

УДК 004.822

Автоматическая генерация структуры Wiki-систем при помощи онтологий

В.К. Шестаков

Институт систем информатики имени А.П. Ершова СО РАН
shestakov@iis.nsk.su

Шестаков В.К. Автоматическая генерация структуры Wiki-систем при помощи онтологий.
В статье описываются инструментальные средства, поддерживающие технологию автоматической генерации структуры систем на основе Wiki-технологии при помощи онтологий предметных областей. Технология позволяет получать Wiki-системы с хорошей структурой и согласованной системой понятий.

Ключевые слова: онтология, Wiki, семантические технологии, генерация структуры информационных систем.

Введение

Для удовлетворения все возрастающих информационных потребностей пользователей разработаны разнообразные средства построения информационных систем. Одним из таких удобных и простых в использовании средств сбора и хранения информации являются Wiki-системы [1]. Они позволяют работать не только с текстовым, но и с мультимедийным контентом, имеют удобный и интуитивно понятный интерфейс, просты в освоении. Однако их большим недостатком является то, что они позволяют отслеживать в создаваемых информационных системах только структурную целостность ссылок, не обеспечивая при этом логической непротиворечивости и семантической согласованности [2] используемых в них понятий и отношений. Общая идея предлагаемого подхода состоит в создании инструментария, который бы обеспечивал автоматическую генерацию структуры Wiki-систем с согласованной системой понятий (семантически согласованных Wiki). Wiki с такими свойствами можно получить, если генерировать ее на основе логически непротиворечивой онтологии, описывающей предметную область будущей системы.

Обзор средств создания информационных Wiki-систем, использующих семантические технологии

На сегодняшний день существует достаточно много проектов создания информационных Wiki-систем, так или иначе связанных с семантическими технологиями. Хотя эти проекты нацелены на решение различных задач, общим для них является использование онтологий на тех или иных этапах построения или

функционирования информационных систем. Так как на данный момент существует большое количество средств создания Wiki-систем, то в этом обзоре мы сосредоточим наше внимание только на тех из них, которые в той или иной мере используют семантические технологии.

Сначала рассмотрим проект из Санкт-Петербурга [3], в котором в одном из первых начали применяться онтологии при построении таких информационных систем, как корпоративные Wiki-порталы. Исходя из того, что онтологии позволяют создавать более концептуально-ясные и целостные модели предметных областей, разработчиками было решено привязать контент Wiki-системы к онтологии, чтобы избежать проблем с плохой структурой портала, т.к. она в этом случае будет такого же качества, как и определяющая ее онтология. В рамках этого проекта был разработан прототип инструментария OntolingWiki, который содержит внутри себя Wiki-движок и для каждого концепта онтологии автоматически создает Wiki-страницу, в которую эксперт вносит информацию об описываемом понятии. При этом конечные пользователи видят онтологию, представленную в виде графа, могут осуществлять навигацию по ее понятиям и видеть содержимое Wiki-страниц, соответствующих выбранному понятию.

Недостатком OntolingWiki является то, что онтология, лежащая в основе созданной с помощью нее Wiki-системы, не может быть изменена или дополнена средствами самой Wiki-системы.

Наиболее распространенным инструментальным средством для создания Wiki-систем является популярный Wiki-движок MediaWiki [4]. Именно для этого универсального движка больше всего строится расширений, позволяющих использовать средства семантических технологий. Одной из первых была создана надстройка над MediaWiki, получившая название Semantic MediaWiki [5]. Это

расширение допускает возможность импорта онтологии [6] и позволяет использовать в Wiki-системе структуры и связи, определенные в ней. Однако предоставляемый Semantic MediaWiki модуль импорта онтологий на данный момент находится в стадии бета-версии и не включен в состав новых версий данного расширения. В связи с этим реализацию импорта онтологий пользователи Semantic MediaWiki должны выполнять самостоятельно.

Другим расширением MediaWiki является BOWiki [7] — основанный на онтологии семантический Wiki-движок. Первоначально он базировался на Semantic MediaWiki, а потом был реализован как «чистое» расширение MediaWiki. Достоинством BOWiki является то, что он позволяет пользователям рассматривать сущности, описанные Wiki-страницами, в качестве экземпляров онтологических категорий, определять новые отношения в Wiki-контенте, связывать Wiki-страницы, в том числе *n*-арными семантическими отношениями, запрашивать Wiki-страницы, удовлетворяющие определенным условиям, допускает импорт еще нескольких био-онтологий для ограниченного использования и обеспечивает экспорт Wiki-контента в OWL-представление.

BOWiki разрабатывался как средство для совместного создания и интеграции знаний в области био-информатики, поэтому его функционирование основано на онтологии высокого уровня General Formal Ontology (GFO) [8] в OWL-версии. В своей работе он использует также две онтологии, основанные на GFO, — базовую биомедицинскую онтологию (GFO-Bio) и онтологию функций (Ontology of Functions или OF). Как было сказано выше, возможность использования других (внешних) онтологий существенно ограничена, поэтому BOWiki не может рассматриваться в качестве инструмента для построения Wiki-систем произвольного вида.

Также для MediaWiki существуют расширения, работающие совместно с Semantic MediaWiki, и добавляющие дополнительную функциональность по импорту семантической информации. Одним из таких является RDFIO [9]. Оно позволяет импортировать произвольные RDF-триплеты [10]. Но это расширение до сих пор имеет статус бета-версии и еще не готово для массового использования. Еще одним подобным расширением является LinkedWiki [11]. Оно применяется для получения данных с внешнего сервиса и отображения этих данных в Wiki-системе. Это расширение не может изменить структуру существующей системы, а только дополняет ее, поэтому имеет довольно ограниченную область применения.

Основные части Wiki-систем и инструменты для работы с ними

В основе любой Wiki-системы лежит так называемый Wiki-движок — комплекс программных средств для преобразования Wiki-разметки в код, предназначенный для отображения в браузере. Одним из самых распространенных движков является MediaWiki (на его основе работает широко известная Википедия). Для него существуют специальные дополнения, называемые расширениями, позволяющие получить определенную функциональность. Так, расширение Semantic MediaWiki дает возможность добавлять семантическую информацию за счет расширения разметки, а также предлагает средства для работы с этой информацией.

Для проведения работ технического характера в Wiki-системах используют ботов — специальные программы для выполнения заданного набора операций. Они являются клиентскими приложениями, поэтому не требуют внесения изменений на стороне сервера (т.е. для их работы не нужно модифицировать движок или ставить какое-либо расширение). Например, в Википедии ботов используют для таких задач, как переименование категорий и статей, расстановка интервики ссылок (ссылок на родственные проекты), исправление ссылок, удаление спама и т.п. Для реализации ботов используются различные языки программирования, а также существуют разные библиотеки для облегчения их написания. Одной из наиболее развитых библиотек является Python WikipediaBot Framework [12]. Она использует MediaWiki API (специальный интерфейс прикладного программирования) для взаимодействия с MediaWiki-системой: авторизации, получения данных и внесения изменений.

Автоматическая генерация структуры Wiki-системы при помощи онтологий: описание работы системы

Предлагаемая в рамках данной работы инструментальная система позволяет на основе ранее разработанной онтологии автоматически сгенерировать Wiki-систему, структура и содержание которой будет определяться этой онтологией. В дальнейшем построенная Wiki-система может расширяться как обычная Wiki-система с использованием традиционных средств Wiki-технологии.

Общая схема функционирования системы, реализованной в рамках предлагаемого подхода, объединяющего Wiki-технологии и онтологии, выглядит следующим образом (см. рис. 1). На первом этапе квалифицированный эксперт описывает онтологию предметной области. На втором при помощи разработанного

автором инструмента (Onto2Wiki) по ней строится интерфейс на основе Wiki-технологии, состоящий из двух частей: конструкторского и пользовательского. При помощи первого эксперты предметной области могут вносить данные в систему, которые тут же становятся

На первом шаге онтология, полученная в готовом виде или разработанная в Protégé [13] или каком-либо другом редакторе онтологий, сохраняется или конвертируется в OWL-формат [14]. После этого файл со спецификацией онтологии подается на вход специально

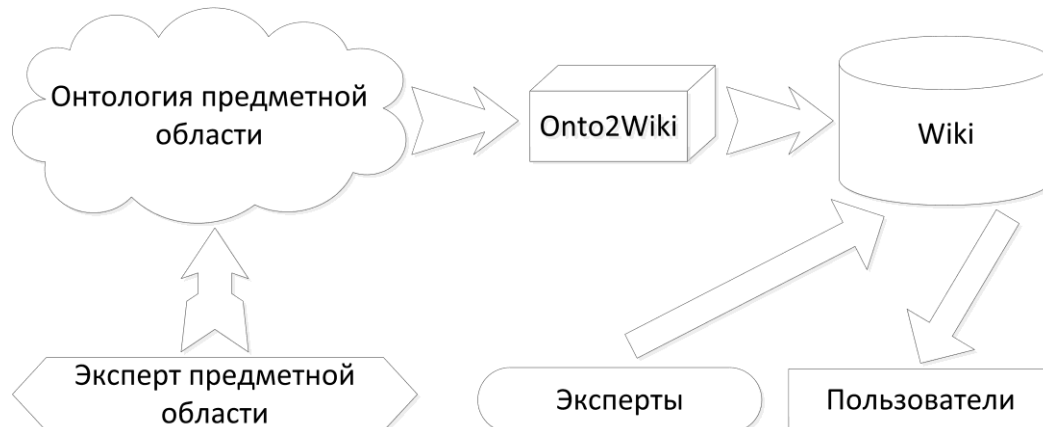
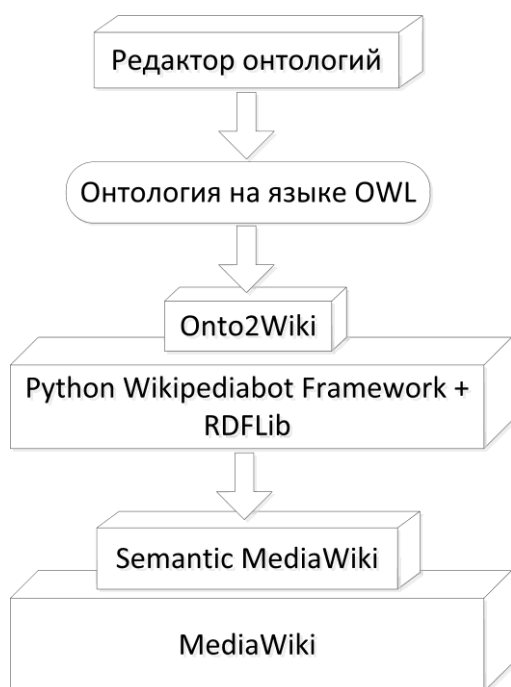


Рисунок 1 – Общая схема функционирования системы

доступны пользователям посредством второго.

Подробная схема функционирования инструментальной системы представлена на рис. 2. Согласно ей разработка Wiki-системы в рассматриваемой инструментальной системе включает следующие этапы.



5 Рисунок 2 – Подробная схема функционирования системы

разработанному в рамках данного проекта программному модулю Onto2Wiki, который разбирает его с использованием библиотеки RDFLib [15], а затем, при помощи среды Python WikipediaBot Framework, создает каркас информационной системы на основе пустого Wiki-сайта, работающего на базе MediaWiki с расширением Semantic MediaWiki. При этом в Wiki-систему добавляются нужные страницы, для которых указываются соответствующие атрибуты, расставляются категории и прописываются нужные связи. После выполнения этих действий Wiki-система готова к использованию.

В таблице 1 представлено соответствие между конструкциями языка OWL и семантической Wiki, которое используется при отображении онтологии в структуры и содержание Wiki-системы.

Таблица 1. Соответствие конструкций языка OWL и Semantic MediaWiki

Конструкция OWL	Semantic MediaWiki
Класс	Категория
Подкласс	Подкатегория
Экземпляр	Страница
Атрибут	Атрибут
Отношение	Типизированная ссылка

В соответствии с этой таблицей для каждого класса онтологии будет создана своя категория, которая будет проставлена на каждой подкатегории, соответствующей подклассу данного класса, и на каждой странице,

соответствующей экземпляру данного класса. Также все атрибуты экземпляра онтологии будут указаны в качестве атрибутов на странице, соответствующей этому экземпляру. При этом все отношения онтологии отобразятся в типизированные ссылки [16] между соответствующими страницами и категориями.

Также при отображении онтологии ее типы данных преобразуются в типы данных Semantic MediaWiki. Пример соответствия некоторых типов данных представлен в таблице 2.

Таблица 2. Пример соответствия типов данных при отображении

Тип OWL	Тип Semantic MediaWiki
string	Строка
double	Число
boolean	Булево
dateTime	Дата

Стоит заметить, что созданную с помощью данного инструмента систему можно дорабатывать в соответствии с возникающими потребностями, взаимодействуя с ней как с обычной Wiki-системой. Т.е. можно создавать

система получается полностью независимой от той онтологии, на основе которой была построена.

Демонстрация работы системы

Рассмотрим работу данной инструментальной системы на примере построения учебной информационной системы «Технология создания экспертных систем». Онтология для этой системы была построена в уже упоминавшемся редакторе онтологий Protégé. Ее визуализация представлена на рис. 3.

Вот фрагмент этой онтологии в формате Turtle:

```
ex:Как_построить_свою_ЭС      rdf:type
ex:Литература_по_ЭС ,
```

```
owl:NamedIndividual ;
    ex:название_книги
"Как построить свою экспертную
систему" ;
    ex:автор_книги
"Нейлор К." ;
    ex:упоминается
ex:MYCIN ,
```

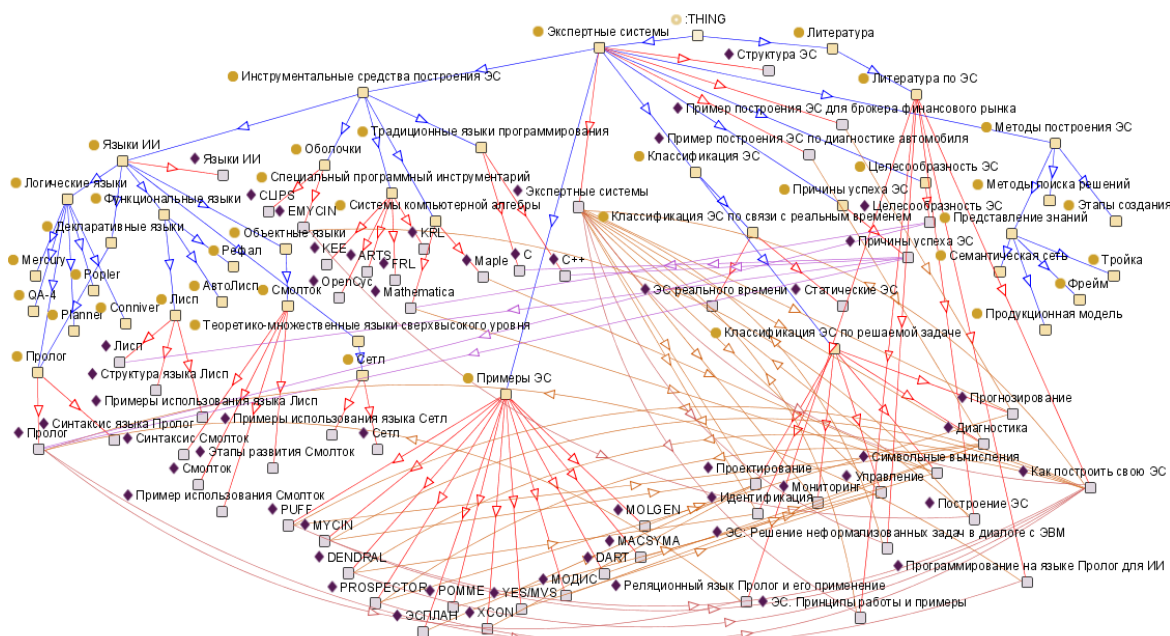


Рисунок 3 – Онтология «Технология создания экспертных систем».

новые страницы, прописывать новые связи, указывать новые категории и т.д.

Таким образом, особенностью данного подхода является то, что полученная Wiki-

ex:DENDRAL ,

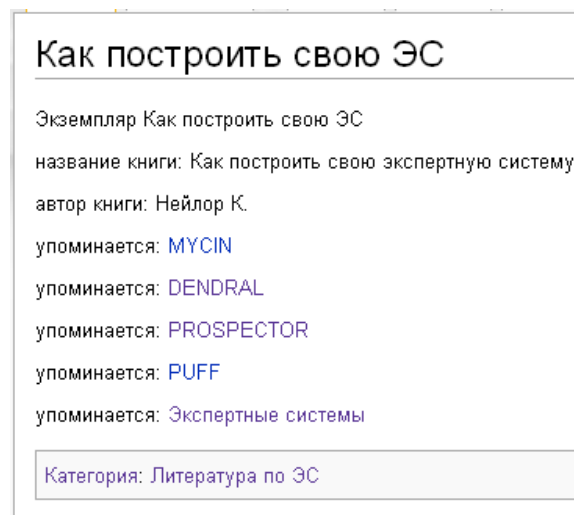
ex:PROSPECTOR ,

ex:PUFF ,

ex:Экспертные_системы .

Подаем эту онтологию на вход системе и после окончания ее работы получаем заполненную Wiki-систему.

На рис. 4 представлен пример одной из страниц полученной системы. На нем видны атрибуты и типизированные ссылки, находящиеся на странице.



6

7 Рисунок 4 – Одна из страниц

А вот так выглядит Wiki-разметка данной страницы:

Экземпляр Как построить свою ЭС

название книги: [[название книги::Как построить свою экспертную систему]]

автор книги: [[автор книги::Нейлор К.]]

упоминается: [[упоминается::MYCIN]]

упоминается:
[[упоминается::DENDRAL]]

упоминается:
[[упоминается::PROSPECTOR]]

упоминается: [[упоминается::PUFF]]

упоминается:
[[упоминается::Экспертные системы]]

[[Категория:Литература по ЭС]]

Заключение

В данной работе рассмотрен способ автоматической генерации структуры систем на основе Wiki-технологии при помощи онтологий предметных областей. В рамках этого подхода предложен метод автоматической генерации структуры Wiki-систем при помощи онтологий, разработан прототип инструментальной системы, реализующий метод в объеме, указанном в четвертом разделе.

Применение метода автоматической генерации при помощи онтологий позволяет получать Wiki-системы с хорошей структурой и согласованной системой понятий.

Литература

1. Leuf B., Cunningham W. The Wiki Way: Quick Collaboration on the Web. Addison-Wesley Professional, 2001. 435 p.
2. Baader F., Nutt W. Basic Description Logics. In the Description Logic Handbook, edited by Baader F., Calvanese D., McGuinness D.L., Nardi D., Patel-Schneider P.F., Cambridge University Press, 2002.
3. Муромцев Д.И., Горовой В.А., Малинин А.А., Гаврилова Т.А., Злобин А.Н., Катков Ю.В. Интеграция wiki-технологии и онтологического моделирования в задаче управления знаниями предприятия // Труды 11-ой национальной конференции по искусственному интеллекту с международным участием КИИ-2008 (г. Дубна, Россия). — М.: ЛЕНАНД, 2008. — Т.3. — С. 360-368.
4. Режим доступа: <http://mediawiki.org> — MediaWiki.
5. Режим доступа: <http://semantic-mediawiki.org> — Semantic MediaWiki.
6. Режим доступа: http://semantic-mediawiki.org/wiki/Help:Ontology_import — описание импорта в Semantic MediaWiki.
7. Backhaus M., Kelso J., Bacher J., Herre H., Hoehndorf R., Loebe F., Visagie J. BOWiki – a collaborative annotation and ontology curation framework // In Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge, CKC Banff, Canada, May 8, of CEUR Workshop Proceedings. Volume 273. Edited by Noy N., Alani H., Stumme G., Mika P., Sure Y.,

Vrandecic D. Aachen, Germany: CEUR-WS.org, 2007.

8. Herre H., Heller B., Burek P., Hoehndorf R., Loebe F., Michalek H. General Formal Ontology (GFO) – A foundational ontology integrating objects and processes [Version 1.0] // Onto-Med Report 8, Research Group Ontologies in Medicine, Institute of Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig (2006).

9. Режим доступа: <http://www.mediawiki.org/wiki/Extension:RDFIO> — RDFIO.

10. Режим доступа: <http://www.w3.org/RDF/> — RDF.

11. Режим доступа: <http://www.mediawiki.org/wiki/Extension:LinkedWiki> — LinkedWiki.

12. Режим доступа: <http://pywikipediabot.sourceforge.net> —

Python WikipediaBot Framework.

13. Режим доступа: <http://protege.stanford.edu> — Protégé.

14. Motik B., Patel-Schneider P.F., Parsia B., eds. OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax // W3C Recommendation, 27 October 2009, <http://www.w3.org/TR/2009/REC-owl2-syntax-20091027/>.

15. Режим доступа: <http://rdflib.net> — RDFLib.

16. Völkel M., Krötzsch M., Vrandecic D., Haller H. and Studer R. Semantic Wikipedia // In Proceedings of the 15th International Conference on World Wide Web (Edinburgh, Scotland, May 23–26, 2006). WWW '06. ACM Press, New York, NY, pp. 585–594, 2006, <<http://doi.acm.org/10.1145/1135777.1135863>>.

Шестаков В.К. Автоматическая генерация структуры Wiki-систем при помощи онтологий. В статье описываются инструментальные средства, поддерживающие технологию автоматической генерации структуры систем на основе Wiki-технологии при помощи онтологий предметных областей. Технология позволяет получать Wiki-системы с хорошей структурой и согласованной системой понятий.

Ключевые слова: онтология, Wiki, семантические технологии, генерация структуры информационных систем.

Shestakov V.K. Automatic ontology-based generation of Wiki-system structure. The paper presents tools supporting technology of an automatic generation of structure of systems based on Wiki-technology using subject domain ontology. The technology allows to create Wiki-systems with a good structure and a consistent system of concepts.

Keywords: ontology, Wiki, semantic technology, generation structure of information system.

Статья поступила в редакцию 20.11.2016
Рекомендована к публикации д-ром физ.-мат. наук А.С. Миненко