

УДК 004.043

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

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

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

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

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

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

### Аннотация

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

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

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

### Введение

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

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

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

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

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

## 1 Проблемно-ориентированные модели

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

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

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

$$(1) CResource = \langle Content\ Type, Type, Verbosity, Role, Language, Complexity, Prerequisites, Results \rangle,$$

где:

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

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

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

где:

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

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

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

где:

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

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

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

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

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

где:

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

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

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

которую составляют перечисляемые далее компоненты.

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

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

где:

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

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

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

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

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

UnR - модель требований обучающего.

## 2 Разработка онтологической модели

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

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

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

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

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

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

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

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

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

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

### 3 Предполагаемые конечные пользователи и сценарии применения онтологии

Для разрабатываемой онтологии выделяются следующие группы пользователей:

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

### 4 Нефункциональные и функциональные требования

Разрабатываемая онтология должна отвечать следующим функциональным требованиям (ФТ):

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

Разрабатываемая онтология должна соответствовать следующим нефункциональным требованиям:

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

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

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



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

## 5 Определение возможности повторного использования существующих онтологий

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

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

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

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

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

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

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

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

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

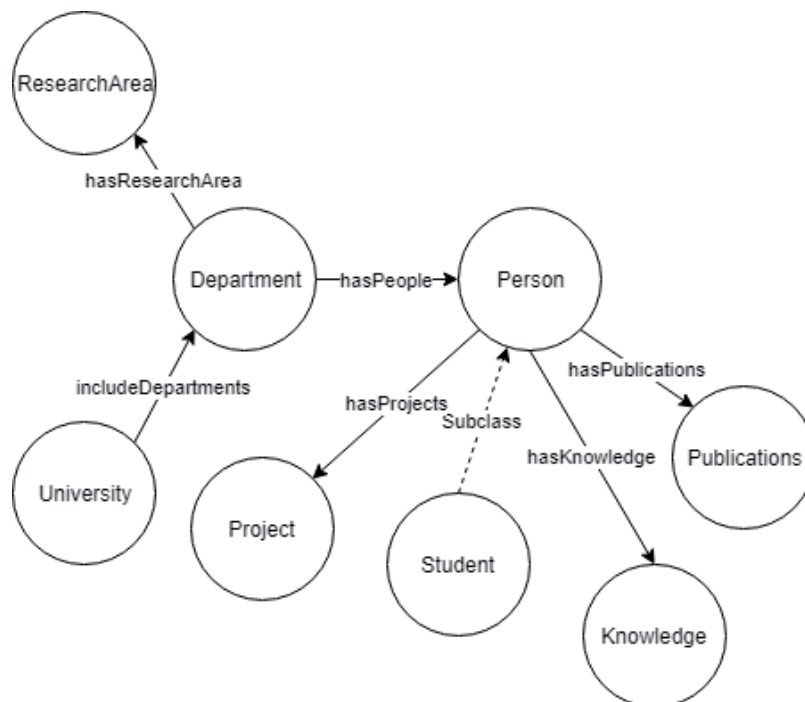


Рисунок 1 - Структура модуля описания пользователя (фрагмент)

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

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

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

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

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

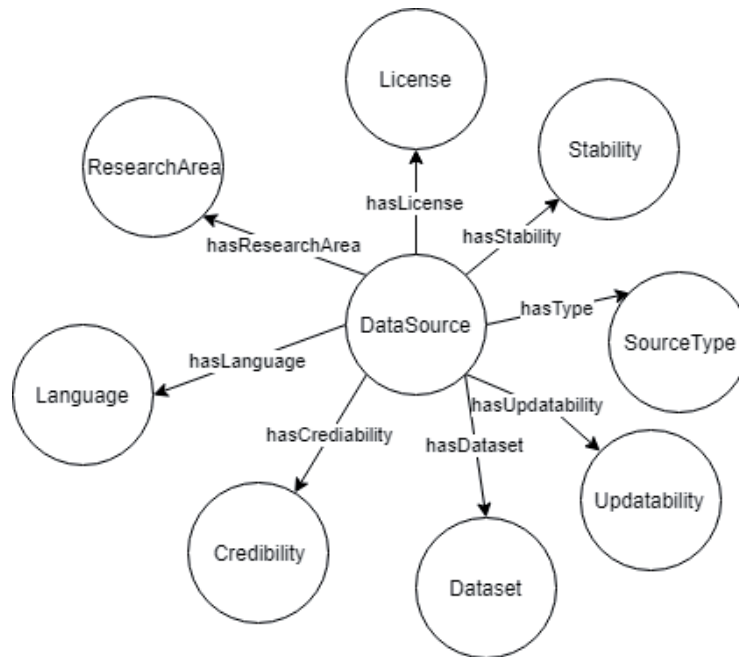


Рисунок 2 - Структура модуля описания источников данных

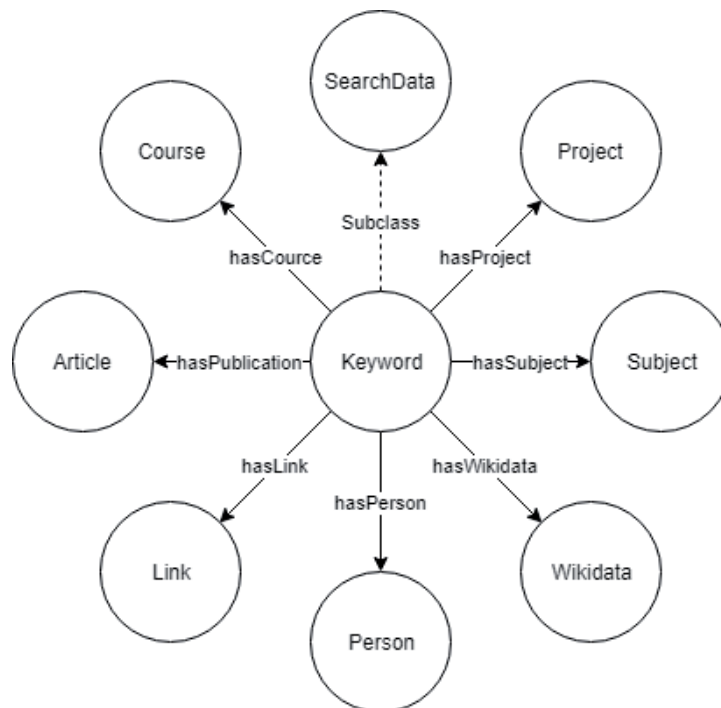


Рисунок 3 - Структура модуля описания поисковой выдачи

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



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

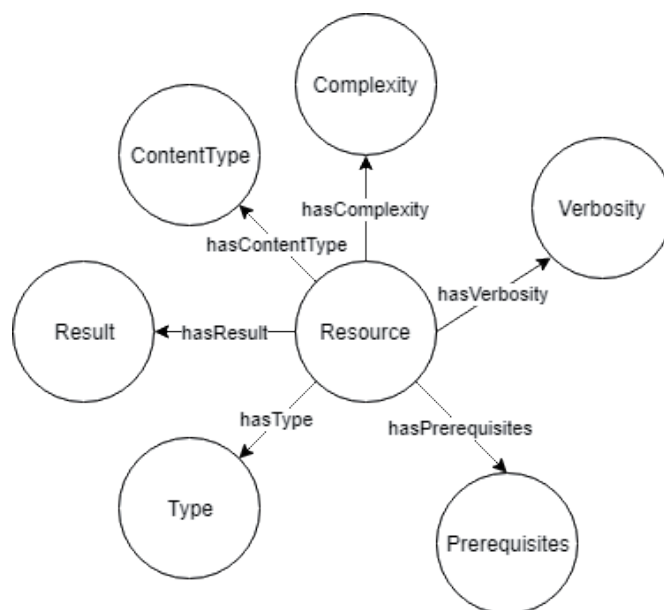


Рисунок 4 – Структура модуля описания учебного ресурса

## 6 Верификация онтологической модели

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

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

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

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

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

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

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

Таблица 1 – Результат выполнение SPARQL-запроса

№	source
1	DBpedia
2	LOD-IFMO
3	Wikidata

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

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

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

Таблица 2 – Результат выполнения SPARQL-запроса (Фрагмент)

№	user
1	Navrotskiy_MA
2	Navrotskiy_RA
3	Mouromtsev_DI

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

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

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

Таблица 3 – Результат выполнения SPARQL-запроса

№	source
1	SemanticWeb
2	Ontology
3	OWL20
4	LinkedOpenData

## 7 Применение онтологической модели

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

Ниже приведён пример *входных данных*.

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

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

где:

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

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

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

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

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

### Заключение

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

## Semantic Web

Data
Data graph
Structure graph

ENG
RUS

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

dbpedia
Wikipedia
W3C

**Publications**

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

**People**

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

dbpedia

Рисунок 5 – Пример выходных данных (фрагмент)

## Список источников

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

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

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

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

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

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

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

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

### Abstract

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

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

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

### References

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



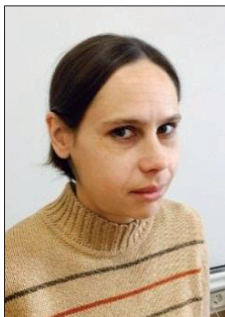
- [7] *Suárez-Figueroa MC, Gómez-Pérez A, Fernández-López M.* The NeOn methodology for ontology engineering //Ontology engineering in a networked world. – Springer Berlin Heidelberg, 2012. – P.9-34.
- [8] *Janowicz K.* et al. Five stars of linked data vocabulary use // Semantic Web. – 2014; 5(3): 173-176.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

## Сведения об авторах



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

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



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

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



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

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

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