

## ESCOLE — СЕМАНТИЧЕСКИЙ АГРЕГАТОР ОТКРЫТЫХ ОБРАЗОВАТЕЛЬНЫХ РЕСУРСОВ

Стафеев Сергей Константинович, Муромцев Дмитрий Ильич,  
Козлов Федор Алексеевич

### Аннотация

В статье описываются разработка и методы наполнения онтологий для системы ESCOLE. Система ESCOLE является системой дистанционного обучения, функционал которой основан на работе с онтологиями. С помощью разработанных алгоритмов в системе производится автоматизация сбора и поддержки учебных материалов из внешних источников. Разработанные методы наполнения онтологий позволяют создавать дополнительные связи между объектами на основе информации в самих объектах. Анализ косвенных семантических связей между объектами учебных курсов позволяет оценивать качество и состояние учебной программы.

**Ключевые слова:** семантические сети, связанные данные, онтологии, анализ учебных материалов, NLP алгоритмы.

### 1. ВВЕДЕНИЕ

В настоящее время повторное использование учебных материалов сетевых образовательных ресурсов является одним из наиболее перспективных подходов для разработки систем дистанционного обучения. Технологии Semantic Web[1] и Linked Data [2] позволяют системам обмениваться данными в сети с использованием онтологий и уникальных идентификаторов ресурсов URI (Uniform Resource Identifier). Системы, используя данные технологии, могут интегрировать и адаптировать данные из сторонних источников. Формальное описание данных и семантические связи между объектами позволяют производить сложный анализ данных системы. В системах дистанционного обучения использование данных в формате RDF (Resource Description Framework) позволяет представлять структуру курса как набор связей и объектов, таких как курс, лекция, модуль и тест [3]. В статье описана разработка онтологий и методов их наполнения для системы дистанционного обучения ESCOLE (Enhanced Course Ontology for Linked Education) [4]. В статье описываются методы анализа содержания учебных материалов на основе оценки связей в онтологиях.

#### 1.1. Обзор аналогов

Семантические технологии широко используются в образовательных ресурсах в большинстве развитых стран. Наиболее известным проектом в области связанных данных является инициатива Linked Universities. Linked Universities является альянсом европейских университетов, распространяющих свои данные, программы, курсы и

учебные материалы в формате Linked Data. Другим успешным примером использования семантических технологий в системе дистанционного обучения является Open University. Open University является исследовательским университетом дистанционного обучения с более чем 240 тысячами студентов. Семантические технологии используются в прикладных образовательных системах. Одной из таких систем является система SlideWiki [5]. SlideWiki — это образовательная платформа, позволяющая составлять курсы на основе презентаций. Платформа реализует возможность повторного использования данных созданных курсов при составлении нового курса.

## 1.2. Цели и задачи

Главной целью системы ECOLE является сбор и связывание учебных материалов из различных источников. Для хранения учебных материалов, тестов, предметных областей и результатов обучения студентов необходимо описать модели данных и онтологии. Для поддержания учебных материалов в актуальном состоянии в системе должны быть реализованы механизмы автоматического наполнения и обновления онтологий. При автоматическом наполнении данными в системе велика вероятность снижения качества учебных курсов. Одной из задач является разработка методов и инструментов для преподавателей, позволяющих на основе анализа связей между учебными материалами контролировать качество учебного курса. Другим подходом по оценке качества учебного курса является анализ деятельности студентов в системе. На основе ответов студентов на тестовые задания может быть сформирован набор предметных терминов, вызвавших наибольшее затруднение у студентов в процессе обучения. Используя данную информацию, преподаватели могут корректировать содержание курса с целью более подробного раскрытия понятий проблемных предметных терминов.

## 2. РАЗРАБОТКА ОНТОЛОГИЙ

Данные в системе дистанционного обучения ECOLE хранятся в формате RDF. Для хранения данных в системе был разработан набор онтологий. Модель данных системы делится на три основных уровня: уровень предметных областей, уровень учебных материалов, уровень деятельности пользователей в системе. Уровни модели связаны друг с другом для обеспечения взаимодействия различных ресурсов системы. Модель данных представлена на рис. 1. Уровень предметных областей является основой модели данных и содержит информацию о предметных областях науки и образования. Сбор данных для этого уровня производится из сторонних баз знаний, таксономий и опубликованных наборов данных, таких как DBpedia [6] и Mathematics Subject Classification [7].

Уровень учебных материалов содержит информацию, необходимую для проведения учебного процесса. Уровень содержит данные по образовательным программам, курсам, тестам и мультимедиа ресурсам. Сбор данных для этого уровня производится из хранилищ университетов, открытых электронных библиотек и мультимедиа ресурсов. Связывание учебных материалов с предметными терминами и областями производится в полуавтоматическом режиме с использованием алгоритмов обработки естественного языка. Уровень деятельности пользователей в системе содержит результаты обучения студентов и статистические данные по активности пользователей. Статистика ведется в системе управления обучением LMS (Learning Management System). При ведении статистики используется информация из социальных сетей. Связывание статистики и ре-

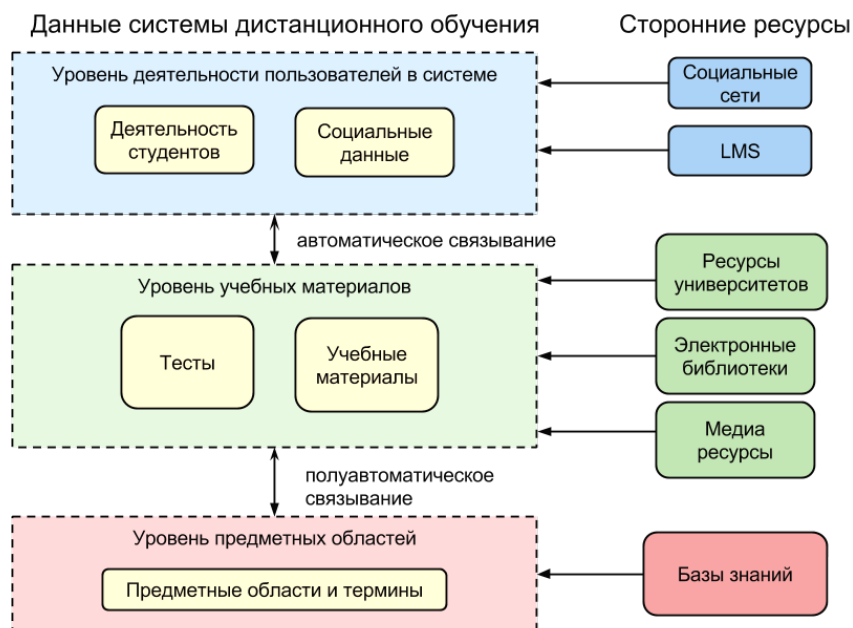


Рис. 1. Общая модель данных системы дистанционного обучения ECOLe

результатов обучения с учебными материалами происходит в автоматическом режиме алгоритмами LMS.

## 2.1. Онтология учебных материалов

Онтология учебных материалов описывает отношения между курсами, модулями, лекциями, тестами, практиками и предметными терминами. Онтология основана на онтологиях верхнего уровня, рекомендованных к использованию при описании учебных материалов [8]. AIISO (The Academic Institution Internal Structure Ontology) является онтологией, описывающей внутреннюю организационную структуру образовательного процесса. Онтология предоставляет классы и свойства для описания курсов и модулей. BIBO (The Bibliographic Ontology) является онтологией, описывающей библиографические ресурсы. В онтологии учебных материалов BIBO используется для описания рекомендованной литературы, научных публикаций, методичек и монографий. MA-ONT (The Ontology for Media Resources) является онтологией, описывающей мультимедиа ресурсы. С помощью классов и свойств MA-ONT в онтологии учебных материалов производится связывание лекций с видеоматериалами. Основными классами онтологии учебных материалов являются: Курс, Модуль, Лекция, Тест, Экзамен, Практика, Предметная область, Предметный термин и Ресурс. Онтология состоит из 32 классов, 42 объектных свойств и 13 свойств-значений.

Одной из главных особенностей разработанной онтологии является возможность произведения прямого и косвенного междисциплинарного связывания объектов в курсах. Например, тест по курсу физики «Интерференция и когерентность» включает в себя использование математических терминов, таких как «Вектор» и «Векторное произведение». Если студент не сможет успешно пройти данный тест, система должна рекомендовать к повторению не только лекции по физике, но и определенные лекции по векторной алгебре. Данный пример демонстрирует косвенное связывание курсов

физики и векторной алгебры с помощью предметных терминов «Вектор» и «Векторное произведение». Пример представлен на рис. 2. Связывание объектов курса с предметными терминами позволяет косвенно связывать лекции, тесты и методические материалы друг с другом.

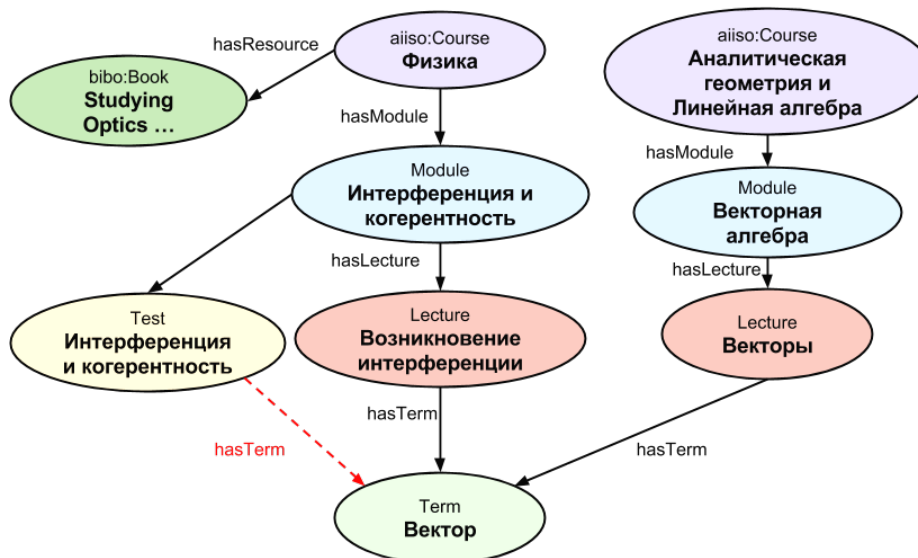


Рис. 2. Пример косвенного связывания курсов с помощью предметных терминов

## 2.2. Онтология тестов

Для описания содержания тестов была разработана онтология тестов. Разработка онтологии производилась методом раскрытия и конкретизации существующих онтологий верхнего уровня. Именно поэтому при разработке онтологии использовался нисходящий подход. Онтология тестов содержит классы, описывающие тесты, группы и варианты вопросов, задания, ответы, вопросы, различные типы вопросов и ответов [9]. Онтология содержит 12 классов, 10 объектных свойств и 6 свойств-значений. Основной целью онтологии тестов является описание структуры тестов и предоставление возможности автоматического семантического связывания заданий тестов с предметными терминами. Онтология описывает тест как набор из вариантов групп заданий [10]. Задания теста состоят из вопроса и набора ответов. В зависимости от типа вопроса у задания может быть различный набор правильных и неправильных ответов. Связывание предметных терминов с заданиями позволяет описать содержание вопроса и ответов задания.

## 2.3. Онтология активности студента в системе обучения

Онтология активности студента в системе обучения была разработана для хранения информации о прогрессе и результатах обучения студентов в системе. При разработке были использованы две онтологии верхнего уровня: онтология тестов и онтология FOAF (Friend Of A Friend). Онтология FOAF используется для описания людей и отношений между ними. В дистанционной системе обучения онтология FOAF может быть использована для описания персоналий студентов, преподавателей и их связей. Онтология активности студента в системе обучения состоит из 10 классов, 15 объектных свойств и 5

свойств-значений. Основной задачей онтологии является хранение действий студентов в системе. В онтологию может быть записана информация о просмотре студентом видеолекции, о прохождении теста или завершении курса. Онтология хранит в себе персональные данные студентов. В онтологию включены классы, описывающие результаты студентов при прохождении тестов и изучении теоретического материала. Для хранения ответов на тесты конкретного студента используется связывание с онтологией тестов. Связи между студентами, их ответами на задания тестов и предметными терминами позволяют создавать косвенные связи между студентом и объектами курса. На основе полученных косвенных связей возможна реализация персонализированной рекомендательной системы для коррекции процесса обучения студентов. После прохождения теста студент может получить не только оценку, но и список предметных терминов и материалов для повторения, составленный на основе ответов на тест.

### 3. МЕТОДЫ НАПОЛНЕНИЯ ОНТОЛОГИЙ

Технологии Semantic Web и Linked Data позволяют использовать онтологии для хранения, сбора и распространения данных. В Интернете существует множество открытых источников с данными, при описании которых были использованы онтологии верхнего уровня. Данные источники могут быть использованы для наполнения системы дистанционного обучения. Автоматизация и разработка методов наполнения онтологий является одной из главных задач в реализации механизмов поддержки учебных материалов в актуальном состоянии в системе дистанционного обучения.

#### 3.1. Интеграция данных из сторонних источников

Для сбора данных в формате RDF в системе ECOLE используются провайдеры данных. Провайдеры данных поддерживают автоматическое обновление связанных данных из сторонних источников. Провайдеры позволяют преобразовывать структурированные данные различных форматов в формат RDF. С помощью провайдеров данных система ECOLE наполняет онтологии учебных материалов и тестов. Система производит сбор данных из электронных библиотек. Одной из таких библиотек является The British National Bibliography (BNB). BNB предоставляет открытый доступ к библиографической информации в формате RDF. Библиографические данные описываются с помощью онтологии верхнего уровня BIBO. Система ECOLE собирает информацию о книгах и публикациях из библиотеки BNB и предоставляет возможность преподавателям связывать курсы с книгами BNB, используя свойство «hasResource». Система ECOLE поддерживает создание определений для предметных терминов. Одним из способов автоматизации наполнения онтологии терминов является использование внешней базы знаний. В системе ECOLE в качестве такой базы знаний используется DBpedia. DBpedia предоставляет точку доступа SPARQL (SPARQL Protocol and RDF Query Language) для получения информации, которая была извлечена из Wikipedia. Провайдер данных автоматически создает определения для предметных терминов, используя запросы к точке доступа SPARQL. Множество источников в Интернете хранят структурированные данные не в формате RDF. Тесты и учебные материалы университета могут храниться в формате XML, а электронная библиотека предоставлять информацию о публикациях через REST API. Система ECOLE использует алгоритмы конвертации данных в провайдерах для интеграции структурированных данных различных форматов в онтологию системы.

### 3.2. Обработка естественного языка

Для наполнения онтологий системы используются не только данные внешних источников, но и данные самой системы. Данные, хранящиеся в онтологии системы и связанные семантическими отношениями, позволяют создавать новые связи на основе predetermined правил. Одним из методов связывания данных в онтологиях является применение алгоритмов обработки естественного языка NLP (Natural Language Processing). Используя данный подход, можно извлекать семантические связи из текстовой информации объекта онтологии. Система ECOLE использует NLP алгоритмы для поиска предметных терминов в текстах заданий тестов. Учитывая небольшой размер образца и предустановленный набор терминов, шаблоны POS-tag в совместном использовании с синтаксическими шаблонами являются наиболее предпочтительным методом извлечения предметных терминов из заданий тестов [11–13]. Для извлечения терминов-кандидатов было использовано более десяти составных шаблонов предметных терминов. После извлечения термины-кандидаты приводились к канонической форме с использованием предустановленных словарей. Извлечение терминов было реализовано в лингвистической платформе NooJ [14]. NooJ обладает мощным механизмом регулярных выражений поиска, позволяющим комбинировать различные POS-tag шаблоны в единую грамматику для запроса к тексту. Для обработки русскоязычного текста авторами статьи был разработан набор грамматик и словарей. Алгоритм извлечения предметных терминов из текста с использованием платформы NooJ состоит из следующих шагов:

- текст задания подгружается в платформу NooJ, что приводит к его лингвистическому анализу, используя разработанные словари;
- в результате анализа платформа NooJ формирует текст с аннотациями, содержащими морфологическую и семантическую информацию о каждом слове;
- применяя запросы на основе POS-tag шаблонов, платформа NooJ формирует список терминов-кандидатов.

Для применения алгоритма в других предметных областях и на других языках необходимо сформировать соответствующие словари, грамматики и шаблоны извлечения предметных терминов. Связывание терминов-кандидатов с предметными терминами системы производится с помощью лемм. Каждый предметный термин системы обладает свойством-значением «lemma». Для реализации алгоритма связывания заданий тестов и предметных терминов был разработан провайдер данных. Провайдер принимает на вход ссылку на объект курса и производит создание ссылок между заданиями и терминами. Общий алгоритм работы провайдера данных представлен на рис. 3

Алгоритм состоит из следующих шагов:

- провайдер собирает список заданий курса, используя запросы SPARQL;
- происходит формирование текстовых данных для каждого задания с использованием информации о вопросах и ответах;
- провайдер данных запускает NLP алгоритмы в платформе NooJ для текстовых данных каждого задания;
- провайдер формирует список терминов-кандидатов, содержащих каноническую форму и набор лемм;
- производится поиск предметных терминов в системе с совпадающими наборами лемм для дальнейшего связывания с терминами кандидатами;

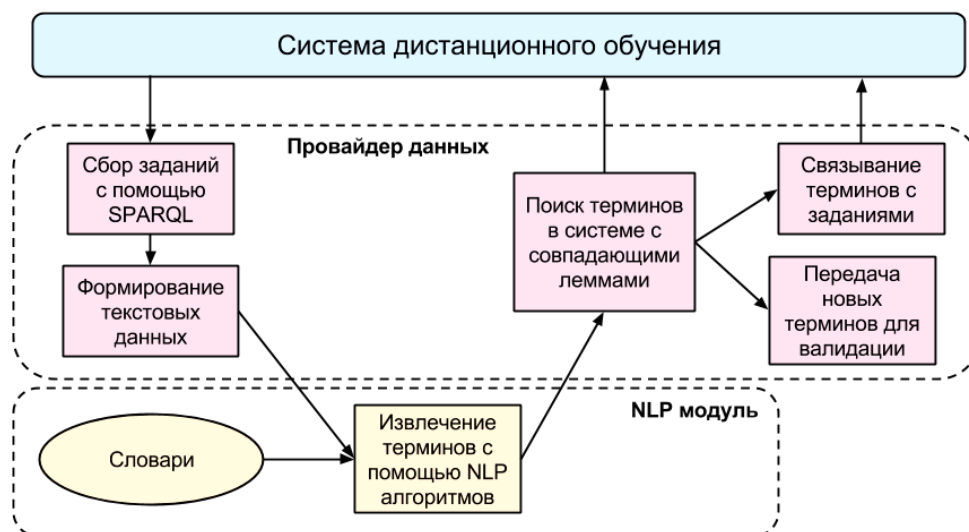


Рис. 3. Алгоритм работы провайдера извлечения терминов с использованием NLP алгоритмов

- провайдер создает связи между найденными предметными терминами системы и заданиями терминов-кандидатов с помощью свойства «hasTerm».

Термины-кандидаты, для которых не были найдены соответствующие предметные термины системы, могут быть записаны в онтологию как новые предметные термины системы. Перед записью в онтологию провайдер производит проверку термина-кандидата на соответствие предметному термину. Для проверки используются запросы к базе знаний DBpedia. В случае совпадения названия термина-кандидата со свойствами «rdfs:label» или «dbpedia-owl:wikiPageRedirects» объекта из DBpedia в онтологии системы создается новый предметный термин на основе термина-кандидата и производится связывание с заданиями теста. При отсутствии совпадений термин-кандидат помечается провайдером данных как ложный термин. При SPARQL запросах к базе знаний DBpedia используется фильтрация по предметным областям и категориям с помощью свойства «dcterms:subject». Данная фильтрация позволяет избежать ошибочных совпадений терминов-кандидатов с терминами из сторонних предметных областей. Для включения нового термина в систему необходима верификация преподавателя или администратора системы. Алгоритм проверки терминов-кандидатов представлен на рис. 4.



Рис. 4. Алгоритм проверки новых терминов-кандидатов

## 4. РЕАЛИЗАЦИЯ И РЕЗУЛЬТАТЫ

### 4.1. Архитектура системы и интерфейс пользователя

Сервер системы дистанционного обучения ECOLE основан на платформе Information Workbench [15]. Платформа Information Workbench предоставляет функционал для работы с открытыми связными данными Linked Open Data. Пользовательский интерфейс сервера ECOLE основан на модуле семантической разметки Semantic MediaWiki [16]. Данный модуль позволяет использовать predetermined шаблоны и визуальные средства для отображения семантических данных в виде Wiki-страниц. Редактирование и управление RDF данными системы реализовано с использованием платформы OpenRDF Sesame. Сервер системы ECOLE предоставляет открытую точку доступа для SPARQL запросов. Внешним интерфейсом системы дистанционного обучения ECOLE является система LMS. LMS предназначена для удобного представления учебных материалов пользователям системы. LMS обладает локальным хранилищем и производит управление пользовательскими данными, настройками и результатами обучения студентов. Внешний интерфейс системы предоставляет функционал по администрированию системы и управлению доступом к данным. В LMS реализованы модули для отображения видеолекций, слайдов, тестов и практических заданий. Внешний интерфейс взаимодействует с сервером системы с помощью запросов к открытой точке доступа SPARQL. Внешний интерфейс получает с сервера данные по учебным материалам и связи между объектами курса. Приватные персональные данные пользователей и настройки LMS хранятся в локальной памяти внешнего интерфейса. Общая архитектура системы дистанционного обучения ECOLE представлена на рис. 5.

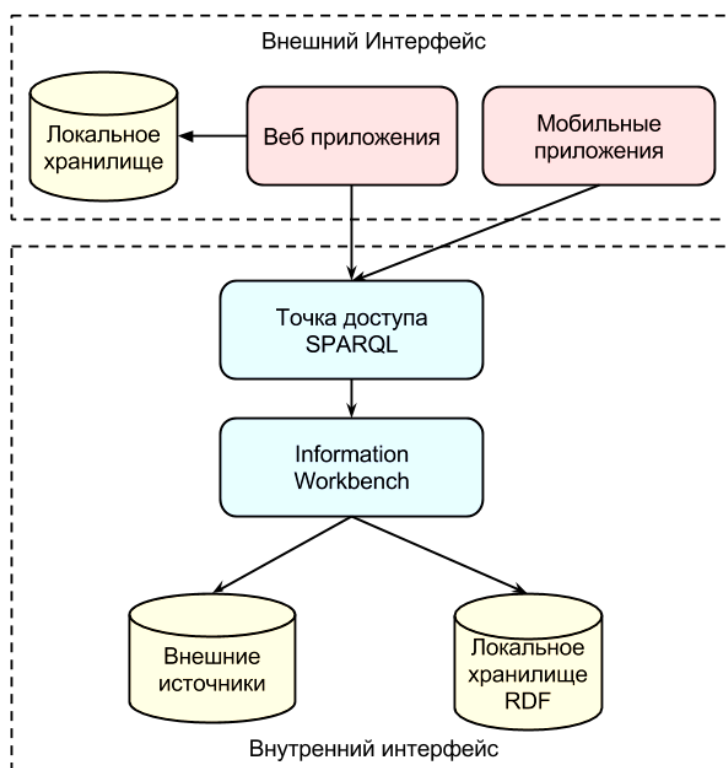


Рис. 5. Общая архитектура системы дистанционного обучения ECOLE



LMS реализована на языке Python с использованием Django Web Framework [17]. Библиотека SPARQLWrapper используется для отправки запросов к точке доступа SPARQL. Когда пользователь завершает тест, LMS собирает результаты теста, ответы на задания и дополнительную статистику и записывает полученные данные на сервер с помощью запроса SPARQL Update Query [18]. В результате прохождения теста студентом на сервере создается объект класса «AttemptToPassTest» с набором правильных и неправильных ответов на задания теста. Данный объект связывается с объектом студента. В целях безопасности персональных данных, объекты студентов идентифицируются с использованием хеш-суммы электронной почты студента. После прохождения теста система предоставляет студенту информацию о количестве и доле правильных ответов на задания теста. Также студенту предоставляется список предметных терминов для повторения. Система генерирует список проблемных терминов для студента, используя результаты теста и связи между предметными терминами системы и заданиями теста. Для каждого предметного термина, связанного с заданиями теста производится расчет рейтинга на основе ответов студента. Лист проблемных терминов сортируется в порядке возрастания рейтинга. Чем выше рейтинг термина, тем больше правильных ответов дал студент на задания, связанные с данным термином, и тем меньше затруднений вызвал у студента данный термин. Рейтинг знаний для каждого термина позволяет студенту выявлять проблемные термины и восполнять знания по ним.

#### 4.2. Анализ связей

Анализ качества учебных материалов производится внутри аналитических модулей системы. Каждый модуль состоит из отдельной аналитической страницы с графиками, таблицами и визуальными компонентами. Аналитические страницы основаны на синтаксисе Semantic MediaWiki и хранятся в платформе Information Workbench. Данные графиков, таблиц и визуальных элементов собираются аналитическим модулем с помощью SPARQL запросов к точке доступа.

Одним из методов анализа качества учебных материалов является анализ покрытия лекций курса тестами и заданиями. В онтологии системы лекции и тесты связаны с определенным модулем курса. В результате работы алгоритмов по наполнению онтологий лекции и задания тестов могут быть связаны с определенными предметными терминами. Таким образом происходит косвенное связывание лекций и тестов через предметные термины. Система может предоставить статистику по количеству предметных терминов лекций, использованных в тестах. Если термин лекции использован в задании теста, он считается покрытым в данном модуле. Каждый модуль имеет аналитическую страницу. Аналитическая страница модуля содержит следующую статистическую информацию:

- количество покрытых и непокрытых предметных терминов в модуле,
- общий процент покрытия модуля тестами на основе отношения покрытых терминов к общему количеству терминов,
- облако тегов предметных терминов модуля, демонстрирующее качество и степень покрытия каждого термина,
- список непокрытых терминов модуля, которые были использованы в лекциях, но не были использованы в заданиях тестов.

Интерфейс аналитической страницы модуля позволяет получать информацию о покрытии тестами каждой лекции в отдельности. Интерфейс представлен на рис. 6. Дру-

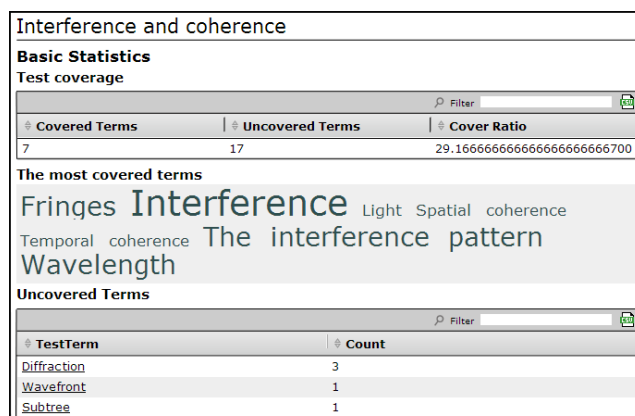


Рис. 6. Интерфейс статистики покрытия лекций тестами в аналитической странице модуля

гим подходом к анализу учебных материалов является выявление проблемных предметных терминов в модуле курса. Проблемными предметными терминами являются термины, при изучении которых у студентов возникают наибольшие затруднения. Статистика по результатам прохождения студентами тестов, правильные и неправильные ответы на задания, связанные с определенными предметными терминами, позволяют рассчитать рейтинг знания каждым студентом определенного термина. Используя данный рейтинг, преподаватель может получить список предметных терминов, которые студенты знают хуже всего. Это позволит преподавателю вносить коррективы в учебные материалы и учебный процесс. В текущей реализации аналитического модуля рейтинг проблемного термина рассчитывается делением количества неправильных ответов на количество правильных ответов в заданиях, связанных с данным термином. В будущем формула будет усложнена с учетом сложности заданий и глобального рейтинга терминов студентов. Рейтинг проблемных терминов для модуля составляется с помощью следующего SPARQL-запроса:

```

SELECT ?term
(count(?correct_answer) AS ?correct_answer_count)
(count(?answer) AS ?answer_count)
((2*?correct_answer_count - ?answer_count)
AS ?rank)
WHERE{
    ?module learningRu:hasTest ?test .
    ?test ifmotest:hasGroupOfTasks
    ?group_of_tasks .
    ?group_of_tasks ifmotest:hasTask ?task .
    ?test_element lres:hasTask ?task .
    ?test_element lres:hasAnswer ?answer .
    ?task learningRu:hasTerm ?term .
    OPTIONAL {
        ?task ifmotest:hasCorrectAnswer
        ?correct_answer
        filter( ?correct_answer = ?answer)
    }
}
GROUP BY ?term
ORDER BY ASC(?rank)
}
    
```

Анализ проблемных терминов реализован на аналитической странице модуля. Аналитическая страница модуля включает в себя список проблемных терминов с рейтингами и диаграмму отношений между пятью самыми проблемными терминами модуля. Аналитика учебного материала, основанная на использовании семантических связей между объектами системы, позволяет строить различные запросы для оценки качества учебных материалов. Полученная информация позволяет преподавателям и авторам курса выявлять и исправлять устаревшие, ошибочные и неточные учебные материалы, основываясь на динамике образовательного процесса и его структуре.

## 5. РЕЗУЛЬТАТЫ

Основной набор данных системы дистанционного обучения ECOLE формировался вручную. Часть данных была создана с помощью методов наполнения онтологии, описанных в главе 3. Набор данных системы состоит из объектов образовательного процесса, таких как курс, модуль, лекция, тест, практика, предметный термин, предметная область и книга. В результате работы NLP алгоритмов по извлечению предметных терминов из тестов были получены результаты, представленные в таблице 1. С одной стороны, с терминами было связано 95% заданий тестов. С другой стороны, более 50% терминов курса остались не связанными с заданиями. Одной из причин данного явления является косвенное употребление предметных терминов в задании. Чтобы выполнить такое задание, необходимо знать предметный термин, который не упомянут ни в тексте задания, ни в тексте ответов. Примером таких заданий являются задачи на поиск длины гипотенузы треугольника при известных катетах. Косвенно связанным предметным термином в данном примере является термин «Теорема Пифагора». В текущей реализации метода извлечение косвенных терминов не производится. В будущем планируется использование семантических связей между терминами для выявления косвенных предметных терминов в заданиях тестов.

**Таблица 1.** Результаты работы NLP алгоритмов по извлечению предметных терминов из тестов

Количество обработанных заданий	20
Процент связанных заданий, %	95
Процент несвязанных заданий, %	5
Количество извлеченных терминов-кандидатов	155
Количество терминов, извлеченных вручную	30
Термины системы, совпавшие с терминами-кандидатами, %	50
Термины-кандидаты, совпавшие с терминами системы, %	8
Термины-кандидаты, добавленные в систему после прохождения проверки, %	6
Ложные термины-кандидаты, %	86

## 6. ЗАКЛЮЧЕНИЕ

Разработанные в данной работе онтологии и методы их наполнения для системы дистанционного обучения ECOLE позволяют преподавателям использовать в электронных курсах различный учебный материал из сторонних источников. Разработанные модули и методы анализа данных системы предоставляют преподавателям инструменты для контроля качества и актуальности учебных курсов. Система ECOLE предоставляет

рейтинг предметных терминов, вызывающих у студентов затруднения при изучении. На основе данного рейтинга преподаватель может изменять теоретическое содержание курса с целью более полного и понятного раскрытия проблемных предметных терминов. Рейтинг знания предметных терминов предоставляется индивидуально для каждого студента. Данный рейтинг позволяет студентам выявить пробелы в своих знаниях и восполнить их изучением дополнительных материалов. Система ECOLE собирает и интегрирует учебные материалы из различных источников. Таким образом, система позволяет реализовать на своей основе обмен учебными материалами между университетами и другими организациями. В будущем планируется увеличение количества внешних источников данных для наполнения онтологий. Планируется интеграция новых баз знаний и таксономий, таких как Freebase, WikiData и Mathematics Subject Classification. Использование внешних баз знаний и таксономий позволит получить новые связи и отношения между предметными терминами и областями. Данные связи могут быть использованы для увеличения качества извлечения предметных терминов из текстов и анализа структуры предметных областей. Точность расчета рейтинга проблемных терминов может быть увеличена путем изменения формулы расчета с учетом дополнительных метрик, таких как сложность задания теста, глобальный рейтинг термина у студента и роль термина в учебном курсе. Внешний интерфейс системы ECOLE доступен по адресу <http://ecole.ifmo.ru>. Интерфейс сервера системы доступен по адресу <http://openedu.ifmo.ru:8888>. Разработанные онтологии доступны по адресу <https://github.com/ailabitmo/linked-learning-datasets>. Код провайдеров данных доступен по адресу <https://github.com/ailabitmo/linked-learning-solution>. Работа выполнена при государственной финансовой поддержке ведущих университетов Российской Федерации (субсидия 074-U01).

### Список литературы

1. *Berners-Lee T. et al.* The semantic web //Scientific american, 2001. Т. 284. №. 5. P. 28–37.
2. *Bizer C., Heath T., Berners-Lee T.* Linked data-the story so far. 2009.
3. *Klyne G., Carroll J.J.* Resource description framework (RDF): Concepts and abstract syntax. 2006.
4. *Mouromtsev D. et al.* Development of an Ontology-Based E-Learning System // Knowledge Engineering and the Semantic Web, 2013. P. 273–280.
5. *Khalili A. et al.* SlideWiki: elicitation and sharing of corporate knowledge using presentations // Knowledge Engineering and Knowledge Management. Springer Berlin Heidelberg, 2012. P. 302-316.
6. *Auer S. et al.* Dbpedia: A nucleus for a web of open data. Springer Berlin Heidelberg, 2007. P. 722-735.
7. *Lange C. et al.* Bringing mathematics to the web of data: the case of the mathematics subject classification // The Semantic Web: Research and Applications. Springer Berlin Heidelberg, 2012. P. 763-777.
8. *Keßler C., d'Aquin M., Dietze S.* Linked Data for science and education // Semantic Web, 2013. Т. 4, № 1. P. 1-2.
9. *Soldatova L., Mizoguchi R.* Ontology of test //Proc. Computers and Advanced Technology in Education, 2003. P. 173-180.
10. *Vas R.* Educational ontology and knowledge testing //The Electronic Journal of Knowledge Management of, 2007. Т. 5, №. 1. P. 123-130.
11. *Hulth A.* Improved automatic keyword extraction given more linguistic knowledge // Proceedings of the 2003 conference on Empirical methods in natural language processing. – Association for Computational Linguistics, 2003. P. 216-223.
12. *Хохлова М.В.* Исследование лексико-синтаксической сочетаемости в русском языке с помощью статистических методов (на базе корпусов текстов) //АКД, СПб., 2010.
13. *Большакова Е.И., Васильева Н.Э.* Формализация лексико-синтаксической информации для распознавания регулярных конструкций естественного языка // Программные продукты и

- системы, 2008. №. 4. С. 103-106.
14. *Silberztein M.* NooJ for NLP: a linguistic development environment. 2002. // URL: <http://www.NooJ4nlp.net/pages/NooJ.html> (дата обращения: 25.02.15).
  15. *Haase P., Schmidt M., Schwarte A.* The Information Workbench as a Self-Service Platform for Linked Data Applications //COLD, 2011.
  16. *Krötzsch M., Vrandečić D., Völkel M.* Semantic mediawiki // The Semantic Web-ISWC 2006. Springer Berlin Heidelberg, 2006. P. 935-942.
  17. *Holovaty A., Kaplan-Moss J.* The definitive guide to Django: Web development done right. Apress, 2009.
  18. *Geareon P., Passant A., Polleres A.* SPARQL 1.1 Update // Working draft WD-sparql11-update-20110512, W3C (May 2011). 2012.

## ECOLE — SEMANTIC AGGREGATOR OF OPEN EDUCATIONAL RESOURCES

Stafeev S. K., Muromcev D. I., Kozlov F. A.

### Abstract

The paper describes the development of ontologies and methods of ontology population for the ECOLE system. ECOLE is the online eLearning system which functionality is based on several ontologies. Designed algorithms allow gathering educational material from external resources. The methods of ontology population allow creating additional relations between objects using the information of the objects. Analysis of semantic relations between the objects of educational courses allows evaluating the quality and status of educational program.

**Keywords:** *semantic web, linked data, ontologies, analysis of educational content, NLP algorithms.*

Стафеев Сергей Константинович,  
доктор технических наук, профессор,  
декан ЕНФ СПбНИУ ИТМО,  
[stafeevs@yahoo.com](mailto:stafeevs@yahoo.com)

Муромцев Дмитрий Ильич,  
кандидат технических наук, доцент,  
руководитель международной  
лаборатории «Интеллектуальные методы  
обработки информации и семантические  
технологии»,  
[mouromtsev@mail.ifmo.ru](mailto:mouromtsev@mail.ifmo.ru)

Козлов Федор Алексеевич,  
аспирант кафедры информатики и  
прикладной математики-1 СПбНИУ ИТМО,  
[kozlovfedor@gmail.com](mailto:kozlovfedor@gmail.com)

© Наши авторы, 2015.  
Our authors, 2015.