

ПОСТРОЕНИЕ ПЕРСОНИФИЦИРОВАННЫХ ОНТОЛОГИЧЕСКИХ БАЗ ЗНАНИЙ НА ОСНОВЕ СЕМАНТИЧЕСКИХ WIKI-РЕСУРСОВ

Юлия Рогушина

Институт программных систем НАН Украины

Аннотация:

Проанализированы особенности Wiki-ресурсов и средства их семантической разметки. Предлагается метод использования семантических Wiki-ресурсов для формирования персонифицированных онтологий. Рассматриваются способы применения этих онтологий в интеллектуальных Web-приложениях.

Ключевые слова:

Wiki-ресурс; онтология; семантический поиск.

Анотація:

Проаналізовано особливості Wiki-ресурсів і засоби їх семантичної розмітки. Запропоновано метод використання семантичних Wiki-ресурсів для формування персоніфікованих онтологій. Розглянуто способи використання цих онтологій в інтелектуальних Web-застосуваннях.

Ключові слова:

Wiki-ресурс; онтологія; семантичний пошук.

Resume:

Rohushyna Yuliia. Design of personalized ontological knowledge databases on the basis of semantic Wiki-resources.

The features of Wiki-resources and their semantic markup means are analyzed. The method of semantic Wiki-resource use for generating personalized ontologies is proposed. The means of application of these ontologies in intelligent Web-based applications are considered.

Key words:

Wiki-resource; ontology, semantic search.

Современные информационные системы (ИС), используемые в науке, образовании и других сферах деятельности, как правило, ориентированы на использование знаний и функционирование в открытой информационной среде. Это предполагает, что они способны извлекать знания, которые содержатся в динамичных и гетерогенных информационных ресурсах (ИР), сохранять их и использовать для получения результата, в котором нуждается пользователь.

Использование ресурсов Web обеспечивается *актуальностью* сведений, используемых при работе ИС, а знания позволяют обрабатывать эти сведения на семантическом уровне знаний. К сожалению, несмотря на развитие проекта Semantic Web и связанных с ним средств интеллектуализации открытой информационной среды, значительная часть ресурсов Web не содержит семантической разметки и метаданных, и поэтому извлечение знаний из них – сложный и трудоемкий процесс, который не может быть полностью автоматизирован. Исключение составляют *семантические Wiki-ресурсы*, которые становятся в последнее время все популярнее. Именно эти ресурсы могут использоваться как источник для формирования базы знаний (БЗ) для других приложений. *Онтологический подход* поддерживает интероперабельность и повторное использование таких БЗ, из которых могут формироваться онтологические репозитории. Онтологии позволяют моделировать интересующие пользователей предметные области (ПрО) или проблемы.

Постановка задачи. Чтобы пользователь мог самостоятельно формировать и пополнять онтологические БЗ, которые непосредственно

связаны с интересующей его проблемой или предметной областью и содержат наиболее актуальные знания о ней, предлагается разработать методы извлечения таких знаний из семантических Wiki-ресурсов. Использование Wiki-ресурсов в качестве источников знаний обуславливается тем, что они динамически обновляются всем сообществом пользователей, имеют четко определённую и простую для понимания структуру, а также обеспечивают обработку информации на семантическом уровне.

Семантические Wiki-ресурсы. *Wiki-технология* представляет собой технологию создания Web-сайта, которая дает пользователям возможность коллективно участвовать в редактировании его контента – исправление ошибок, добавление новых материалов. Эта технология позволяет аккумулировать знания, представляя их в интероперабельной форме, обеспечить навигацию по этой базе знаний и средства ее актуализации. При этом использовать Wiki могут сообщества различного объема и тематической направленности, создавая базы знаний от глобальных Википедий и электронных энциклопедий крупных корпораций до легко обновляемых справочных систем небольших организаций, предприятий и учебных заведений [3].

Широко известная Википедия – наибольшая из бесплатных онлайн-энциклопедий – создана на основе технологии Wiki. Разрабатываемая в настоящее время электронная версия Большой украинской энциклопедии также ориентирована на использование Wiki-технологий [2]. Влияние Semantic Web на подобные проекты постоянно возрастает: движки Wiki-сайтов поддерживают онтологическое представление знаний

и семантическую разметку ресурсов. Это позволяет включать семантические аннотации в Wiki-разметку в виде OWL и RDF и явно разделять структурированную и неструктурированную информацию. Эти динамически пополняемые новыми знаниями БЗ могут интероперабельно использоваться другими знание-ёмкими приложениями Semantic Web [5], и в первую очередь теми, которые ориентированы на поддержку научных исследований и образовательного процесса.

Существует много различных движков для создания и поддержки Wiki-ресурсов, которые обеспечивают семантизацию информации. Например, для MediaWiki [4] (это свободное программное обеспечение с открытым кодом для гипертекстовой Wiki-среды, являющееся платформой для создания справочников, энциклопедий и каталогов) существует семантическое расширение Semantic MediaWiki (SMW), которое позволяет пользователям добавлять семантические аннотации к Wiki-страницам, превращая MediaWiki в семантический ресурс [6]. В нем используются такие дополнительные элементы разметки, как категории, семантические свойства и запросы. В основу SMW положена концепция семантических свойств (для создания данных) и семантических запросов (для использования данных). Пользователи размечают статьи категориями и свойствами для того, чтобы информация становилась доступной для запросов. Система SMW написана с помощью механизма расширений MediaWiki (<https://www.mediawiki.org/wiki/MediaWiki>) [4]. Это упрощает интеграцию в существующие

приложения MediaWiki. Необходимые коллекции семантических данных в SMW пополняются пользователями, которые могут добавить аннотации к Wiki-текстам статей с помощью специальной разметки. Semantic MediaWiki предоставляет инструменты для обработки данных, полученных из этих аннотаций.

Рассмотрим основные компоненты Wiki. Эти ресурсы (энциклопедии, справочники и т. д.) состоят из набора страниц. Некоторые из них являются обычными (тематическими), а некоторые – служебными (страницы категорий, семантических свойств, шаблонов, обсуждения и т. д.). Каждая из служебных страниц относится к тому или иному пространству имен. Страница из каждого пространства имен включается в любую категорию путем добавления тэга-ссылки.

Категории предназначены для того, чтобы классифицировать обычные страницы с помощью различных иерархий. Чтобы отнести страницу к определенной категории, нужно добавить ее в код [[Категория:Название категории]]. Страницы категорий, содержащие описание и назначение соответствующей категории, создаются и редактируются так же, как и обычные страницы. На страницах категорий сортировки вырабатываются автоматически, а на страницах статей категории располагаются в порядке включения соответствующих тэгов. Страницы категорий позволяют устанавливать иерархию категорий: для этого следует указать, что определенная категория является подкатегорией другой категории. Например, категория «Научные работники» – подкатегория «Персоналии».



Рис. 1. Работа в SMW с семантическими свойствами

Для привязывания данных к Wiki-страницам используются *семантические свойства*. Каждое свойство имеет тип, название и значение, кроме того, ему соответствует отдельная Wiki-страница в специальном пространстве имен, которая позволяет задавать тип свойства, определять его положение в иерархии свойств, а также документировать его использование. Свойства, в отличие от категорий MediaWiki, имеют не только название, но и значение. В текст Wiki-страницы семантическое свойство вставляется в формате [[Имя свойства::Значение свойства]]. Семантические свойства используются для того, чтобы определить смысловую нагрузку ссылок на другие страницы. Например, если на странице «Украина» встретится [[столица:Киев]], то это означает, что понятия «Украина» и «Киев» связаны отношением «столица». Имена семантических свойств могут использоваться в шаблонах и в запросах к контенту Wiki-страниц (рис. 1).

Стандартизация семантических свойств и категорий. MediaWiki позволяет пользователям создавать произвольные наборы категорий и семантических свойств, не накладывая на их имена и связи между ними какие-либо ограничения. Однако если использовать разные названия для семантических свойств и категорий, то это значительно усложнит создание запросов к ним. Поэтому часто при создании Wiki-ресурсов вначале разрабатываются различные рекомендации относительно структуры свойств и категорий. Например, в [1] предложены типичные классификаторы и схемы статей для различных направлений наук и типичные схемы биографических статей (в частности, статей об ученых в области общественных наук или техники, писателей, государственных и политических деятелей). Если разработать на их основе единую систему категорий и семантических свойств, то можно предоставить пользователю Wiki-ресурса развитые возможности по интеграции имеющихся сведений и семантическому поиску среди имеющейся информации. Например, можно автоматически генерировать перечни исследователей, работавших в определенное время над той или иной проблемой, находящихся в выбранной местности, организациях и т. д. При внесении изменений в те ресурсы, которые служат основой для построения таких списков, сами перечни будут динамично обновляться.

Извлечение онтологических знаний из Wiki-страниц. Приведенный выше анализ публикаций показал, что часто для персонализации поиска в Web

целесообразно использовать онтологии, характеризующие область интересов пользователя. При этом, как правило, речь идет не о крупной онтологии, предоставляющей описание ПрО в целом, а о сравнительно небольшой онтологии с несложной структурой, которая формализует непосредственную информационную потребность пользователя, – онтологии проблемы (ОП). ОП позволяет трансформировать запросы пользователя, расширяя или уточняя их, использовать синонимию и устанавливать меру семантической близости между ИР и проблемой, решаемой пользователем. Благодаря относительно простой структуре, для ОП можно использовать достаточно простые и быстрые алгоритмы обработки.

В ОП можно выделить следующие элементы:

- классы (понятия);
- экземпляры классов (экземпляры);
- иерархические отношения между классами («быть подклассом»);
- отношения синонимии между классами («является синонимом»);
- отношение принадлежности между экземпляром и классом («быть экземпляром класса»);
- отношения между экземплярами, специфичные для ПрО (например, «Дисциплина_А *входит* в состав Специальности_Б»), – свойства объекта, устанавливающие связи между двумя экземплярами классов;
- специфические для ПрО свойства экземпляров (например, «Дисциплина_А *занимает* X часов») – свойства данных, устанавливающие связи между экземпляром класса и данными.

Такую онтологию пользователь может достаточно легко построить сам на основе семантически размеченного Wiki-ресурса. При этом категории Wiki (и связанные с ними страницы) соответствуют классам ОП, обычные страницы – экземплярам ОП, семантические свойства типа «Страница» характеризуют отношения между экземплярами, а семантические свойства других типов – свойства экземпляров.

При формировании ОП пользователю не надо использовать все Wiki-страницы, достаточно проанализировать несколько выбранных и тех, которые связаны с ними. При этом созданная ОП будет, с одной стороны, достаточно компактной и отражающей персональные потребности пользователя, а с другой – благодаря использованию знаний, содержащихся в семантической разметке, учитывающей объективные знания о выбранной ПрО. Предполагается, что пользователь обладает

достаточной квалификацией в выбранной ПрО, чтобы найти пертинентный ей Wiki-ресурс.

Предлагается использовать следующий алгоритм построения ОП, при котором пользователю надо анализировать не только информацию, которая выводится на экран, но и код страниц, доступный при редактировании:

1. Пользователь выбирает Wiki-страницу, соответствующую ключевому слову его запроса. На этой странице нужно найти категории страницы (текст [[Категория:Название категории]]). Если страница относится к более чем одной категории, то с помощью подраздела «Категории» в меню «Специальные страницы» уточняются связи между ними.

2. В ОП создаются соответствующие категориям классы и экземпляр класса, соответствующий выбранной странице.

3. Если пользователя заинтересовали другие страницы этих категорий, то для них также создаются экземпляры в ОП. Их легко найти при переходе в навигации к соответствующей категории.

4. Чтобы учитывать связи между экземплярами ОП, необходимо проанализировать ссылки и семантические свойства, указанные на этих страницах. В коде страницы может присутствовать простая ссылка на другую Wiki-страницу – имя страницы в двойных квадратных скобках. [[имя страницы]]. Тогда можно добавить в ОП отношение (свойство объекта) с фиксированным именем («Связь страниц»), связывающее текущую страницу со страницей, указанной в «имя страницы». Для этого надо найти в коде страниц элементы типа [[Название свойства::значение свойства,| что выводить в тексте статьи]]. В ОП добавляется отношение (свойство объекта) с именем («Название свойства»), связывающее текущую страницу со страницей, указанной в «значение свойства».

5. Затем аналогичные действия нужно применять рекурсивно к найденным страницам, связанным с выбранной, до тех пор, пока пользователь не посчитает ОП сформированной.

Пользователь может применять указанные действия не ко всем обнаруженным категориям, Wiki-страницам и связям между ними, а только к тем, которые он посчитает относящимися к интересующей его задаче. Таким образом, обеспечивается персонификация строящейся ОП.

Использование онтологии проблемы для решения прикладных задач. Наличие онтологии обеспечивает интеграцию различных ресурсов и приложений. Рассмотрим это на примере задачи разработки интеллектуальной системы информационного сопровождения Национальной рамки квалификаций. Можно

предположить, что сведения об объектах и субъектах этой рамки – учебных заведениях, их программах, курсах, о различных национальных и международных стандартах, нормативах, компетенциях специальностях и т. п. должны находиться в Web. Но анализ неструктурированных естественно-языковых документов является сложной и трудоемкой задачей, поэтому целесообразнее извлекать необходимые сведения о них из соответствующих семантически размеченных Wiki-ресурсов. В частности, наиболее предпочтительно обращаться при этом к национальным энциклопедиям, поскольку содержащаяся там информация более достоверна и создана с учетом современных теоретико-методологических разработок.

Модель семантического поиска достаточно легко адаптировать для решения этой и других прикладных задач. Рассмотрим это на примере проблемы сопоставления компетенций, которая является составной частью таких задач, как поиск работодателем подходящих исполнителей работ; сравнение квалификации специалистов, имеющих различные специальности (в частности, соответствующие нормативам различных стран); выбор абитуриентом учебного заведения, предлагающего необходимый ему набор дисциплин; оценка возможности перехода студента из одного учебного заведения в другое (какие дисциплины из ранее изученных можно засчитывать) и др. При этом учитываются только те категории и семантические свойства страниц, которые соотносятся с понятиями онтологии данной задачи. Модель задачи содержит описание базовых понятий – «дисциплина», «специальность», «содержательный модуль», «компетенция», «кандидат» и др., а также связи между ними и структуру ИО – результат поиска (рис. 2). Это позволяет отображать экземпляры искомого класса в терминах онтологии ПрО. Для сопоставления специальностей, умений и компетенций людей и организаций разных стран целесообразно использовать набор эталонных атомарных экземпляров каждого класса. Экземпляр считается атомарным, если ни один другой экземпляр этого класса не является его подмножеством.

Наличие сформированной по Wiki-ресурсам онтологии ОП позволяет пользователю формировать семантические запросы к этим ресурсам. Семантические Wiki поддерживают возможность построения многокритериальных запросов на формальном языке. Используя семантические запросы, можно автоматически строить списки страниц по ряду условий. Страницы всегда будут актуальными, поскольку строятся автоматически.

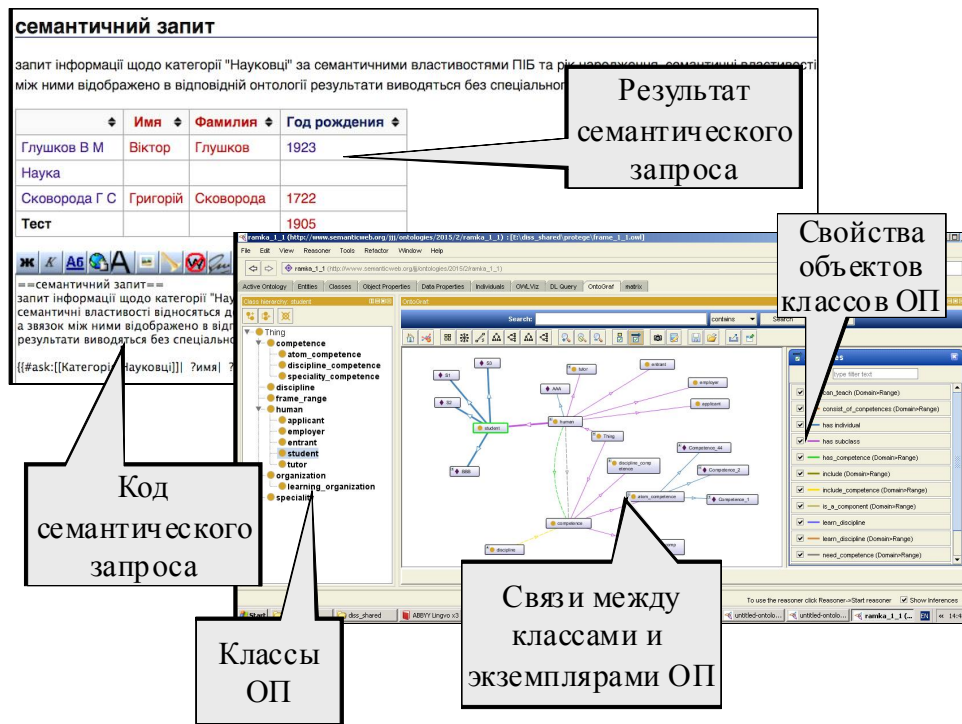


Рис. 2. Семантический запрос к Wiki и соответствующая ему ОП

Любой семантический запрос можно представить как набор условий, ограничивающих выбор; необходимых параметров выборки и формата вывода данных:

```
[[#ask: [[условия]] | ?что_выводить | format=форма_вывода]]
```

Напрмер, запрос `[[#ask: [[Категория:Дисциплина]] [[Специальность:Информатика]] | ?Название | ?вуз | ?Количество_часов | ?Рейтинг]]` позволяет вывести название и другие характеристики дисциплин, входящих в специальность «Информатика». ОП применяется при построении запроса для того, чтобы пользователь мог употребить правильные названия интересующих его семантических свойств и категорий, не просматривая все их значения, используемые в соответствующем Wiki-ресурсе.

В результате выполнения такого пользовательского запроса, динамически формируется Wiki-страница, которая содержит необходимую пользователю информацию.

Список использованных источников

1. Методичні рекомендації з підготовки, редагування та оформлення статей до Великої української енциклопедії (проект) / за ред. д. і. н., проф. Киридон А. М. – К.: ДНУ «Енциклопедичне видавництво», 2015. – 120 с.
2. Наукові засади та теоретико-методологічні принципи створення сучасних енциклопедій: колективна монографія / за ред. д. і. н., проф. Киридон А. М. – К.: ДНУ «Енциклопедичне видавництво», 2015. – 160 с.
3. Рогущина Ю. В. Знание-ориентированные средства поддержки семантического поиска в Web / Ю. В. Рогущина. – LAP LAMBERT Academic Publishing, 2014. – 214 с.

По этой странице можно также автоматически обновлять ОП, генерируя ее RDF-код средствами SMW или обрабатывая ее с помощью предложенного выше алгоритма. Таким образом, онтологии являются как инструментом для создания семантических запросов, так и результатом обработки Wiki-ресурсов.

Выводы. Предложен подход к построению персонифицированной онтологической БЗ предметной области, которая интересует пользователя. Эта модель формируется на основе обработки семантически размеченных Wiki-ресурсов, которые самостоятельно выбирает пользователь на основании собственных представлений о той проблеме, которую он решает. При построении этой онтологии используются семантические свойства, ссылки и категории Wiki, на основании которых создаются классы и экземпляры классов в онтологии. Такой подход позволяет обеспечить прикладные системы актуальными персонифицированными знаниями.

References

1. Kyrydon, A. M., ed. (2015). *Methodical guidelines on preparation, editing and formatting of articles to the Great Ukrainian Encyclopedia (project)*. Kyiv: DNU "Entsyklopedychnе vydavnytstvo". [in Ukrainian]
2. Kyrydon, A. M., ed. (2015). *Scientific principles and theoretical-methodological principles of modern encyclopedias: collective monograph*. Kyiv: DNU "Entsyklopedychnе vydavnytstvo". [in Ukrainian]
3. Rohushyna, Yu. V. (2014). *Knowledge-oriented means to support the semantic search in the Web*. LAP LAMBERT Academic Publishing. [in Russian]
4. Davies, J., Harmelen, F. (2002). *Towards the Semantic Web: Ontology-driven knowledge management*. John

4. Barrett D. J. *MediaWiki (Wikipedia and Beyond)*. O'Reilly Media, 2008 [Электронный ресурс] / D. J. Barrett. – Режим доступа : <http://www.nhmnc.info/wp-content/uploads/fbpdfs2014/MediaWiki-Wikipedia-and-Beyond-by-Daniel-J-Barrett-Good-Basic-Primer.pdf> (23. 03. 2016).
5. Davies J. *Towards the Semantic Web: Ontology-driven knowledge management* / J. Davies, F. Harmelen. – John Wiley & Sons Ltd, England, 2002. – 288 p.
6. Krotzsch M. *Semantic MediaWiki* [Электронный ресурс] / Krotzsch M., Vrandecic D., Volkel M. – Режим доступа : <http://c.unik.no/images/6/6d/SemanticMW.pdf> (23. 03. 2016).
- Wiley & Sons Ltd, England. [in English]
5. Barrett, D. J. (2008). *MediaWiki (Wikipedia and Beyond)*. O'Reilly Media, 2008. Retrieved from : <http://www.nhmnc.info/wp-content/uploads/fbpdfs2014/MediaWiki-Wikipedia-and-Beyond-by-Daniel-J-Barrett-Good-Basic-Primer.pdf>. [in English]
6. Krotzsch, M., Vrandecic, D., Volkel, M. *Semantic MediaWiki*. Retrieved from : <http://c.unik.no/images/6/6d/SemanticMW.pdf> [in English]

Рецензент: Функтикова О.О.– д.пед.н., профессор

Сведения об авторе:

Рогущина Юлия Витальевна

ladamandraka2010@gmail.com

Институт программных систем НАН Украины

пр. Академика Глушкова, 40, г. Киев,

03680 МСП, Украина

doi: <http://dx.doi.org/10.7905/nvmdpu.v0i16.1408>

Материал поступил в редакцию 26. 03. 2016 р.

Принят к печати 28.04.2016 р.