

Д. И. МУРОМЦЕВ, В. А. ГОРОВОЙ, А. Н. ЗЛОБИН, Ю. В. КАТКОВ, И. Н. ПОЧИНОК

## АРХИТЕКТУРА СИСТЕМЫ УПРАВЛЕНИЯ ЗНАНИЯМИ НА ОСНОВЕ WIKI-ТЕХНОЛОГИИ И ИНТЕГРИРОВАННЫХ ОНТОЛОГИЧЕСКИХ МОДЕЛЕЙ

Описывается архитектура и рассматриваются ключевые концепции построения системы управления корпоративными знаниями на базе онтологического моделирования и технологии Wiki. В качестве примера реализации подобного рода технологии рассмотрен прототип системы “OntolingeWiki”.

*Ключевые слова:* управление знаниями, онтология, инженерия знаний, Wiki-технология.

**Введение.** Потребность в системах управления корпоративными знаниями существует в различных областях — от добывающей и перерабатывающей отраслей промышленности до сфер образования и культуры. Среди причин значительного роста интереса к системам управления знаниями можно выделить следующие [1]:

- поиск необходимой информации работниками предприятий сопряжен со значительными временными затратами;
- опыт ведущих и наиболее квалифицированных сотрудников используется лишь в отдельных подразделениях предприятия и не распространяется на корпорацию в целом;
- ценная информация сокрыта в большом количестве документов и данных, доступ к которым затруднен;
- дорогостоящие ошибки повторяются вследствие недостаточной информированности сотрудников и игнорирования ими предыдущего опыта.

В настоящей статье описывается архитектура и рассматриваются ключевые концепции построения системы управления корпоративными знаниями на базе онтологического моделирования и технологии Wiki. В качестве примера реализации подобного рода технологии рассмотрен прототип системы “OntolingeWiki”.

Основная цель описываемой разработки — создание прототипа интегрированной среды накопления знаний [2]. Эта среда призвана объединить методы и средства формального управления знаниями (онтологического инжиниринга [3, 4]) и неформальный подход к структурированию информации (контента открытого Wiki-портала [5]). Использование онтологий в качестве основы для портала знаний имеет два преимущества: формирование хорошо структурированного портала, отражающего предметную область, и возможность предоставления данных, которые могут обрабатываться программными агентами. В последнее время тенденции к открытию хранилищ структурированных данных, представляемых в различных форматах, таких как RDF и OWL, обрели форму движения Linking Open Data [6]. В рамках этого движения разработчики предоставляют доступ к своим „машиночитаемым“ данным для всех желающих, что позволяет осуществлять различные совместные проекты с использованием

разнообразных источников данных. Безусловно, наличие удобных инструментов, позволяющих формировать порталы на основе онтологий, будет способствовать развитию движения Linking Open Data. Одним из таких инструментов и является система “OntolingWiki”.

**Анализ требований к системе.** Первая версия системы “OntolingWiki” [7] была предназначена для организации управления знаниями при разработке образовательного портала музея оптических технологий Санкт-Петербургского государственного университета информационных технологий, механики и оптики. Прежде всего были определены основные элементы — акторы (в терминах методологии Rational Unified Process), взаимодействующие с системой (применялись шаблоны определения акторов и варианты использования системы, предложенные в работе [8]). В таблице приведено краткое описание действий акторов, анализ поведения которых учитывался при формировании функциональных требований к системе.

Актор	Действие
Пользователь-школьник	Взаимодействует с системой в познавательных целях, используя минимальный набор функций
Пользователь-студент	Взаимодействует с системой в учебных и исследовательских целях
Пользователь-преподаватель	Использует систему для проведения лекций обычно по заранее известному сценарию
Пользователь-эксперт	Предоставляет знания для системы
Администратор	Осуществляет установку/обновление/восстановление системы
Браузер	Обеспечивает доступ к системе для всех групп пользователей
Инженер по знаниям	Производит оценку онтологии, отображенной средствами системы, по критериям когнитивной эргономичности, полноты и непротиворечивости

Следующим этапом проектирования стало определение процессов, наличие которых является обязательным для большинства систем управления знаниями. На рис. 1 представлена схема, отображающая основные этапы жизненного цикла контента и знаний внутри системы.

Последний этап формализации требований к системе заключался в создании сценария работы с порталом знаний. Основными стадиями этого этапа стали следующие:

- разработка инженером по знаниям с помощью экспертов онтологии предметной области портала;
- автоматическое создание системой структуры портала (названий/имен разделов и документов, формирование ссылок между ними);
- разработка контента портала в Wiki-стиле экспертами предметной области (к разработке также допускаются преподаватели по смежным областям);
- перевод портала в режим „только для чтения“; контент портала и графическое представление понятий системы могут служить учебным и демонстрационным материалом для пользователей-школьников и пользователей-студентов.

**Функции системы.** В соответствии с анализом задач, поставленных перед системой, а также ранее полученными результатами [2] были выделены следующие основные функции, объединенные в три группы.

*Ввод контента и наполнение базы знаний:*

- импорт одной или нескольких онтологий в качестве шаблона структуры портала;
- создание, редактирование и хранение тематических статей (контента портала).

*Управление контентом:*

- поддержка нескольких разделов, описываемых разными онтологиями;
- отображение тематической статьи;
- создание системы ссылок между тематическими статьями на основе представляемых ими концептов онтологии и связей между ними;
- вывод результатов поисковых запросов по тематической базе данных Интернет-ресурсов;

— отображение информации о концепте, описываемом определенной статьей.

*Представление структуры знаний:*

— визуализация структуры знаний, представленных в портале, в виде дерева отношений наследования и отношений класс—индивид между концептами онтологии;

— визуализация множества отношений между концептами в виде „облака тегов“ (набора ключевых слов);

— визуализация множества близких к определенной тематической статье концептов в форме графа отношений.

Важно отметить, что реализация всех пунктов последней группы подразумевает интерактивную визуализацию: т.е. их введение в портал направлено в первую очередь на поддержку навигации по portalу и его структуре, а не на поддержку обзора формальной модели, лежащей в его основе.



Рис. 1

**Классификация и форматы данных.** Обрабатываемые в системе данные можно разделить на три основных класса:

- 1) формальная модель знаний в виде онтологии;
- 2) контент портала;
- 3) онтологическая структура портала, представляемая в виде направленного графа отношений между статьями.

Логика работы системы управления порталом знаний заключается в преобразовании данных первого и второго классов в данные третьего класса. Важно отметить, что данные имеют множество видов представления.

*Модель знаний.* Для описания формальных моделей знаний применяются специализированные языки онтологий. В системе “OntolingWiki” используется основанный на XML язык OWL — Web Ontology Language [9], что обусловлено:

- распространённостью языка в области web-онтологий;

— поддержкой со стороны консорциума W3C — организации, решающей формальные вопросы развития Интернета (в частности, поиска новых путей развития и внедрения новых технологий);

— широкой поддержкой, осуществляемой различными инструментальными средствами разработки онтологий.

**Контент портала.** Контент портала представляет собой форматированный текст, содержащий гиперссылки и дополнительные мультимедиа-материалы. Для описания статей все Wiki-инструменты предоставляют свои, сравнительно простые языки разметки, используемые системе в качестве формата хранения контента.

**Онтологическая структура портала.** По сути, этот класс данных формируется динамически на основе хранимой онтологии предметной области путем исключения не используемых в выбранном контексте понятий. Так как алгоритм получения ссылок между элементами онтологии и Wiki-страницами зависит только от выбранной онтологии и является вполне детерминированным, хранение данных этого класса оказывается избыточным: они формируются в процессе обработки запросов и используются при поиске или создании страниц портала.

**Логическая структура системы.** Сложность и разнородность задач, решаемых системой, повлекла за собой значительные трудности на этапе построения и формализации структуры программных средств. На верхнем уровне система разделена на модули в соответствии с выполняемыми функциями. Данное разделение обусловлено желанием максимально использовать для решения некоторых задач готовые программные средства с минимальной модификацией.

Схема, отображающая основные компоненты системы “OntolingeWiki” и потоки данных между ними и пользователями, представлена на рис. 2.



Рис. 2

**Wiki-система.** Этот компонент системы “OntolingeWiki” представляет собой стандартный Wiki-инструмент с базовым для таких средств набором функций. Была выбрана Wiki-система с открытым исходным кодом DokuWiki [10]; основные причины такого выбора:

- большое количество свободно распространяемых подключаемых модулей;
- значительный опыт разработчиков;
- простой и переносимый формат хранения контента.

Как было отмечено выше, Wiki-разметка имеет довольно простой формат. Приведем несколько примеров разметки текста в DokuWiki:

**Заголовки:**

```
===== Заголовок 1-го уровня =====
```

```
...
```

```
== Заголовок 5-го уровня ==
```

**Форматирование:**

```
**жирный**, //курсив//, __подчеркнутый__ текст
```

**Оформление ссылок:**

```
[[http://google.com|Google]] — внешние
```

```
[[optics:personal:newton|Ньютон]] — внутренние
```

В современных Wiki-системах приняты два основных способа группировки контента: по категориям или по пространствам имен. В DokuWiki доминирующим средством являются пространства имен, но также поддерживается способ группировки данных по тэгам. Поддержка в DokuWiki подключаемых модулей (плагинов) обеспечивает такие преимущества, как поддержка типов содержимого Wiki-портала, расширение функциональности системы без необходимости перехода на новую версию ядра.

Wiki-система потребовала минимальной доработки, которая заключалась в создании версии, функционирующей в режиме „только контент“, без использования дополнительных элементов редактирования и навигации по portalу.

*Средство разработки онтологий.* Этот инструмент используется инженерами по знаниям в процессе создания онтологии предметной области. В соответствии с предложенной архитектурой системы единственная связь данного ее компонента с остальными может осуществляться через файлы формата OWL. В качестве данного инструмента был выбран свободно распространяемый редактор онтологий Protégé [11], разрабатываемый группой сотрудников Стэнфордского университета (США) при поддержке большого сообщества волонтеров. Пригодность Protégé для редактирования онтологий определяется следующими возможностями:

- стабильностью и поддержкой всех средств текущей версии языка OWL;
- распространением по лицензии GPL;
- несколькими видами визуализации онтологий;
- интуитивно понятным интерфейсом пользователя и наличием большого количества документации;
- поддержкой подключаемых модулей.

*Портал.* Основная идея, заложенная в функционирование данного блока, — загрузка „декораций“ для Wiki-контента, таких как:

- представление общей структуры знаний и формализованных знаний о концепте;
- отображение близких и связанных концептов;
- интеграция поисковых систем в портал.

Данный блок реализован как web-модуль на языке Java [12]. На уровне web-приложения можно выделить несколько компонентов: аналитический модуль, модуль визуализации и модуль декорирования.

Аналитический модуль отвечает за формирование структуры портала на основе онтологии, в том числе определение места хранения (адреса) фрагмента контента, связанного с определенным концептом. Этот модуль осуществляет первичную обработку OWL-документа (непосредственно для работы с этим форматом и получения модели онтологии используется библиотека Jena [13]). В результате создается ряд промежуточных представлений онтологии, которые используются остальными модулями приложения.

Модуль визуализации структуры знаний состоит из двух частей. Первая — апплет (англ. applet, букв. от *application* — приложение — и *let* — суффикс: ≈ программный компонент) HyperGraph [14] — загружается браузером клиента и осуществляет визуализацию структуры портала в виде графа в гиперболическом пространстве [15]. Данный апплет поддерживает отображение надписей и изображений в узлах графа, а также встраивание в них

гиперссылки. Таким образом, он полностью обеспечивает возможность интерактивной навигации по структуре портала. Всевозможные настройки внешнего вида портала (цвета, шрифты и т.п.) также поддерживаются этим модулем. Вторая часть модуля визуализации — собственно модуль приложения — позволяет выполнять трансляцию онтологических описаний в формат данных, совместимый с представлением в виде графа.

Модуль „декорирования“ контента представляет собой набор динамических страниц (JSP), объединяющих следующие элементы отображения: меню навигации, область отображения Wiki-контента, описание структуры онтологии, описание отдельных концептов, результаты поисковых запросов, сформированных по имени концепта. С использованием результатов анализа онтологии и сформированной структуры портала модуль „декорирования“ поддерживает пользовательскую сессию посредством запоминания текущей онтологии, выбранного концепта и режима отображения. На основе этих данных отображается общая структура онтологии, статья о выбранном концепте, его формальное описание, результаты поиска по соответствующему запросу в Интернете.

**Интерфейс пользователя.** Пример интерфейса пользователя системы “OntolingWiki” приведен на рис. 3.

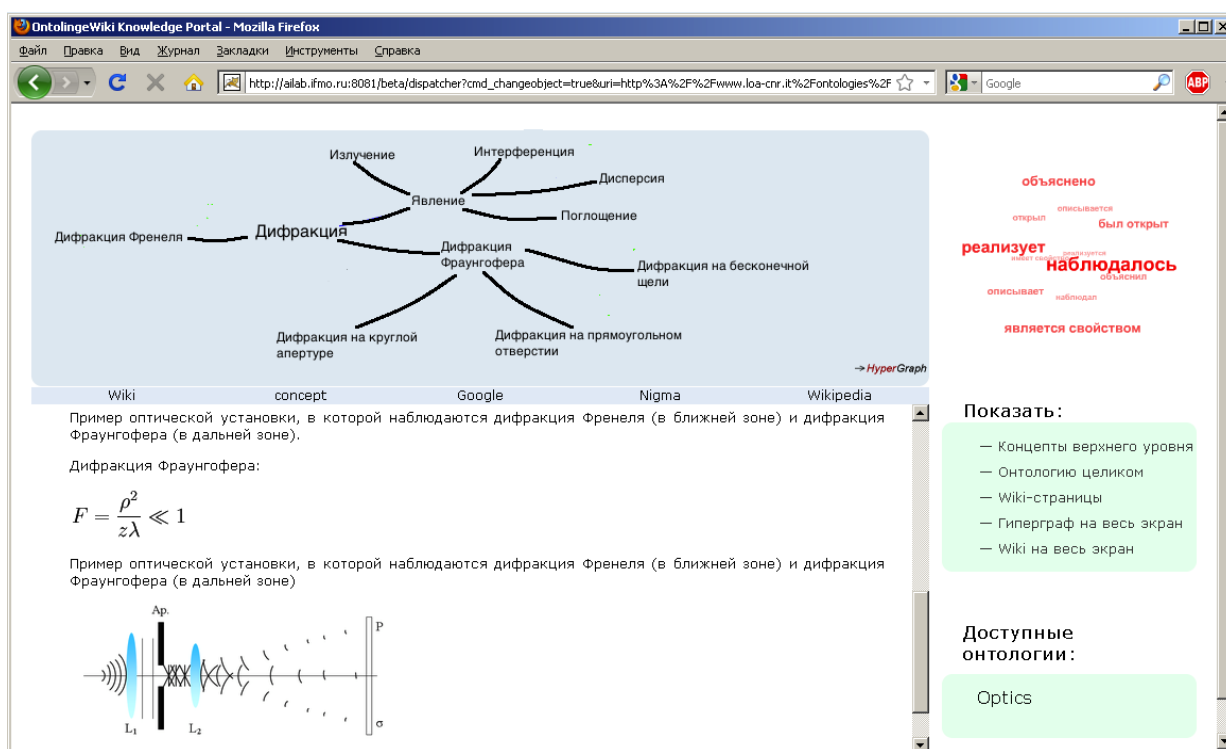


Рис. 3

Для организации данных портала используется блочная верстка. В верхнем блоке отображены средства визуализации структуры портала. Первое средство — это граф концептов, отображающий отношения наследования между классами и отношения принадлежности между классами и индивидами. Узлы графа являются ссылками на соответствующие Wiki-страницы (открываются также внутри портала). Второе средство — это „облако“ типов связей. Входящие в него элементы являются ссылками на страницы с формальным описанием типа связи.

В блоке, расположенном справа, находится меню портала. В нем содержатся основные команды навигации (перейти к главной странице раздела, перейти к описанию онтологии, сменить раздел, перейти к разделу помощи) и команды управления режимом просмотра (показывать только граф, показывать только статью).

В основном блоке в зависимости от режима просмотра могут быть отображены: общие сведения о разделе, описание онтологии или описание конкретного концепта. Описание концепта может быть осуществлено в нескольких режимах:

— формальное описание — в этом режиме отображается информация из онтологии: имя, связи и т.д.;

— Wiki-статья — в этом режиме отображается соответствующая Wiki-страница;

— результаты поиска — здесь отображаются результаты тематического поиска по заранее подготовленному списку источников информации (используется Google Custom Search [16]).

**Заключение.** Возможность доступа к „машиночитаемым“ структурированным данным для всех желающих, обозначившаяся в форме движения Linking Open Data, демонстрирует важность и актуальность интеграции разнородных источников информации в единое семантическое пространство — базу знаний или корпоративную память. Очевидно, что подобное объединение информационных ресурсов позволит поднять многие проекты на качественно новый уровень за счет использования различных источников данных и знаний. Однако для работы в этом новом информационном пространстве требуется построение специализированных систем управления знаниями аналогично тому, как в свое время системы управления данными позволили совершить качественный скачок в работе с базами данных.

Рассмотренные в настоящей статье архитектура и ключевые концепции построения систем управления корпоративными знаниями на базе онтологического моделирования и технологии Wiki были успешно применены на практике при создании прототипа интегрированной среды накопления знаний “OntolingеWiki”. Показано, что успеха при разработке подобного класса систем можно добиться за счет объединения методов и средств формального управления знаниями, включая онтологический инжиниринг, и неформального подхода к структурированию контента, например, на основе открытого Wiki-портала. Система “OntolingеWiki” была успешно реализована и опробована при разработке справочно-обучающего портала по оптике для музея оптических технологий СПбГУ ИТМО [17].

#### СПИСОК ЛИТЕРАТУРЫ

1. Гаврилова Т. А., Муромцев Д. И. Интеллектуальные технологии в менеджменте. СПб: Изд-во ВШМ СПбГУ, 2007.
2. Гаврилова Т. А., Горовой В. А., Злобин А. Н. и др. Интеграция Wiki-технологии и онтологического моделирования в задаче управления знаниями предприятия // Тр. 11-й Нац. конф. по искусственному интеллекту „КИИ-2008“. Дубна, 2008. Т. 2. С. 360—368.
3. Gruber T. R. A translation approach to portable ontologies // Knowledge Acquisition. 1993. N 5 (2). P. 199—220.
4. Guarino N., Giaretta P. Ontologies and Knowledge Bases: Towards a Terminological Clarification. Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing. 1995. IOS Press: 25—32 [Электронный ресурс]: <<http://www.loa-cnr.it/Papers/KBKS95.pdf>>.
5. The Wiki Principle. The Economist, 2006 [Электронный ресурс, англ.]: <[http://www.economist.com/surveys/displaystory.cfm?story\\_id=6794228](http://www.economist.com/surveys/displaystory.cfm?story_id=6794228)>.
6. Linking Open Data, 2008 [Электронный ресурс, англ.]: <<http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData/>>.
7. Горовой В. А., Муромцев Д. И. Реализация технологии активного обучения на базе онтологического моделирования // Науч.-техн. вестн. СПбГУ ИТМО. 2009. № 2(60). С. 107—114.
8. RUP Glossary 7.2.0. IBM Rational Corp., 2010 [Электронный ресурс, англ.]: <<http://www-01.ibm.com/software/awdtools/rup/>>.
9. OWL — Web Ontology Language. Overview, 2004 [Электронный ресурс, англ.]: <<http://www.w3.org/TR/owl-features/>>.
10. DokuWiki, 2010 [Электронный ресурс, англ.]: <<http://www.dokuwiki.org/dokuwiki>>.

11. Protégé , 2009 [Электронный ресурс, англ.]: <<http://protege.stanford.edu/>>.
12. Java EE Web Application Technologies, 2010 [Электронный ресурс, англ.]: <<http://java.sun.com/javaee/technologies/webapps/index.jsp>>.
13. Jena – A Semantic Web Framework for Java, 2010 [Электронный ресурс, англ.]: <<http://jena.sourceforge.net/>>.
14. The HyperGraph Homepage, 2003 [Электронный ресурс, англ.]: <<http://hypergraph.sourceforge.net/>>.
15. *Munzner T.* 3D Hyperbolic Quasi-Hierarchical Graphs., 2003 [Электронный ресурс, англ.]: <[http://graphics.stanford.edu/papers/munzner\\_thesis/html/node8.html](http://graphics.stanford.edu/papers/munzner_thesis/html/node8.html)>.
16. Система пользовательского поиска, 2010 [Электронный ресурс]: <<http://www.google.com/cse/>>.
17. *Муромцев Д. И., Баландин Е. А., Катков Ю. В., Починок И. Н.* Опыт использования онтологий верхнего уровня при проектировании базы знаний музея оптических технологий // Материалы Всерос. конф. „Знания— Онтологии — Теории“ (ЗОНТ-09). Новосибирск, 2009. Т. 1. С. 165—172.

**Сведения об авторах**

- Дмитрий Ильич Муромцев** — канд. техн. наук, доцент; Санкт-Петербургский государственный университет информационных технологий, механики и оптики, кафедра проектирования компьютерных систем; E-mail: [mouromtsev@mail.ifmo.ru](mailto:mouromtsev@mail.ifmo.ru)
- Владимир Андреевич Горовой** — Высшая школа менеджмента Санкт-Петербургского государственного университета, кафедра информационных технологий в менеджменте; ассистент; E-mail: [a.malinin@gmail.com](mailto:a.malinin@gmail.com)
- Алексей Николаевич Злобин** — аспирант; Санкт-Петербургский государственный университет информационных технологий, механики и оптики, кафедра проектирования компьютерных систем; E-mail: [alexey.zlobin@gmail.com](mailto:alexey.zlobin@gmail.com)
- Юрий Валериевич Катков** — аспирант; Санкт-Петербургский государственный университет информационных технологий, механики и оптики, кафедра проектирования компьютерных систем; E-mail: [ganqurgon@gmail.com](mailto:ganqurgon@gmail.com)
- Ирина Николаевна Починок** — аспирант; Санкт-Петербургский государственный университет информационных технологий, механики и оптики, кафедра проектирования компьютерных систем; E-mail: [i.pochinok@gmail.com](mailto:i.pochinok@gmail.com)

Рекомендована кафедрой  
проектирования компьютерных систем СПбГУ ИТМО

Поступила в редакцию  
08.02.10 г.