Критерий – весь жизненный цикл
	2(16)	Первый в списке мудрецов
	1(15)	Понимание греческой философии
4	4(14)	Триада Витрувия
	3(13)	О первых началах
	2(12)	Scientometric Introspection
	1(11)	Точка опоры
3	4(10)	«Standing on the shoulders of giants»
	3(09)	«Начала» современной онтологии
	2(08)	Выборы будущего российской науки
	1(07)	В тени учеников о цели бытия
2	4(06)	Поиск точек отсчета
	3(05)	Устремлённые в будущее... Принципы генерала Кузнецова
	2(04)	«Догонять тех, кто впереди...»
	1(03)	«Платон мне друг, но истина дороже»
1	2(02)	Уподобимся Сократу или апология онтологии проектирования
	1(01)	«Последняя цель науки в том, чтобы доставлять человеку как можно больше удовольствия...»

В обзорной статье Боргеста Н.М. «Онтологии проектирования от Витрувия до Виттиха» в журнале «Онтология проектирования» за 2018 год №4 (DOI: 10.18287/2223-9537-2018-8-4-487-522) на рисунке 3 в формате инфографики по временной шкале отмечен ряд учёных, внесших значительный вклад в онтологию и проектирование. Имена этих учёных упомянуты в журнале в статьях «От редакции» в период 2011-2018 гг.

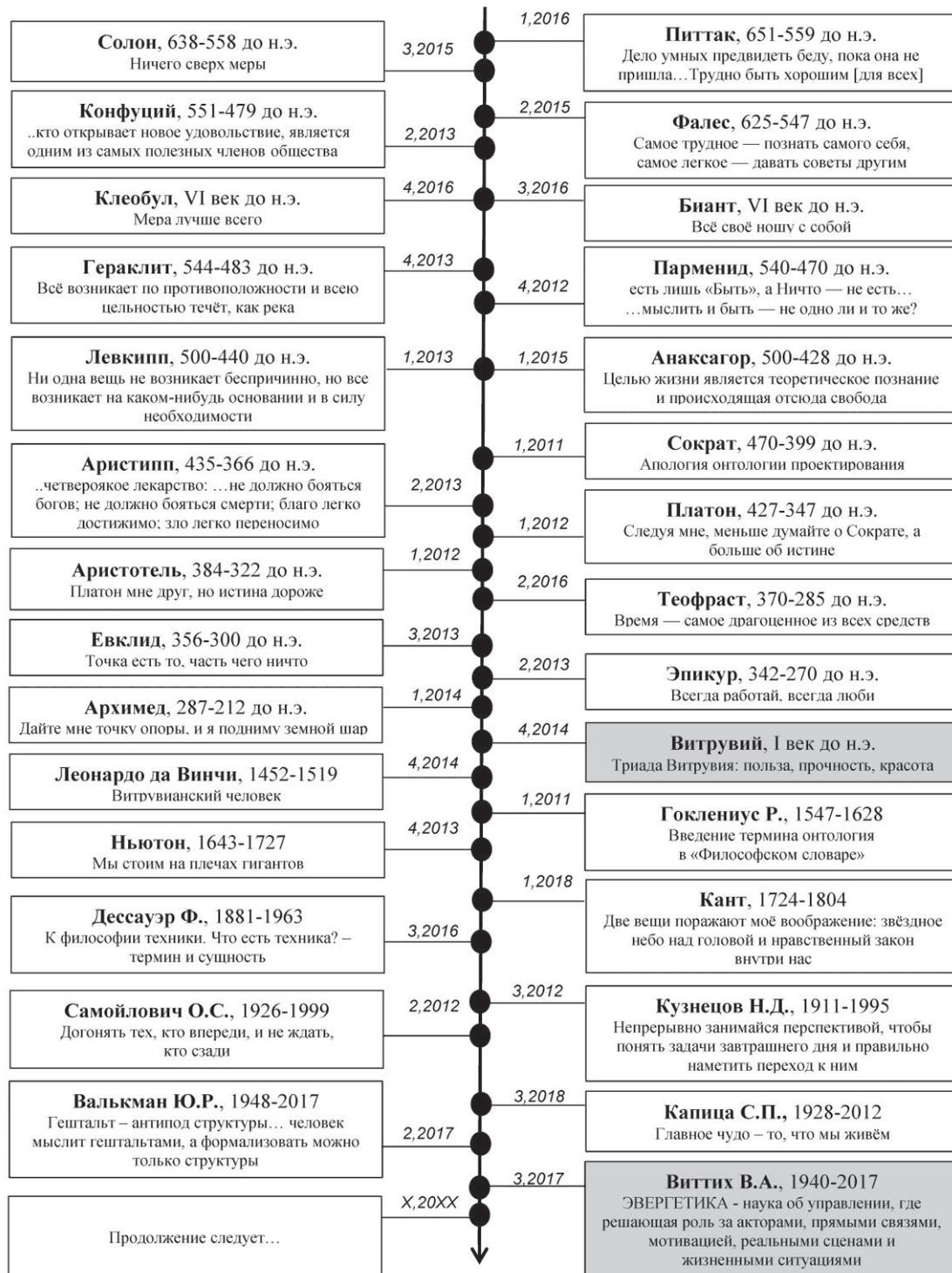


Рисунок 3 – Учёные (далеко не полный перечень по временной шкале), внесшие вклад в онтологию и проектирование, чьи имена упоминались в журнале «Онтология проектирования» в статьях «От редакции» за период 2011-2018 гг.

За последние 2,5 года (в период 2019-2021 гг) к этому списку в журнале прибавилось упоминание имен корифеев науки и рассказы о них. Это: Анаксимандр, Цицерон, Галилео Галилей, Фома Аквинский, Айзек Азимов, Альберт Эйнштейн, Дмитрий Пospelов, Марио Бунге, Стивен Хокинг, Александр Фридман, Андрей Линде, Нассим Талеб, Виктор Соифер и мн.др. Продвижение заслуг гигантов, на плечах которых стоит и развивается наука, *привлекает внимание молодых учёных, популяризирует достижения науки, вносит вклад в поддержку престижа профессиональной научной деятельности.*

## Термины, их определения и принципы классификации в новом стандарте ГОСТ Р 59277-2020

Любой стандарт фиксирует не только лучшие практики, но в первую очередь термины, их определения и принципы классификации, принятые в исследуемой предметной области. Целью нового ГОСТ Р 59277-2020 является установление принципов классификации систем искусственного интеллекта (СИИ). Внедрение данного стандарта необходимо для повышения эффективности использования СИИ при решении прикладных задач.

### НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ Системы искусственного интеллекта. Классификация систем искусственного интеллекта *Artificial intelligence systems. Classification of artificial intelligence systems* (выдержки)

Дата введения — 2021-03-01

#### Термины и определения

*Автоматизированная система*: Система, состоящая из персонала и комплекса средств автоматизации его деятельности, реализующая информационную технологию выполнения установленных функций.

*Автоматическая система*: Совокупность управляемого объекта и автономной СИИ, функционирующая самостоятельно, без участия человека.

*Автономность* (autonomy): Характеристика СИИ, связанная с её способностью самостоятельно (без участия человека) выполнять возложенные на неё функции в течение заданного времени и с заданными показателями качества, надёжности, безопасности.

*Агент* (agent): Физический/программный объект, который оценивает собственное состояние, состояние других объектов и окружающей среды для выполнения своих действий, включая прогнозирование и планирование, которые максимизируют успешность, в том числе при неожиданном изменении оцениваемых состояний, достижения своих целей.

*Алгоритм* (algorithm): Конечное упорядоченное множество точно определенных правил для решения конкретной задачи.

*Библиотека знаний* (knowledge library): Набор информационных (знаковых, символьных) моделей, которые выражают знания (также могут включать в себя определение моделей и их требования) о ряде вещей (понятий) и хранятся и воспроизводятся в электронном виде.

*Большие данные* (big data): Обширные наборы данных — главным образом, по таким характеристикам данных, как объём, разнообразие, скорость генерации и/или изменчивость, — которые требуют использования технологии масштабирования для эффективного хранения, обработки, управления и анализа.

*Данные* (data): Предоставление информации в формальном виде, пригодном для передачи, интерпретации или обработки людьми или компьютерами.

*Действие* (activity): Множество связанных задач процесса.

*Доверие к системе искусственного интеллекта* (trustworthiness): Уверенность потребителя и при необходимости, организаций, ответственных за регулирование вопросов создания и применения СИИ, и иных заинтересованных сторон в том, что система способна выполнять возложенные на неё задачи с требуемым качеством.

*Знания* (в искусственном интеллекте) (knowledge): Совокупность фактов, событий, убеждений, а также правил, организованных для систематического применения.

*Искусственный интеллект* (artificial intelligence): Комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение, поиск решений без заранее заданного алгоритма и достижение инсайта) и получать при выполнении конкретных практически значимых задач обработки данных результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека.

*Интернет вещей* (internet of thing): Инфраструктура взаимосвязей сущностей, систем и информационных ресурсов совместно с сервисами, которые снимают с вещей первичные данные, обрабатывают и выдают информацию для физического или виртуального мира.

*Интероперабельность* (interoperability): Способность двух или более информационных систем или компонентов к обмену информацией, в том числе на организационном, семантическом и техническом уровнях, и к использованию информации, полученной в результате обмена.

*Интероперабельная система* (interoperable system): Система, в которой входящие в неё подсистемы работают по независимым алгоритмам, не имеют единой точки управления, все управление определяется единым набором стандартов — профилем интероперабельности.

*Информационная технология* (information technology): Методы, способы, приёмы и процессы обработки (сбора, накопления, ввода-вывода, приема-передачи, хранения, поиска, регистрации, преобразования, анализа и синтеза, предоставления, отображения, распространения и уничтожения) информации с применением программных и технических средств.

*Класс* (class): Описание множества объектов, для которых имеются одни и те же атрибуты, операции, методы, взаимосвязи и семантики.

*Классификация* (classification): Способ и результат упорядочения, структуризации некоторого множества объектов, разделения его на определённые подмножества путём артикуляции, выделения некоторого признака объектов исходного множества как основания их структуризации по данному признаку. Такого рода признак называется основанием классификации.

*Объект классификации*: Элемент классификационного множества.

*Классификационная схема*: Классификационная структура, основанная на отношениях подчинения.

*Многоагентная система* (multiagency system): Система, состоящая из множества взаимодействующих интеллектуальных агентов. Многоагентные системы могут решить проблемы, которые трудны или невозможны для отдельного агента или для единой (монолитной) системы.

*Модель знаний* (knowledge model): Информационная модель, которая выражает знания в структуре, интерпретируемой компьютером.

*Онтология* (ontology): Формализованное представление набора наименований понятий в предметной области и отношений между этими наименованиями понятий.

*Процесс* (process): Совокупность взаимосвязанных или взаимодействующих видов деятельности, преобразующая входы в выходы.

## Принципы классификации систем искусственного интеллекта

Искусственный интеллект (ИИ) охватывает все области человеческой деятельности, включая: информатику, математику, философию, психологию, лингвистику, когнитивные науки и др. ИИ используется в таких приложениях, как системы управления, системы принятия решений, многоагентные системы, системы обработки естественного языка, распознавание образов и речи, интеллектуальный анализ данных и др. приложения.

В стандарте установлена схема классификации, отражающая основные особенности СИИ для решения прикладных задач, помогающая определить направления их стандартизации.

Базовые классы СИИ целесообразно группировать на основе следующих принципов:

- 1) по классам и категориям объектов в управлении;
- 2) по технологиям построения, приобретения и использования знаний;
- 3) по функциям, которые выполняет СИИ в контуре управления;
- 4) по методам и технологиям, используемым в СИИ;
- 5) по методам и средствам взаимодействия СИИ с другими системами и человеком-оператором.

Дополнительные классификации могут быть связаны со специальными требованиями к объектам, процессам, контуру управления, архитектуре, ресурсам с учётом окружающей среды (интероперабельность, нормы регулирования, безопасность, действия стандартов, этические требования и т.д.). Классификация, связанная с описанием каждого класса, представляет собой перечень объектов, соответствующих данному классу.

В стандарте рассмотрены следующие основания для классификации;

- 1) по степени автономности;
- 2) по степени автоматизации;
- 3) по архитектурному принципу;
- 4) по структуре и процессам обработки знаний:
  - а) по модели знаний;
  - б) по управлению знаниями;
  - в) по методу обучения;
- 5) по специализации СИИ:
  - а) специализированные (используют единый домен знаний);
  - б) комплексные (используют множество доменов знаний);
- 6) по методам обработки информации;
- 7) по функциям в контуре управления;
- 8) по методам достижения интеграции и интероперабельности СИИ;
- 9) по опасности последствий;
- 10) по конфиденциальности;
- 11) по видам деятельности;
- 12) по взаимодействию с человеком-оператором.

## Рекомендуемые издания 2021 гг. по тематике журнала



Лексин В.Н.

### **Искусственный интеллект в экономике, политике и частной жизни:**

*Опыт системной диагностики.* - М.: URSS. 2021. 336 с.

Представлены результаты исследования феноменально быстрого становления искусственного интеллекта (ИИ) предметом большой экономики, большой политики и все более значимой составляющей частной жизни. Критически оцениваются различные подходы к определению сущности ИИ, обосновываются причины его вхождения в центр общественного внимания, анализируются факты стимулирования развития ИИ. Выявляется специфика рынка разработок ИИ, характеризуются основные поставщики разработок ИИ, рассматриваются механизмы стимулирования прикладных разработок ИИ и создания

альянсов крупнейших корпораций для ускорения таких разработок. Проанализированы последствия воздействия ИИ на рынок труда. Монография предназначена для исследователей, занимающихся ИИ и связанными с ним проблемами.



Белов М.В., Новиков Д.А.

### **Модели деятельности:**

*Основы математической теории деятельности.* - М.: URSS. 2021. 216 с.

Книга содержит математические основания методологии (учения об организации деятельности) и включает общую модель любой человеческой деятельности, а также её частные случаи, учитывающие следующие факторы: действие и результат деятельности, опыт и технологии деятельности, внутреннее состояние субъекта деятельности и используемые ресурсы. Значительное внимание уделяется специфике креативной деятельности. Книга ориентирована на специалистов, занимающихся исследованием общих принципов организации деятельности и управления сложными социальными системами.



Малинецкий Г.Г., Войцехович В.Э., Вольнов И.Н., Колесников А.В.,

Скиба И.Р., Сороко Э.М.

### **Красота и гармония в цифровую эпоху:**

*Математика — искусство — искусственный интеллект.*

*Будущее и гуманитарно-технологическая революция.* – М.: URSS. 2021. 240 с.

В книге раскрываются философско-математические основания красоты и меры в контексте теории развития. В работах классиков мы находим подсказки для нахождения путей эволюции цивилизации, зашедшей в духовный тупик. Авторы предлагают новое понимание математики, музыки, живописи, информационной культуры и обсуждают возможные гармонические траектории развития человека и человечества с целью выхода из глобального кризиса. Предложено несколько образов возможного и желаемого будущего. Книга будет интересна как профессионалам в области науки, технологии и культуры, так и начинающим исследователям.



Шимукович П.Н.

### **ТРИЗ-идеи в системном изложении:**

*Путь к творческому результату: целостное представление и пошаговый анализ.* - М.: URSS.

2021. 432 с.

Изложенный в книге материал включает в себя как заимствованные сведения о существовании творческих процессов, так и впервые публикуемые результаты, раскрывающие механизм их сопряжения с основной творческой процедурой — авторскими рекомендациями по системному решению проблем. Процессы анализируются пошагово, что позволяет получить детальное представление как о каждом из них, так и об их взаимосвязи. Комплексный взгляд на процесс создания нового позволяет исследователю

фокусировать внимание на тех сторонах творчества, которые ранее оставались фоновыми, второстепенными, лучше понимать процесс творения в целом и в итоге — более эффективно творить. Книга рассчитана на широкий круг читателей.



Целищев В.В.

### **Алгоритмизация мышления:**

*Гёделевский аргумент.* - М.: URSS. 2021. 304 с.

Профессор Дж. Лукас из Оксфорда показал, что согласно Второй теореме Гёделя человеческий ум принципиально превосходит возможности ИИ. К нему присоединился лауреат Нобелевской премии сэр Р. Пенроуз. Началась дискуссия, в ходе которой ревнителю строгого мышления обвиняли их в нарушении канонов математической логики. Всё поменялось, когда из расшифрованных книжек Гёделя стало известно, что он на стороне этой пары. Книга описывает дискуссии менталистов и механицистов по поводу того, кого Вторая теорема Гёделя считает умнее — человека или машину. Книга предназначена для широкого круга читателей, интересующихся проблемами ИИ и связанными с ними дискуссиями.

## Национальные стратегии в области искусственного интеллекта: европейская перспектива, издание 2021 г.

(краткий обзор 150-ти страничного отчёта)

Источник:

Van Roy, V., Rossetti, F., Perset, K. and Galindo-Romero, L., *AI Watch - National strategies on Artificial Intelligence: A European perspective, 2021 edition, EUR 30745 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-39081-7, doi:10.2760/069178, JRC122684<sup>1</sup>*.

Искусственный интеллект (ИИ) станет «технологией общего назначения» нашей эпохи и окажет огромное влияние в ближайшие десятилетия на то, как люди живут и работают. Хотя технологии ИИ привлекают свою долю «шумихи», но их значимость реальна, и примеров полезных приложений ИИ уже предостаточно. И всё же полное влияние ИИ ещё впереди. Вот почему важно отслеживать прогресс разработки и внедрения ИИ. Все государства-члены ЕС разрабатывают и реализуют политику и национальные стратегии, чтобы воспользоваться преимуществами ИИ для экономики и общества. Отслеживание того, как страны поддерживают развитие ИИ и насколько успешны различные стратегии, имеет решающее значение для лиц, определяющих государственную политику. Это позволяет странам оценивать свой прогресс и извлекать уроки из передового опыта, применяемого в других странах. Чтобы подготовить эту публикацию, ЕС и ОЭСР объединили усилия, продолжая сотрудничество в области политики ИИ. База данных ЕС-ОЭСР о национальных политиках и стратегиях в области ИИ, которую начали создавать в 2019 году, содержит более 650 политик и стратегий в области ИИ более 60 стран, включая все страны ЕС.

К июню 2021 года 20 государств-членов ЕС и Норвегия приняли свои национальные стратегии ИИ, а 7 государств-членов находились на финальной стадии разработки и готовы опубликовать свою стратегию в ближайшие месяцы. В таблице 1 представлен обзор национальных стратегий ИИ в государствах-членах ЕС и Норвегии.

Таблица 1 - Обзор национальных стратегий ИИ в странах-членах ЕС и Норвегии.

Country	Status	Date	Country	Status	Date
 Austria	In progress		 Italy	In progress	
 Belgium	In progress		 Latvia	Published	Feb. 2020
 <b>Bulgaria</b>	Published	Dec. 2020	 Lithuania	Published	Mar. 2019
 Croatia	In progress		 Luxembourg	Published	May 2019
 <b>Cyprus</b>	Published Last update	Jan. 2020 Jun. 2020	 Malta	Published	Oct. 2019
 Czech Republic	Published	May 2019	 Netherlands	Published	Oct. 2019
 Denmark	Published	Mar. 2019	 Norway <sup>AC</sup>	Published	Jan. 2020
 Estonia	Published	Jul. 2019	 <b>Poland</b>	Published	Dec. 2020
 <b>Finland</b>	Published Last update	Oct. 2017 Nov. 2020	 Portugal	Published	Jun. 2019
 France	Published	Mar. 2018	 Romania	In progress	
 <b>Germany</b>	Published Last update	Nov. 2018 Dec. 2020	 Slovakia	Published	Jul. 2019
 Greece	In progress		 <b>Slovenia</b>	Published	May 2021
 Hungary	Published	Sept. 2020	 <b>Spain</b>	Published	Dec. 2020
 Ireland	In progress		 Sweden	Published	May 2018

Источник: JRC - Европейская комиссия.

<sup>1</sup> Это совместная публикация Объединенного исследовательского центра Европейской комиссии (JRC), Службы науки и знаний Европейской комиссии и Организации экономического сотрудничества и развития (ОЭСР). Представленные результаты не подразумевают политической позиции Европейской комиссии или ОЭСР... Мнения, выраженные в отчёте, принадлежат исключительно авторам и не могут рассматриваться как изложение позиции Европейской комиссии.

*Примечание:* Последнее обновление таблицы выполнено 1 июня 2021 года<sup>2</sup>. В таблице представлены даты выпуска национальных стратегий ИИ на их родном языке. Страны, выделенные полужирным шрифтом, опубликовали или обновили свои национальные стратегии ИИ с момента выпуска предыдущего отчёта *AI Watch* в феврале 2020 года. Помимо стран-членов ЕС, в таблице указана Норвегия как ассоциированная страна, выделенная надстрочным индексом *АС*. Швейцария не намерена выпускать национальную стратегию ИИ.

После публикации отчёта *AI Watch* за 2020 год о национальных стратегиях ИИ, Болгария, Венгрия, Польша и Испания приняли свои стратегии в третьем и четвёртом кварталах 2020 года, а Словения опубликовала свою стратегию в мае 2021. Другие государства-члены, такие как Кипр, Финляндия и Германия, пересмотрели свои первоначальные стратегии. Кипр объявил тендер на разработку Плана действий по отслеживанию прогресса и реализации своей национальной стратегии ИИ. В ноябре 2020 года Финляндия запустила программу «Искусственный интеллект 4.0» для содействия развитию ИИ в компаниях. Германия дважды пересматривала свою стратегию ИИ и в декабре 2020 года обязалась увеличить первоначальные государственные инвестиции в ИИ с 3 до **5 миллиардов евро**. Восемь государств-членов (Австрия, Бельгия, Греция, Ирландия, Италия, Румыния, Хорватия) имеют окончательные проекты своих стратегий и намерены опубликовать их до конца 2021 года.

## Политики в области ИИ

Инициативы национальной политики представлены в стратегиях в области ИИ по следующим направлениям (см. рисунок 1):

- *Человеческий капитал*: включает в себя все политики, способствующие образовательному развитию людей в использовании и разработке решений ИИ. Он включает аспекты формального образования и обучения, профессионального и непрерывного обучения, а также информацию о потребности рынка труда;
- *От лаборатории к рынку*: включает политические инициативы, направленные на поощрение исследований и инноваций в области ИИ для роста бизнеса в частном секторе и повышения эффективности государственных услуг. Эта область также включает инструменты политики для облегчения тестирования и экспериментов с недавно разработанными пилотными проектами и услугами в области ИИ;
- *Сетевое взаимодействие*: охватывает все политические инициативы, связанные с сотрудничеством в области ИИ в частном и государственном секторах и направленные на повышение (международной) национальной привлекательности страны. Эта область также включает политики, связанные с распространением и внедрением ИИ, а также рекламные кампании приложений ИИ;
- *Регулирование*: охватывает политику по разработке этических принципов, законодательных реформ и международной стандартизации;
- *Инфраструктура*: охватывает инициативы по поощрению сбора, использования и обмена данными, а также по развитию цифровой и телекоммуникационной инфраструктуры.

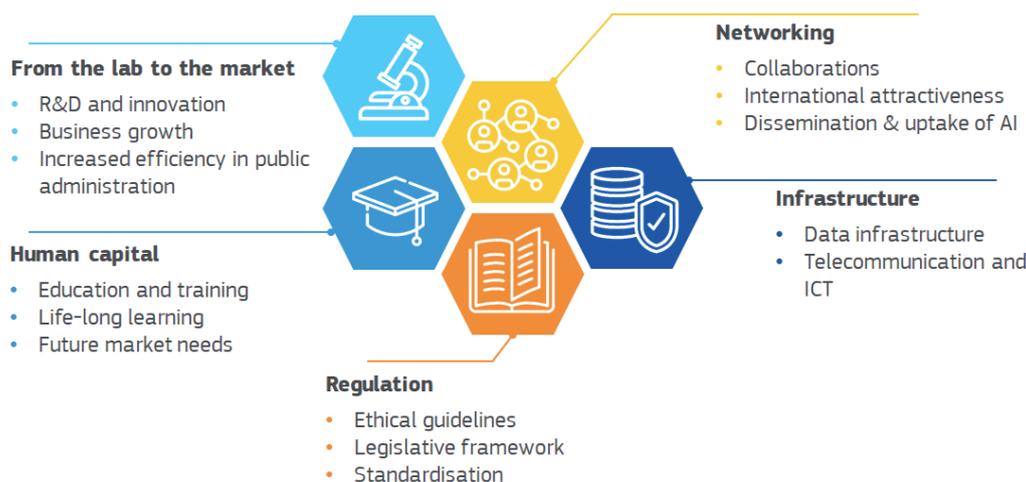


Рисунок 1 - Обзор соответствующих политик в области ИИ

<sup>2</sup> В журнале «Онтология проектирования» ранее рассматривались обзоры стратегий развития ИИ, принятые в ведущих странах мира, при этом подробно анализировались стратегии, принятые в России и США. См. статью: *Боргест, Н.М.* Стратегии интеллекта и его онтологии: попытка разобраться / Н.М. Боргест // Онтология проектирования. – 2019. – Т. 9, №4 (34). – С.407-428. – DOI: 10.18287/2223-9537-2019-9-4-407-428.

# Онтологический Саммит 2021: генерация и гармонизация онтологий<sup>1</sup>

## Ontology Summit 2021: Ontology Generation and Harmonization

Июнь 2021

Онтологический Саммит - это ежегодная серия мероприятий, в которых участвуют сообщества онтологов и сообщества, связанные с темой каждого года, выбранной для Саммита. Онтологический Саммит проводят *Ontolog*<sup>2</sup> и *NIST*<sup>3</sup>, а сама программа организована совместно *Ontolog*, *NIST*, *NCOR*<sup>4</sup>, *NCBO*<sup>5</sup>, *IAOA*<sup>6</sup>, *NCO\_NITRD*<sup>7</sup> наряду с другими организациями, которые поддерживают цели и задачи Саммита.

### 1 Темы Саммита 2021

Онтологии - это богатая и универсальная конструкция. Их можно извлечь, изучить, разбить на модули, связать, преобразовать, проанализировать и согласовать, а также разработать в рамках формального процесса. На Саммите изучаются многие виды онтологий и способы их использования. Цель состоит в том, чтобы познакомить как состоявшихся, так и потенциальных пользователей онтологий с возможностями того, как онтологии могут использоваться для решения проблем.

Различные типы онтологий находят разное применение и разные методы использования. Можно выделить следующие основные типы онтологий: базовые, справочные, предметные и прикладные. Различные их типы могут пересекаться. На Саммите исследуются типы онтологий, предлагаются рекомендации о том, как определить тип онтологии и как её использовать.

Определение - это формальное заявление о смысле или значении объекта, включая слова, фразы, классы и свойства. Определения могут служить связями между формальными онтологиями и неформальными онтологиями, а также между различными формальными онтологиями. Исторические попытки стандартизировать термины включали создание основных моделей метаданных и общих концептуальных моделей для объединения данных в единое представление. Однако они в значительной степени не были приняты из-за ошибочной концепции, отсутствия согласия сообщества, недостаточного представительства и, таким образом, оказались разрознены. Некоторый прогресс достигнут благодаря использованию пере-

<sup>1</sup> Как и в прошлом году, редакция журнала решила не дожидаться итоговой версии очередного Коммюнике Онтологического Саммита за 2021 год, который обсуждается с осени 2020 года. Читателю предлагается ознакомиться лишь с общей концепцией исследуемой тематики. - <https://ontologforum.org/index.php/OntologySummit2021>. Редакция планирует разместить перевод Коммюнике Онтологического Саммита за 2021 год после его принятия.

<sup>2</sup> ONTOLOG («Форум онтологов») - это открытое международное виртуальное сообщество практиков, занимающихся продвижением онтологий, онтологической инженерии и семантических технологий, которое выступает за их внедрение в основные приложения и международные стандарты. - <https://ontologforum.org/index.php/WikiHomePage>.

<sup>3</sup> Национальный институт стандартов и технологий, США – *NIST (US National Institute of Standards and Technology)* - <https://www.nist.gov/>.

<sup>4</sup> Национальный центр онтологических исследований, США – *NCOR (National Center for Ontological Research)* - <https://ubwp.buffalo.edu/ncor/>.

<sup>5</sup> Национальный центр биомедицинской онтологии, США – *NCBO (National Center for Biomedical Ontology)* - <https://www.bioontology.org/>.

<sup>6</sup> Международная ассоциация по Онтологиям и ее приложениям – *IAOA (The International Association for Ontology and its Applications)* - <https://iaoa.org/>.

<sup>7</sup> Национальное координационное бюро для Программы исследования и развития сетевых и информационных технологий США - *NCO NITRD (The (US) National Coordination Office (NCO) for the Networking and Information Technology Research and Development (NITRD) Program)*. - <https://www.nitrd.gov/>.

дового опыта, включая использование онтологического анализа и проектирования. На Саммите рассматриваются различные понятия и уровни формальности определений и практические методы согласования различных семантических ресурсов.

Теперь онтологии можно создавать с использованием автоматизированных методов, таких как обработка естественного языка и машинное обучение. На Саммите рассмотрены современные методы, включая нейросимволические подходы и подходы здравого смысла, сочетающие логику и языковую обработку. Цель состоит в том, чтобы предложить рекомендации по наиболее подходящему использованию техник.

Многие организации, включая правительственные агентства, органы по стандартизации и коммерческие фирмы, используют онтологии и разработали инструменты для различных онтологических операций, таких как создание, эволюция, отображение и другие формы гармонизации. На Саммите исследуется степень, в которой организации заботятся об устойчивости. Одним из аспектов устойчивости является экосистема, в которой мы все живем, и то, как онтологии могут помочь в достижении устойчивости. Другой аспект - достижение устойчивых онтологий в сообществах, обслуживаемых организациями.

## 2 Цель Саммита 2021

В рамках продвижения онтологии целью Саммита является вывод онтологической науки и инженерии в передовое направление за счёт обобщений, которые могут облегчить обсуждение и обмен знаниями среди заинтересованных сторон. Выводы должны быть подкреплены примерами из различных областей, а результаты отражены в форме *Коммюнике Саммита 2021 года* с расширенным вспомогательным материалом и размещены в Интернете и в журнальных статьях.

## 3 Темы предыдущих Саммитов

- 2020 - «*Knowledge Graphs*»
- 2019 - «*Explanations*»
- 2018 - «*Contexts in Context*»
- 2017 - «*AI, Learning, Reasoning, and Ontologies*»
- 2016 - «*Framing the Conversation: Ontologies within Semantic Interoperability Ecosystems*»
- 2015 - «*Internet of Things: Toward Smart Networked Systems and Societies*»
- 2014 - «*Big Data and Semantic Web Meet Applied Ontology*»
- 2013 - «*Ontology Evaluation Across the Ontology Lifecycle*»
- 2012 - «*Ontology for Big Systems*»
- 2011 - «*Making the Case for Ontology*»
- 2010 - «*Creating the Ontologists of the Future*»
- 2009 - «*Toward Ontology-based Standards*»
- 2008 - «*Toward An Open Ontology Repository*»
- 2007 - «*Ontology, Taxonomy, Folksonomy: Understanding the Distinctions*»
- 2006 - «*Upper Ontology Summit*»



The  
International  
Association for  
Ontology and  
its Applications



**Индекс 29151**

## **ГОД НАУКИ И ТЕХНОЛОГИЙ В РОССИИ - 2021**

### **СОДЕРЖАНИЕ НОМЕРА**

#### **ОТ РЕДАКЦИИ**

Оппозиция знания и незнания

#### **ОБЩИЕ ВОПРОСЫ ФОРМАЛИЗАЦИИ ПРОЕКТИРОВАНИЯ: ОНТОЛОГИЧЕСКИЕ АСПЕКТЫ**

**А.Д. Редозубов** (*Санкт-Петербург*)

Формализация смысла. Часть 1

#### **ПРИКЛАДНЫЕ ОНТОЛОГИИ ПРОЕКТИРОВАНИЯ**

**Г.Ю. Гуськов, А.М. Наместников,  
А.А. Романов, А.А. Филиппов** (*Ульяновск*)

Формирование базы знаний для поддержки процесса архитектурного проектирования программных средств

**С.В. Кузнецов, Д.В. Кознов** (*Санкт-Петербург*)

Управление мастер-данными в рамках итеративного подхода

#### **ИНЖИНИРИНГ ОНТОЛОГИЙ**

**Н.В. Максимов, А.А. Лебедев** (*Москва*)

Онтологическая система «знания-деятельность»

**А.В. Видия, Н.О. Дородных, А.Ю. Юрин** (*Иркутск*)

Подход к созданию онтологий на основе электронных таблиц с произвольной структурой

#### **МЕТОДЫ И ТЕХНОЛОГИИ ПРИНЯТИЯ РЕШЕНИЙ**

**В.П. Офицеров, С.В. Смирнов** (*Москва - Самара*)

Оптимизация при ограничении числа проектных переменных

#### **Год науки и технологии в России - 2021**

Премия «За верность науке». Заявка участника

Термины, определения и принципы классификации в новом стандарте

Рекомендуемые издания 2021

**Национальные стратегии в области искусственного интеллекта:  
европейская перспектива 2021**

**Онтологический Саммит 2021**

---

*Ontologists and designers of all countries and subject areas, join us!*