

Разработка базы знаний по оптике для образовательных веб-приложений¹

Аннотация. В статье рассмотрены результаты работ по созданию прототипа on-line базы знаний по оптике для систем e-Learning. Рассмотрены возможности онтологического моделирования для построения систем e-Learning нового поколения. Приведена методика создания базы знаний, включая этап инженерии знаний. Описана архитектура программной реализации базы знаний, центральным компонентом которой является модуль графической визуализации онтологий, обеспечивающий простой и наглядный способ доступа к образовательному контенту.

Ключевые слова: технология дистанционного обучения, базы знаний, онтологический инжиниринг.

Введение

Создание баз знаний (БЗ) для решения практических задач уже давно привлекает внимание разработчиков как эффективный способ доставки до потребителя информации, источником которой традиционно считались специалисты-эксперты [1]. В последнее время БЗ становятся неотъемлемой частью образовательных систем и технологий [2]. Однако большинство разработок в данном направлении посвящены созданию электронных архивов и каталогов. К сожалению, при таком подходе задача определения критериев поиска необходимой информации ложится на пользователя, который может не иметь ни малейшего понятия о предметной области, а, следовательно, и не в состоянии самостоятельно осуществлять целенаправленный поиск. Единственной возможностью в этой ситуации является, по сути, фасетная классификация, что нельзя считать приемлемым образовательным инструментом.

Говоря об образовательных веб-приложениях, прежде всего, имеют в виду различные системы дистанционного образования. В последнее время все чаще упоминается термин e-Learning 2.0 [3, 4], подразумевающий использование в процессе обучения средств Web 2.0 [5]. В отличие от

традиционного подхода, когда ученику предлагается изучить некий набор материалов, выполнить тестовые задания, которые затем проверяются преподавателем, новая форма e-Learning 2.0 предполагает широкую социализацию процесса обучения и использование такого программного обеспечения как блоги, wiki, социальные сети и пр., поддерживаемые как педагогами, так и учениками. В литературе этот феномен также упоминается под названием «Длинный хвост» (Long Tail Learning) [6], так как в процесс преподавания вовлекаются не только учителя, но и большое количество учеников, обменивающихся опытом в решении частных задач, таким образом, обучая друг друга.

Вопросы реализации технологии активного обучения на базе онтологического моделирования рассмотрены авторами в работе [7]. Однако применение технологий Web 2.0 само по себе не обеспечивает решение вышеупомянутой задачи «квалифицированного ассистирования» пользователю при поиске и работе с информацией.

Одним из подходов к построению интеллектуальной образовательной системы является разработка в качестве центральной компоненты БЗ с онтологической структурой, т.е. содержащей модель предметной области. Использование стандартных форматов представления

¹ Работа поддержана грантом Правительства Санкт-Петербурга

онтологий позволяет применять широкий спектр инструментов инженерии и визуализации знаний. Также становится возможным применение методов логического вывода для удовлетворения сложных семантических поисковых запросов пользователя на основе формальных рассуждений (логического вывода).

Авторы данной статьи применили данный подход при разработке прототипа образовательной системы для центра оптических технологий СПбГУ ИТМО [8], включающего онтологическую базу знаний по оптике в качестве формальной модели предметной области и семантический wiki-портал, содержащий образовательный контент. Целевой аудиторией рассматриваемой в данной статье БЗ являются ученики старших классов средней школы, студенты младших курсов и все интересующиеся историей и проблемами оптики.

Задачами БЗ являются:

- предоставление учащимся школ, младших курсов вузов и другим пользователям оптических знаний в форме семантической сети;
- информационная поддержка лекционных и практических занятий;
- справочно-энциклопедическое обеспечение;
- повышение информационной культуры и мотивации школьников и студентов во время обучения за счет концептуализации знаний предметной области.

При этом основные функции системы, реализующие перечисленные задачи, включают:

- моделирование предметной области «оптика»;
- визуальное представление сложной картины взаимосвязей между понятиями, объектами и персоналиями в оптике, включая различные контекстные «срезы»;
- предоставление научно-образовательного мультимедийного wiki-контента;
- автоматическое построение связей между выбранными концептами предметной области;
- обеспечение доступа пользователей к ресурсам БЗ через сеть Интернет.

1. Применение онтологического моделирования в системах e-Learning

Использование онтологий в образовании не является абсолютно новой идеей. Так, напри-

мер, онтологический инжиниринг был успешно применен в Японии для улучшения ИТ-образования [9]. Одна из лучших карт памяти, описывающих применение онтологий в образовании, предлагается в работе [10]. Преимуществом использования онтологий в качестве инструмента познания является системный подход к изучению предметной области [11]. При этом достигаются:

- системность — онтология представляет целостный взгляд на предметную область;
- единообразие — материал, представленный в единой форме, гораздо лучше воспринимается и воспроизводится;
- научность — построение онтологии позволяет восстановить недостающие логические связи во всей их полноте.

Существует множество определений понятия «онтология», однако с точки зрения программных приложений существенным является то, что онтологии используются как когнитивный инструмент для таких задач, как формализация знаний, экстернализация знаний, передача знаний и оценка знаний [12]. Онтологический инжиниринг может предоставить эффективную модель процесса обучения, помогающую как преподавателю, так и ученикам получать знания, использовать и проверить их.

Применение онтологий в современных системах e-Learning основывается на прикладном и технологическом аспектах. С точки зрения приложений можно выделить следующие задачи:

- как когнитивный инструмент - создание знаний, передача знаний, оценка знаний;
- так называемые e-portfolios (электронные портфолио студентов).

С точки зрения технологии выделяются следующие группы:

- технологии представления знаний - организация и структурирование знаний, вывод знаний;
- технологии семантического веба - шаблоны метаданных, применяемые для семантического аннотирования образовательного контента, инструкций, авторской деятельности; архитектура приложений и сервисов, основанных на онтологиях, предоставляющая программным агентам осуществлять обмен знаниями, их повторное использование и интероперабельность.

Наконец, следует отметить последние тенденции к созданию так называемых систем

Linked Learning, предоставляющих возможности использования распределенного образовательного контента в рамках движения Linked Open Data [13].

2. Методика создания БЗ

В качестве основы процесса разработки БЗ была выбрана стандартная методика создания экспертных систем [1]. После определенной адаптации основными этапами методики явились:

1. На первом этапе была сформирована команда проекта, состоявшая из одного эксперта в области оптики, двух аспирантов с развитыми аналитическими способностями, выполнявших роль инженеров по знаниям. Также в команду вошли два студента-предметника, не принимавшие участие в разработке онтологии, а отвечавшие за разработку и привязку образовательного вики-контента к концептам онтологии с помощью URI.

2. Извлечение знаний из эксперта при помощи его интервьюирования инженерами по знаниям. В процессе интервью создавались визуальные карты памяти [14], описывающие определенный раздел оптики.

3. Этап формализации, в котором происходил переход от модели карт памяти к модели концепт-карт (к-карты) [15]. Полученные к-карты содержат ясный набор концептов и связей между ними.

4. Создание части онтологии и добавление ее к существующей на этот момент версии силами инженеров по знаниям. На этом этапе используется один из множества доступных на рынке инструментов (Protege, Top Braid Composer и др.) [16]. Единственным требованием к редактору онтологий является возможность сохранять онтологию в формате OWL. Полученная онтология становится доступной для просмотра и навигации в разрабатываемой системе.

5. Обсуждение полученной версии онтологии с экспертом в предметной области и ее доработка. На этом этапе с когнитивной точки зрения очень важна хорошая визуализация. Эксперт может определить и исправить некоторые отношения, которые могут портить когнитивный баланс и ясность. Не имеет большого значения, используются ли при этом специальные инструменты создания к-карт, специальные средства визуализации онтологии или же просто бумага и карандаш.

6. Шаги 2–5 повторяются итеративно до тех пор, пока проект не будет завершен. Эксперт в предметной области и инженеры по знаниям переходят к следующей итерации разработки после устранения всех неопределенностей на текущем шаге.

3. Структура онтологической базы знаний

Современный подход к разработке баз знаний основан на использовании онтологий высокого уровня, фактически представляющих собой множество классов, индивидов и свойств, присущих миру с философской точки зрения. Онтологическая схема высокого уровня заимствуется и адаптируется с учетом требований конкретной предметной области, что позволяет избежать повторов и противоречий при разработке новых онтологий.

Для создания базы данных по оптике была использована онтология DOLCE Lite (упрощенная версия DOLCE 2.0) [17]. Эта онтология хорошо подходит в качестве верхнеуровневой основы для создания описательной онтологии различных предметных областей (в нашем случае оптики), поскольку она содержит общие для всех наук абстрактные классы и отношения, такие как «явление», «понятие», «метод», «опыт», «принцип» и пр. Использование полной версии DOLCE Lite в качестве основы для онтологии оптики представлялось нецелесообразным и избыточным, так как ряд абстрактных классов никогда бы не использовался и не имел бы производных концептов в силу специфики предметной области. Поэтому перед непосредственным использованием из нее были удалены ненужные классы, она была русифицирована и пополнена специфическими для оптики отношениями.

При разработке онтологии оптики инженеры по знаниям использовали концепцию смысловых срезов, представляющих собой логическую группировку свойств онтологии, позволяющую рассматривать тот или иной вопрос с разных точек зрения, например, с точки зрения истории, практического применения, научного обоснования. Были выделены следующие смысловые срезы:

- исторический;
- практический (практическое применение);

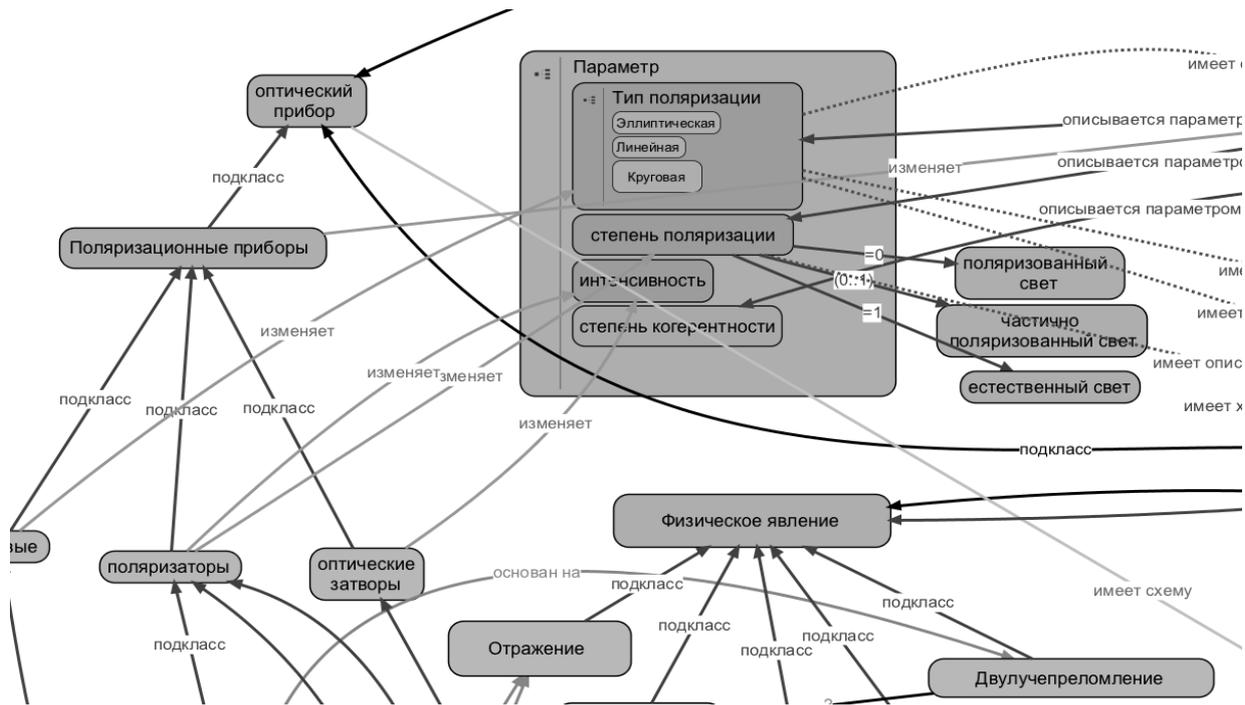


Рис. 1. Фрагмент концепт-карты, полученной в процессе инженерии знаний по оптике

- теоретический (например, понятия, лежащие в основе какого-нибудь явления).

К историческому срезу отнесены такие свойства, как «объяснил», «провел», «открыл», «доказал», «ввел понятие». Этот срез позволяет взглянуть на некоторое явление в его историческом контексте – выяснить, какие ученые его изучали, какие опыты они проводили.

К практическому срезу отнесены такие свойства, как «основано на», «использует». Этот срез позволяет, например, оценить применение явления – просмотреть основанные на нем приборы, установки или практические технологии.

К теоретическому срезу отнесены такие свойства, как «есть условие», «является параметром», «является частью». Этот срез позволяет увидеть концепты, с помощью которых рассматриваемое понятие может быть объяснено. Например, в случае интерференции это такие понятия, как «когерентность», «длина волны». Одно свойство может входить в несколько срезов. Например, свойство «основано на» входит в практический и в теоретический срез.

В процессе инженерии знаний важную роль играет использование визуальных моделей. В случае с базой знаний по оптике наиболее привлекательной моделью оказались концепт-карты [15]. В отличие от интеллект-карт Бузана

[14] концепт-карты не требуют наличия одного центра, зато позволяют вводить различные типы отношений между индивидами. Фрагмент разработанной концепт-карты приведен Рис. 1.

Преобразование к-карты в формат OWL потребовало решения ряда концептуальных вопросов, среди которых:

- в каких случаях мы должны считать концепт классом, а в каких индивидом?
- как наиболее корректно определить отношения между классами, и обеспечить наследование отношений в условиях динамической классификации?
- какие структуры в онтологии позволят минимизировать вычислительную сложность работы алгоритма рассуждений (reasoner)?

Так, например, при определении принадлежности концепта к классу или индивиду, в конечном итоге был выбран принцип «уникальности». Т.е. любой ученый является индивидом класса «персоналии», законы оптики, различные оптические схемы также следует считать индивидами, поскольку они существуют в единственном экземпляре. Аналогичные рассуждения верны и для параметров. Хотя в последнем случае дело обстоит не столь очевидно: параметр «интенсивность» однозначно может быть представлен в онтологии как инди-

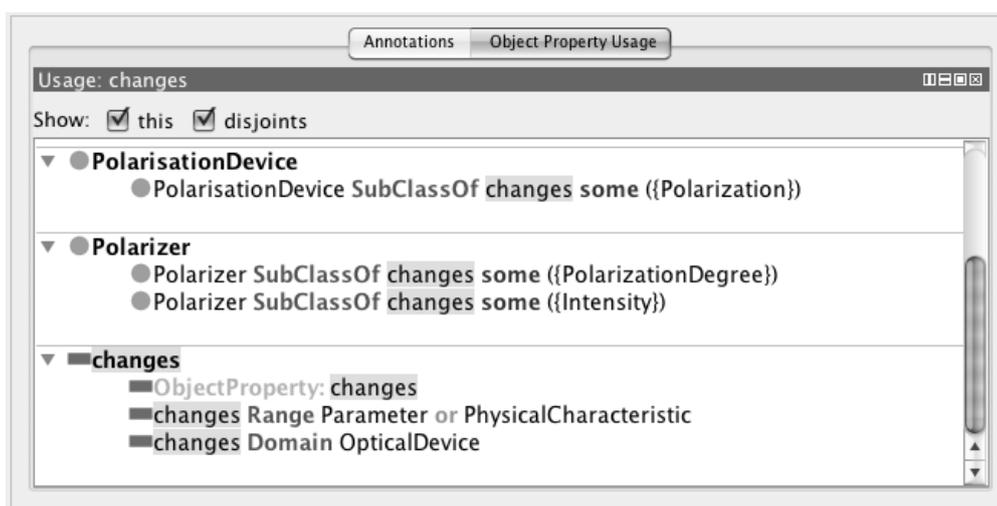


Рис. 2. Описание отношения «изменяет» в формате OWL

вид, а, например, «когерентность» бывает двух видов - «пространственная» и «временная». Также в онтологии могут присутствовать и «чистые абстракции», которые, несомненно, должны быть классами, но при этом индивиды для них создаваться не будут. Например, к таким абстракциям относится концепт «Свет».

На Рис. 2 показан фрагмент онтологии, формализующий отношение «Поляризатор изменяет интенсивность и степень поляризации» (см. фрагмент на Рис. 1). Отношение «изменяет» имеет область определения «Оптические устройства» и область значений «Параметры» или «Физические характеристики». Но в данном конкретном случае нужно указать, что поляризаторы изменяют только «степень поляризации» и «интенсивность», что и показано на Рис. 2.

4. Архитектура и принцип работы системы

Структура онтологической базы знаний предполагает значительное количество ссылок на электронные документы, ассоциированные с концептами. Задача поиска конкретной информации в этом массиве без применения специальных средств навигации может оказаться весьма трудоемкой для пользователя. Одним из возможных решений этой проблемы видится создание интерактивного модуля визуализации онтологических структур. В то же время следует отметить наличие определенных количественных и качественных сложностей. В первом случае речь идет о значительном числе концеп-

тов различных типов в онтологии. В настоящий момент их порядка тысячи, а окончательная БЗ может содержать до нескольких десятков тысяч понятий. Следовательно, для отображения каждый раз необходимо динамически выбирать некоторый фрагмент онтологии, объем которого зависит от контекста. К качественным сложностям визуализации онтологии относятся проблемы представления в виде графа всего множества возможных отношений, среди которых присутствуют различные ограничения, логические выражения и пр. Кроме того определенные сложности накладывает и динамическая классификация индивидов, а также наличие абстрактных категорий в онтологии. В совокупности, перечисленные сложности требуют специальных процедур при формировании графа для навигации по онтологии.

Для осуществления поставленной задачи была выбрана библиотека интерактивной визуализации графов SpringGraph [18], обладающая следующими важными особенностями:

- написана на Flash (используется Flex, ActionScript), поддерживается большинством платформ;
- открытый исходный код и свободная лицензия, возможности расширения;
- развитый интерактивный функционал для управления отображением графа, возможность программирования дополнительных функций;
- возможность инициализации описания графа на языке XML;
- широкие возможности настройки стиля отображения.

Рассмотрим подробнее физическое устройство модуля визуализации. Пользователю модуль представляется в виде страницы портала, на которой отображается граф. Для генерации страницы (на языке HTML [19]) применена технология JSP [20]. Сам граф рисуется flash-компонентом, построенным на основе библиотеки SpringGraph, которая обсуждалась выше. Построение и отрисовка графа производится на основе исходных данных, получаемых от портала отдельным запросом.

Логически модуль визуализации состоит из следующих компонентов:

- Конвертер данных — используется для преобразования структур онтологии в интерпретируемое flash-компонентом описание графа. Также этот модуль хранит ассоциации между именами концептов и адресами страниц с соответствующим им контентом.

- Модуль работы с онтологиями — используется для доступа к формализованным OWL-моделям онтологии.

- Сессия — служит для временного хранения пользовательских настроек отображения графа и получения доступа к местоположению пользователя на портале.

Очевидно, что на выходе требуется получить представление, включающее в себя основные элементы, составляющие онтологию. Однако flash-компонент визуализации позволяет отображать только вершины и ребра с возможностью задания стилей отображения. Для приведения сложного и многообразного языка онтологий в соответствие с имеющимися возможностями отображения конвертер данных выполняет следующие преобразования:

- объекты онтологии преобразуются в вершины с определенным стилем надписи (например, подчеркнутый шрифт);

- классы онтологии преобразуются в вершины с определенным стилем надписи, отличным от объектов (например, шрифт без изменений);

- отношения наследования между классами преобразуются в ребра графа между вершинами, соответствующими классам с определенным стилем оформления (например, черный цвет);

- отношения между объектами преобразуются в ребра графа, соответствующими объектам со стилем оформления определяемым классом отношения (например, отношение «является частью» отображается зеленым цветом).

Описанный способ преобразования данных позволяет решить задачу визуализации. При этом в процессе преобразований не теряется информация о типе связей между концептами. Так как входными данными является список визуализируемых связей различного типа: класс-подкласс, ограничение, свойство-значение, то важным становится наличие в системе таблицы ассоциаций каждого типа свойств со стилем отображения на графе. Учитывая все выше сказанное, основным звеном, разрабатываемым в рамках решения задачи, является конвертер данных из модели онтологии в xml-формат описания графа. Основное внимание в процессе реализации уделено разработке этого компонента и обеспечению его данными.

На Рис. 3. показана последовательность формирования отображения страницы для текущего концепта онтологии. При первом входе в систему текущим элементом является некий начальный концепт (задаваемый явно). В процессе ра-

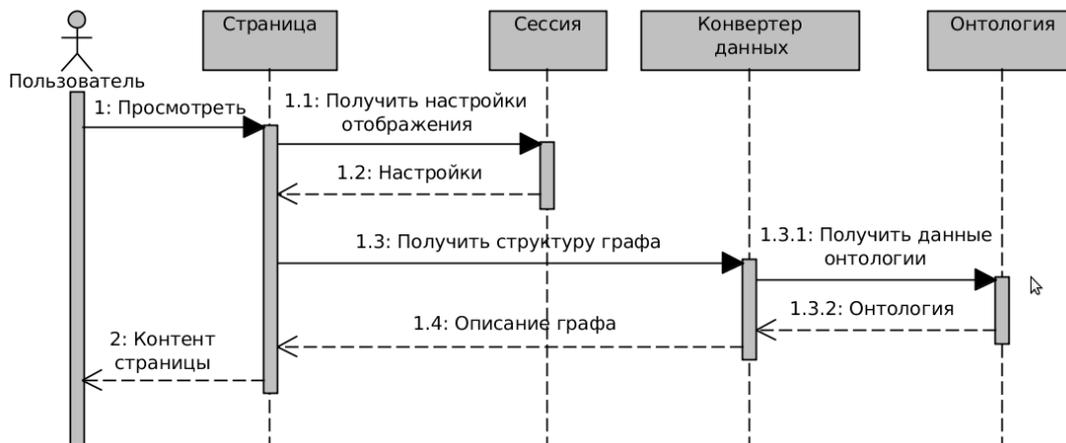


Рис. 3. Процесс построения визуализации онтологии

боты с системой пользователь, используя средства навигации по БЗ, переходит от одной страницы к другой. Соответственно меняется и фокус «проекции» графа для каждой из страниц. При этом БЗ содержит ссылки на все страницы портала, являющиеся также URI концептом онтологии (в предлагаемой реализации URI концептов совпадают с URL соответствующих страниц). Переход по этим ссылкам эквивалентен непосредственному обращению к странице. Если для какого-либо концепта страница не была указана, то система предложит создать искомую страницу по соответствующему URL, что гарантирует устойчивость системы.

В описываемом процессе отправной точкой для работы всей системы является инициируемое пользователем выполнение JSP-страницы. Во время сеанса работы сервер автоматически создает объект сессии, в котором содержатся настройки, необходимые для отображения компонентов портала (страниц). При отображении каждая страница связывается с сервисом преобразования данных из онтологии и получает требуемые данные. В том числе и корректные ссылки на другие страницы. Таким образом, портал, в целом, ведет себя как набор страниц, независимых на уровне алгоритмов отображения, но связанных на уровне онтологии.

Увеличение сложности исходной онтологии неизбежно влечет за собой проблему сохранения баланса между полнотой ее представления на графе и воспринимаемостью общей картины пользователем портала. Наличие множества типов отношений значительно усложняет граф визуализации, делая его перегруженным большим количеством информации. В качестве одного из возможных решений этой проблемы предлагается введение вспомогательного средства визуализации - облака тегов, позволяющего пользователю портала осуществлять переключение между основными смысловыми срезами онтологии. Смысловые срезы, являясь логической группировкой онтологических концептов, дают возможность пользователю базы знаний взглянуть на структуру онтологии с различных точек зрения. Облако тегов предоставляет дополнительную информацию о насыщенности смыслового среза концептами, устанавливая соответствие между количеством концептов и размером соответствующего тега. Такое комбинирование таксономического подхода с категоризацией отношений между кон-

цептами позволяет организовать более удобную семантическую навигацию между страницами портала.

Заключение

Описанный в данной статье прототип базы знаний по оптике является результатом выполнения первого этапа по созданию образовательного on-line инструментария для систем e-Learning. В ходе исследования были проработаны вопросы интеграции вики-контента и онтологической модели, визуализации различного рода отношений в форме интерактивного графа, а также методики онтологического инжиниринга с применением концепт-карт. Уже первое знакомство педагогов школ и преподавателей вузов с данной БЗ вызвало значительный интерес, так как предложенный подход обладает высокой наглядностью, когнитивной эргономичностью, значительной гибкостью программной реализации и масштабируемостью. Авторы предполагают продолжить развитие данного прототипа до полноценного образовательного ресурса. Важным достижением также является разработанная архитектура, фактически представляющая собой универсальную оболочку баз знаний, которая без особых изменений может быть адаптирована для создания баз знаний в других предметных областях.

Литература

1. Гаврилова Т.А., Муромцев Д.И. Интеллектуальные технологии в менеджменте. СПб: Изд-во ВШМ СПбГУ, 2008.
2. Gavrilova T. Knowledge Mapping for Teaching and Learning // Int. Journal "The 21st Century: a scientific quarterly", Nr 2(24), Warsaw, Poland, 2007. – pp. 21–32.
3. Karrer T. What is e-Learning 2.0?, 2006 [электронный ресурс] – <http://elearningtech.blogspot.com/2006/02/what-is-elearning-20.html>, свободный.
4. Karrer T. Understanding eLearning 2.0, 2007 [электронный ресурс] – <http://www.learningcircuits.org/2007/07/07/karrer.html>, свободный.
5. Downes S. E-Learning 2.0., 2005 [электронный ресурс] – <http://www.downes.ca/post/31741>, свободный.
6. Karrer T. Corporate Long Tail Learning and Attention Crisis, 2008 [электронный ресурс] – <http://elearningtech.blogspot.com/2008/02/corporate-learning-long-tail-and.html>, свободный.
7. Горовой В.А., Муромцев Д.И. Реализация технологии активного обучения на базе онтологического моделирования // Научно-технический вестник СПбГУ ИТМО, №2(60)/2009. С. 107-114.

8. Музей оптики СПбГУ ИТМО [электронный ресурс] – <http://optimus.edu.ru/>.
9. Kasai, T., Yamaguchi, H., Nagano, K. and Mizoguchi, R. 'Building an ontology of IT education goals', Int. J. Continuing Engineering Education and Lifelong Learning, 2006 Vol. 16, Nos. 1/2, pp.1–17.
10. Gavrilova, T., Dicheva D., Sosnovsky S., Brusilovsky P. Ontological Web Portal for Educational Ontologies // In Proc. of Workshop "Applications of Semantic Web Technologies for E-Learning. (SW-EL'05)" in conjunction with 12th Int.Conf. on Artificial Intelligence in Education (AI-ED'05), Amsterdam, 2005. – pp. 19–29.
11. Gavrilova T. Knowledge Mapping for Teaching and Learning // Int. Journal "The 21st Century: a scientific quarterly", Nr 2(24), Warsaw, Poland, 2007. – pp. 21–32.
12. Toshinobu Kasai A1, Haruhisa Yamaguchi A2, Kazuo Nagano A3, Riichiro Mizoguchi A Semantic Web system for supporting teachers using ontology alignment A International Journal of Metadata, Semantics and Ontologies Volume 2, Number 1 / 2007 pp.35 – 44.
13. Linked Learning 2011: 1st International Workshop on eLearning Approaches for the Linked Data Age [электронный ресурс] – <http://projects.kmi.open.ac.uk/meducator/linkedlearning/>.
14. Бьюзен Т., Бьюзен Б. Супермышление. Минск: Белорусский дом печати. 2002.
15. Novak J. & Casas A. The Theory Underlying Concept Maps and How to Construct Them. Technical Report IHMC Map. 1, Florida Institute for Human and Machine. 2006.
16. Сравнение популярности редакторов онтологий. [электронный ресурс] – <http://ais-portal.ru/?p=161>.
17. Онтология DOLCE Lite. Laboratory for Applied Ontology LOA [электронный ресурс] – <http://www.loa-cnr.it/ontologies/DUL.owl>.
18. SpringGraph Flex Component [электронный ресурс] – <http://mark-shepherd.com/blog/springgraph-flex-component>.
19. HTML 4.01 Specification. [электронный ресурс] – <http://www.w3.org/TR/html4/>.
20. JSR-000245 JavaServer Pages 2.1 [электронный ресурс]– <http://jcp.org/aboutJava/communityprocess/final/jsr245/index.html>.

Муромцев Дмитрий Ильич. Доцент кафедры ПКС СПбГУ ИТМО. Окончил Санкт-Петербургский государственный политехнический университет в 1999 году. Кандидат технических наук, доцент. Автор 60 печатных работ. Область научных интересов: проектирование интеллектуальных систем, экспертные системы, онтологические базы знаний. E-mail: mouromtsev@mail.ifmo.ru.

Злобин Алексей Николаевич. Аспирант СПбГУ ИТМО. Окончил Санкт-Петербургский государственный электротехнический университет в 2009 году. Автор 4 печатных работ. Область научных интересов: СУЗ, модели знаний, semantic web. E-mail: alexey.zlobin@gmail.com.

Катков Юрий Валериевич. Аспирант СПбГУ ИТМО. Окончил Санкт-Петербургский государственный электротехнический университет в 2010 году. Автор 12 печатных работ. Область научных интересов: smart spaces, искусственный интеллект, e-learning, онтологии, semantic-web, инженерия знаний. E-mail: i.pochinok@gmail.com.

Починок Ирина Николаевна. Аспирантка СПбГУ ИТМО. Окончила Санкт-Петербургский государственный электротехнический университет в 2009 году. Автор 7 печатных работ. Область научных интересов: искусственный интеллект, онтологии, semantic-web, инженерия знаний. E-mail: i.pochinok@gmail.com.