

УДК 004.853, 004.55

РАЗРАБОТКА РАСПРЕДЕЛЁННЫХ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ НА ОСНОВЕ ОНТОЛОГИЧЕСКОГО АНАЛИЗА И СЕМАНТИЧЕСКИХ WIKI-ТЕХНОЛОГИЙ

Ю.В. Рогушина

*Институт программных систем НАН Украины, Киев, Украина
ladamandraka2010@gmail.com*

Аннотация

Рассматривается необходимость автоматизированного получения знаний интеллектуальными информационными системами из распределённой среды Web. Оценивается целесообразность использования онтологического анализа для представления таких знаний. Обосновывается необходимость пополнения этих знаний из более динамичных источников информации, предусматривающих автоматизированную обработку информации, к которым относятся семантические Wiki-ресурсы. Такая обработка информации требует разработки модели и методов сопоставления элементов Wiki с элементами онтологии. Проанализированы выразительные возможности технологической среды Semantic MediaWiki, которые могут использоваться для семантической разметки естественно-языковой и мультимедийной информации. Предложен метод семантической разметки Wiki-ресурсов с помощью элементов онтологической модели предметной области, которая базируется на установлении соответствий между элементами формальной модели такой онтологии и элементами Semantic MediaWiki (категориями, семантическими свойствами, ссылками и т.п.). Разработан метод усовершенствования онтологии предметной области на основе Wiki-ресурса, семантически размеченного элементами этой онтологии. Приведена апробация разработанных моделей и методов на примере разработки электронной версии Большой украинской энциклопедии на платформе Semantic MediaWiki. Новизна работы заключается в способе сопоставления элементов онтологической модели предметной области с элементами семантического Wiki-ресурса, а также в методе усовершенствования онтологии предметной области, применяемой для семантической разметки Wiki-ресурса, на основе изменений и дополнений, которые вносятся в этот ресурс.

Ключевые слова: *Wiki-ресурс, онтология, информационный объект, семантическая разметка.*

Цитирование: *Рогушина, Ю.В.* Разработка распределённых интеллектуальных систем на основе онтологического анализа и семантических Wiki-технологий / Ю.В. Рогушина // Онтология проектирования. – 2017. – Т.7, №4(26). – С. 453-472. – DOI: 10.18287/2223-9537-2017-7-4-453-472.

Введение

Интеллектуальные информационные системы (ИИС), которые функционируют в открытом распределённом информационном пространстве, постоянно пополняют и обновляют знания, которые поступают из внешней среды. В современных ИИС использование знаний о *предметной области (ПрО)* и методы их обработки играют решающую роль и определяют полезность получаемых результатов. Это в свою очередь обуславливает актуальность исследований, направленных на получение и накопление знаний, решение проблем распознавания, логического вывода новых знаний для поддержки принятия решений, то есть на развитие средств управления знаниями, в том числе – онтологическими. Из-за того, что извлечение знаний из естественно-языковых и мультимедийных информационных ресурсов (ИР) требует много времени и усилий, целесообразно отдавать предпочтение тем ИР, которые уже содержат семантическую разметку и сопровождаются семантическими метаописаниями. Однако в

связи с гетерогенностью информации в Web различные виды таких ИР требуют разработки специфичных средств и методов извлечения знаний.

Поэтому для достижения практически полезных результатов целесообразно ориентироваться на те ИР, которые широко распространены, имеют достаточно высокую выразительную мощность, надёжную реализацию и удобный пользовательский интерфейс. Этим требованиям удовлетворяют семантические Wiki, в частности, ресурсы, созданные на основе Semantic MediaWiki. Эта платформа сегодня широко используется, а сформированные на её основе ИР часто обновляются и быстро увеличиваются. Это обуславливается тем, что Wiki-ресурсы динамично обновляются всем сообществом пользователей, имеют чётко определённую и простую для понимания структуру и технологическую платформу для группового управления знаниями. Но для того, чтобы Wiki-ресурсы можно было использовать в качестве распределённой базы знаний (БЗ), необходимо уже в процессе создания этих ресурсов применять к ним подходы, которые используются в системах управления знаниями.

1 Семантические Wiki-ресурсы

Одним из наиболее распространённых сегодня Wiki-движков является MediaWiki – свободное программное обеспечение с открытым кодом, которое является платформой для создания различных справочников, энциклопедий и каталогов. Эта программа позволяет создавать и редактировать разнообразные Wiki-ресурсы, предоставляет пользователям удобный и интуитивно понятный интерфейс и поддерживает большинство функций, необходимых для коллективной работы. В частности, на этой платформе работает Википедия.

Удобным инструментом классификации знаний, представленных в Wiki-ресурсе, является категоризация страниц. В Wiki для категоризации может использоваться произвольное количество классификаторов, которые позволяют сформировать множество Wiki-страниц, объединённых общими свойствами, и осуществлять поиск статей по классификаторам. Каждая страница Wiki-ресурса может быть отнесена к одной или к нескольким категориям любого из классификаторов (в зависимости от специфики самого классификатора).

По своей сущности Wiki-технология предполагает сотрудничество и сосуществование разных точек зрения без конкурентного вытеснения, поэтому система категорий в любой области может представлять собой не дерево, принадлежащее лишь к одной классификации, а систему древовидных структур, которые накладываются одна на другую. Одна и та же категория или статья может быть отнесена к категориям из разных классификаций, и при категоризации статьи в рамках одной классификационной системы практически не важно, классифицирована она или нет в рамках других систем. Например, Wiki-страница, которая описывает определённую организацию, может быть классифицирована по типу деятельности учреждения («принадлежит к НАНУ»), по её местонахождению («организации города Киева»), по сфере деятельности («научные исследования»). В Wiki могут использоваться как стандартные классификаторы, такие как УДК, так и введённые самими пользователями.

Semantic MediaWiki (SMW) [1] – плагин MediaWiki, который обеспечивает семантизацию информации в Wiki-ресурсе. SMW превращает MediaWiki в семантический ресурс, позволяя автоматически интегрировать информацию из разных Wiki-страниц, генерировать ответы на сложные семантические запросы, строить БЗ и визуализировать их результаты, которые отображают знания относительно семантических свойств и категорий информационных объектов (ИО), выполнять логический вывод и т.п., то есть обрабатывать информацию на уровне знаний. Кроме того, SMW позволяет экспортировать онтологические знания в OWL [2] – языке представления онтологий, разработанного в рамках Semantic Web [3].

В Semantic MediaWiki используются такие дополнительные элементы разметки, как *семантические свойства* и *семантические запросы*. Для того чтобы превратить Wiki-ресурс в распределённую базу знаний, необходимо определить базовые принципы классификации страниц и построить систему категорий и семантических свойств, которые позволяют формализовать и интегрировать знания, которые содержатся на отдельных Wiki-страницах.

Semantic MediaWiki имеет достаточно высокую выразительную мощь, но по сравнению с онтологиями выразительные возможности Semantic MediaWiki значительно слабее и не позволяют, например, задавать отношения между самими семантическими свойствами или определять такие их характеристики, как транзитивность и рефлексивность, которые обеспечивают логический вывод новых знаний из имеющихся.

Из-за того, что встроенные к Semantic MediaWiki средства семантизации не поддерживают анализ и визуализацию этой системы знаний, целесообразно применять внешние средства управления распределёнными знаниями на основе онтологического анализа. Для этого необходимо разработать методы сопоставления семантической разметки Wiki-страниц с элементами онтологий.

Проблема сопоставления онтологий и семантических Wiki-ресурсов возникает в нескольких случаях. Во-первых, при создании семантических Wiki-ресурсов необходимо сначала сформировать набор категорий и семантических свойств. Но встроенные средства Semantic MediaWiki не позволяют ни визуализировать эту информацию, ни оценить её целостность и непротиворечивость. Поэтому целесообразно сначала построить онтологию той ПрО, которая отображается в Semantic MediaWiki, а потом использовать эту онтологию как основу для семантической разметки. Во-вторых, семантически размеченные Wiki-ресурсы намного более динамичны в сравнении с онтологиями – в их усовершенствовании и восстановлении может принимать участие широкий круг пользователей, и потому они могут быть полезны для усовершенствования соответствующей онтологии ПрО.

Чтобы обеспечить ИИС возможностью импортировать, модифицировать и обновлять онтологические знания, которые приобретаются из внешних семантически размеченных Wiki-ресурсов, необходимо разработать методы и средства, которые обеспечивают решение следующих задач:

- построить взаимно-однозначные соответствия между основными элементами онтологии ПрО (классы и подклассы, экземпляры, свойства объектов и данных, их значения) и элементами Semantic MediaWiki (категории, семантические свойства, их значения и связи);
- разработать средства и критерии сопоставления онтологий, которые являются усовершенствованиями одной и той же онтологии ПрО, пополненной знаниями из разных Wiki-ресурсов, которые семантически размечены элементами этой начальной онтологии.

Семантические свойства обеспечивают привязывание данных к Wiki-страницам. Каждое свойство имеет тип, название и значение, кроме того, ему соответствует отдельная Wiki-страница в специальном пространстве имён, которая позволяет задавать тип свойства, определяет его положения в иерархии свойств, а также документирует его использование. Свойства, в отличие от категорий MediaWiki, имеют не только названия, но и значения.

В текст Wiki-страницы семантическое свойство вставляется в формате `[[Имя свойства:Значение свойства]]`. Свойства могут быть разных типов, и именно эти типы определяют операции, которые могут выполняться со значениями свойств, например, сортировка и сравнение. SMW поддерживает 15 типов свойств: *Число*, *Строка*, *Дата*, *Географические координаты* и т.п. Семантические свойства типа «Страница» подобны гиперссылкам между Wiki-статьями, но позволяют определить смысловую нагрузку этих ссылок. Например, для страницы «Украина» связь со страницей «Киев» можно определить свойством «столица».

Для того, чтобы зафиксировать такую связь, достаточно заменить текст *[[Украина]]*, который содержится на странице «Киев», на *[[является столицей::Украина]]*.

Значительный интерес представляет следующее: если в текст шаблона поместить семантическое свойство, то это свойство будут иметь все страницы, которые используют такой шаблон.

В Semantic MediaWiki реализован простой, но мощный язык запросов SMW-QL, что открывает широкие возможности для семантического поиска в Wiki-ресурсах. В то время как семантические свойства и категории позволяют структурировать данные в Wiki, запросы нужны для того, чтобы использовать эту информацию: они помогают Wiki-пользователям и Wiki-администраторам интегрировать нужные сведения.

Язык запросов SMW-QL позволяет отфильтровывать страницы по заданным критериям и выводить на страницу только интересующую пользователя информацию (значения семантических свойств), а не весь контент Wiki-страницы.

Результаты таких запросов позволяют динамично интегрировать актуальную информацию из других Wiki-страниц и отображать её в удобном для пользователя формате – в виде диаграмм, географических карт, таблиц и схем и т.п. При внесении изменений в страницы, из которых извлекаются данные, результаты запросов также автоматически обновляются, обеспечивая непротиворечивость и согласованность данных.

Наиболее часто используются встроенные запросы с функцией *ask*. Её вызов обозначается двойными фигурными скобками, перед именем стоит символ *#*, а после – двоеточие. Сначала передаётся сама строка запроса, который отбирает нужную информацию из Wiki, а потом все параметры запроса, разделённые символами вертикальной черты *|*. Запрос на SMW-QL обычно встраиваются в код Wiki-страницы в следующей форме:

```
{{#ask:Критерии выбора страниц  
|?Запрашиваемое свойство  
|Другие параметры  
}}
```

Функция *ask* имеет три параметра. Первый параметр отбирает из всего множества страниц лишь те страницы, среди которых надо осуществлять поиск. Например, это может быть категория или множество категорий, в которые входит страница, определённые условия, которым должна соответствовать страница. В качестве критериев выбора страниц указываются значения семантических свойств и категорий, которые имеют страницы. Второй параметр определяет, какие именно семантические свойства нужно найти на отобранных страницах. Третий параметр указывает на то, какую информацию нужно предоставить пользователю, и какова форма её представления.

По выразительным возможностям языка встроенных запросов Semantic MediaWiki значительно уступает языку запросов SPARQL [4], но использование SPARQL требует от пользователей значительно больших знаний. В запросах к Semantic MediaWiki могут использоваться категории и семантические свойства, а в запросах к онтологиям – классы и свойства, которые делятся на *object properties* и *data properties*.

Оба языка позволяют выполнять сложные запросы (сортировать, ограничивать, упорядочивать результаты и т.д.), но язык встроенных запросов SMW даёт возможность оперировать только теми данными, которые размещены на одном Wiki-ресурсе: он позволяет работать только со специальными категориями и свойствами самой Semantic MediaWiki, а также категориями и свойствами, которые создали пользователи конкретной Semantic MediaWiki. Язык

SPARQL, наоборот, изначально ориентирован на интероперабельность и позволяет обрабатывать одновременно разные онтологии.

2 Постановка задачи

При разработке интеллектуальных распределённых систем, базирующихся на семантических Wiki-технологиях, предлагается использовать онтологии ПрО в качестве основы для семантической разметки. Это позволяет более чётко формализовать знания ПрО и обеспечивает автоматизированное пополнение такой онтологии в процессе развития Wiki-ресурса. Для этого необходимо разработать средства сопоставления основных элементов Wiki-ресурса и онтологии, а также метод пополнения онтологии знаниями, извлекаемыми из семантически размеченных Wiki-страниц.

3 Онтологическая модель предметной области

Для интероперабельного представления распределённых знаний в ИИС широко применяются онтологии, которые обеспечивают повторное использование знаний [5]. В последние годы онтологии активно развиваются как семантические модели, которые позволяют отображать различные аспекты реального мира [6].

Онтологии базируются на фундаментальном теоретическом базисе дескриптивных логик, для них уже существуют общепринятые стандарты описания, языки и программные средства. Особый интерес представляет использование онтологического анализа как основы для обработки распределённых знаний. В связи с этим значительное количество исследований связано с теоретическим базисом онтологий, их построением, усовершенствованием, с получением знаний из онтологических структур, а также с другими важными аспектами менеджмента онтологий, которые в значительной мере зависят от цели и назначения такого анализа.

Одним из важных направлений таких исследований является интеграция онтологий с другими ИР Web, в частности – с Wiki-ресурсами. Такие свойства Web, как гетерогенность и динамичность, обуславливают ряд проблем, связанных с пополнением, использованием и оценкой онтологий, и вызывают потребность использования более динамических семантических Wiki-ресурсов, на основе которых могут обновляться эти онтологии.

Онтологический анализ позволяет превращать описание представления о внешнем мире в набор терминов и правил их использования, пригодных для машинной обработки. Онтология представляет собой явную спецификацию концептуализации. Её можно рассматривать как БЗ специального вида с семантической информацией о ПрО. Компоненты, из которых складываются конкретные онтологии, зависят от парадигмы представления, но практически все модели онтологий содержат классы и их экземпляры, свойства классов, отношения между классами и экземплярами классов и ограничения на их использование.

В общем случае формальная модель онтологии ПрО – это упорядоченная тройка $O = \langle T, R, F \rangle$, где T – множество понятий ПрО; R – множество отношений между ними; F – множество функций интерпретации понятий и отношений.

В дальнейшем будем использовать следующую формальную модель

$$(1) \quad O = \langle X, R, F, T \rangle$$

Эта модель состоит из следующих элементов:

- $X = X_{cl} \cup X_{ind}$ – множество концептов онтологии, где X_{cl} – множество классов, X_{ind} – множество экземпляров классов, таких, что $\forall a \in X_{ind} \exists A \in X_{cl}, a \in A$;

- $R = r_{ier_cl} \cup \{r_i\} \cup \{p_j\}$ – множество отношений между элементами онтологии, где r_{ier_cl} – иерархическое отношение, которое может устанавливаться между классами онтологии и свойствами классов и характеризуется такими свойствами, как антисимметричность и транзитивность, $r_{ier_cl} : X_{cl} \rightarrow X_{cl}$; $\{r_i\}$ – множество объектных свойств, которые устанавливают отношения между экземплярами классов: $r_i(a, a \in X_{ind}) = b, b \in X_{ind}, r_i : X_{ind} \rightarrow X_{ind}$; $\{p_j\}$ – множество свойств данных, которые устанавливают отношение между экземплярами классов и значениями: $p_j(a, a \in X_{ind}) = t, t \in T, p_j : X_{ind} \rightarrow Const$, такие, что внутри множеств объектных свойств и свойств отношений также могут существовать иерархические отношения $r_{ier_obj}, r_{ier_obj} : \{r_i\} \rightarrow \{r_i\}$ и $r_{ier_data}, r_{ier_data} : \{p_j\} \rightarrow \{p_j\}$;
- F – множество характеристик классов онтологии, экземпляров классов и их свойств, которые могут применяться для логического вывода (например, эквивалентность, отличие, отсутствие пересечения, область определения и область значения);
- T – множество типов данных (например, строка, целое).

Выбор именно такой модели онтологии обуславливается следующими причинами. Во-первых, такая модель имеет достаточную выразительность для решения тех задач, которые рассматриваются в этой работе. Во-вторых, она соответствует интуитивному представлению об онтологиях, которое содержится в пользовательском интерфейсе редактора онтологий Protégé [7] и потому легко объединяется с визуализациями элементов онтологии в этом программном продукте. В-третьих, эта модель довольно легко интегрируется с различными приложениями, которые поддерживают семантическую обработку информации (в частности, с семантическими Wiki-ресурсами).

Как правило, в ИИС различают внутренние и внешние онтологии [8]. Внутренние онтологии создаются разработчиками ИИС и отображают их представление относительно свойств и взаимоотношений основных ИО определённой ПрО. Структура внутренней онтологии и её объём целиком определяются целями создания ИИС и теми функциями, которые она может выполнять в системе. Структура внутренней онтологии остаётся неизменной в процессе функционирования ИИС и может лишь пополняться новыми экземплярами класса или значениями их свойств.

Внешние онтологии импортируются ИИС из внешних информационных источников в процессе работы системы и позволяют динамично возобновлять сведения относительно соответствующей ПрО. Например, такие онтологии могут быть найдены по определённым условиям в репозиториях онтологий или построены путём анализа доступных ИР. Важным свойством внешних онтологий является то, что в процессе работы ИИС для решения одной и той же задачи в разные моменты времени могут применяться разные онтологии. Это объясняется динамичностью информационной среды (это в особенности характерно для ИИС, ориентированных на работу в Web): онтологии в репозитории или внешние ИР могут значительно изменяться независимо от разработчиков и пользователей ИИС.

Такие свойства Web, как гетерогенность и динамичность, обуславливают ряд проблем, связанных с использованием внешних онтологий. В частности, возникает проблема сравнения внешних онтологий, сформированных в разные моменты работы ИИС: если отличия между онтологиями превышают определённую количественную меру, тогда необходимо снова выполнить их обработку.

Рассмотрим некоторые частные случаи внешних онтологий: если внешнюю онтологию можно отнести к одному из таких типов, то это значительно упрощает её обработку и использование.

Таксономия T – это онтология, которая, кроме классов и экземпляров, содержит только отношение «подкласс», иерархически объединяющее классы, и отношения «экземпляр класса». Эта онтология не требует сложных средств обработки, методы её построения не предполагают наличия циклов в классах (каждый класс создаётся как подкласс определённого класса, экземпляр – как экземпляр определённого класса). Её формальная модель является частным случаем модели (1):

$$(2) \quad T = \langle X = X_{cl} \cup X_{ind}, R = \{r_{ier_cl}\}, \emptyset, \emptyset \rangle.$$

Wiki-онтология – это онтология, построенная по семантически размеченному Wiki-ресурсу (набору Wiki-страниц, которые содержат семантическую разметку). Она содержит только те знания, которые можно непосредственно получить из семантической разметки. Поэтому в этой онтологии отсутствуют, например, такие характеристики классов и свойств, как эквивалентность, отсутствие пересечения и т.п. При построении этой онтологии могут возникать неоднозначности, связанные с классификацией экземпляров, которые требуют дополнительной проверки [9].

Её формальная модель также является частным случаем модели (1):

$$(3) \quad T_{wiki} = \langle X, R, \emptyset, T \rangle.$$

В модели (3) множество концептов строится как объединение таких элементов Wiki, как страницы и категории $X = X_{wiki_categ} \cup X_{wiki_page}$, связанные различными видами отношений с $R = \{r_{ier_cl}\} \cup \{r_{link}\} \cup \{r_{sem_prop}\}$, то есть множество классов – это множество категорий Wiki X_{wiki_categ} , между которыми существуют иерархические отношения r_{ier_cl} ; множество экземпляров – множество Wiki-страниц X_{wiki_page} , между которыми существуют ссылка r_{link} и семантические отношения $r_{sem_prop_i}$, $i = \overline{0, m}$; а множество типов данных дополняется специфическим классом – «Wiki-страница». Эта модель может быть усовершенствована с учётом таких элементов Wiki, как шаблоны, формы, специальные страницы и т.п.

4 Формальная модель основных семантических компонентов Semantic MediaWiki

Формальная модель Wiki-ресурса

$$(4) \quad W = \langle P, L \rangle$$

состоит из следующих элементов:

$P = P_{user} \cup P_{categ} \cup P_{template} \cup P_{spec}$ - множества Wiki-страниц, где

P_{user} – множество страниц, созданных пользователями,

P_{categ} – множество страниц, которые описывают категории,

$P_{template}$ – множество страниц, которые описывают шаблоны,

P_{spec} – множество других специальных страниц;

$L = \{ "link" \}$ – множество из одного элемента, который описывает ссылку одной Wiki-страницы ресурса на другую Wiki-страницу того же ресурса (хотя в Wiki-ресурсах предусмотрены и ссылки на другие виды страниц, в рамках данной модели они не учитываются).

Формальная модель семантически обогащённых Wiki-ресурсов является более сложной по сравнению с (4) и расширяет её элементами, связанными с семантическими свойствами:

$$(5) \quad W_s = \langle P_s = P \cup P_{sem_prop} \cup P_{spec}, L_s = L \cup L_{sem_prop} = \{l_i\} \rangle,$$

где P_{sem_prop} – множество страниц, которые описывают семантические свойства Wiki-страниц, некоторые из которых являются семантически определёнными ссылками на другие Wiki-страницы: $P_{sem_prop_page} \subseteq P_{sem_prop}$, а другие связывают страницы со значениями разных типов данных (эти типы данных определяются на страницах семантических свойств).

4.1 Использование онтологий для семантизации Wiki-ресурса

Если на Wiki-ресурсе обрабатываются большие объёмы знаний (например, электронная версия Большой украинской энциклопедии – e-BUE – предусматривает десятки тысяч статей), то возникает потребность в автоматизации этой обработки на основе подходов, методов и средств, которые используются сейчас для менеджмента знаний.

В частности, предлагается использовать методы и инструменты онтологического анализа, рассматривая элементы семантической разметки как классы, экземпляры и отношение онтологии соответствующей ПрО.

Для этого нужно преобразовать элементы онтологии в конструкции Semantic MediaWiki (категории и семантические свойства). На следующем этапе необходимо обеспечить возможность уточнения и усовершенствования начальной онтологии ПрО на основе анализа семантически размеченных Wiki-ресурсов, то есть разработать метод преобразования конструкций Semantic MediaWiki в онтологию, представленную на языке OWL. Далее следует проанализировать полученную онтологию, оценить её свойства и соответствие представлениям пользователя относительно ПрО. Итеративное повторение этих действий должно обеспечить формирование адекватной онтологии ПрО, которую можно использовать в разных ИИС как БЗ.

4.2 Сопоставление онтологий и семантических Wiki-ресурсов

В терминах описанных выше формальных моделей (1) и (5) можно достаточно просто определить некоторые соответствия между элементами онтологии ПрО и элементами страниц Semantic MediaWiki. Часть таких соответствий можно выявлять автоматически, без участия эксперта, но некоторые неоднозначны и требуют дополнительных уточнений. Более детально это описано в [8].

Важно отметить, что знания, содержащиеся в онтологии ПрО и в аналогичном семантическом Wiki-ресурсе, не являются полностью эквивалентными. Semantic MediaWiki не содержит сведений о характеристиках классов и свойств (в частности, об эквивалентности классов и свойств, их пересечении, об их области значения и определения), а в онтологии не отображаются сведения о страницах, использующих одни и те же шаблоны (т.е. описывающих ИО одного типа). Поэтому целесообразно использовать и регулярно синхронизировать оба эти источника знаний о ПрО. При изменении Wiki-ресурса следует усовершенствовать онтологию ПрО, а при изменении онтологии – соответственно вносить изменения в семантическую разметку Wiki-страниц.

5 Усовершенствование онтологии ПрО на основе сведений из семантически размеченных Wiki-ресурсов

Данный метод предназначен для усовершенствования онтологии ПрО, которая использовалась для семантической разметки анализируемого Wiki-ресурса. Основой для этого могут стать зависимости между элементами онтологии и семантическими элементами Wiki-ресурсов, приведенные в таблице 1. Если нужно усовершенствовать какую-либо другую онтологию, то вначале надо выполнить её сопоставление с той онтологией, которая использовалась для разметки.

Таблица 1 – Отображение между основными элементами онтологий и Wiki-ресурсов

| Семантические Wiki | Онтология | Из Wiki в онтологию | Из онтологии в Wiki |
|--|------------------------------|--|--|
| Категория Wiki | Класс онтологии | Однозначное $P_{categ} \rightarrow X_{cl}$ | Многозначное $X_{cl} \rightarrow P_{categ} \cup P_{template}$ |
| Иерархия категорий Wiki | Иерархия классов онтологии | Многозначное | Однозначное |
| Wiki-страница | Экземпляр класса онтологии | Многозначное $P_{user} \rightarrow X_{ind}$ | Однозначное $X_{ind} \rightarrow P_{user}$ |
| Ссылка на Wiki-страницу | Объектное свойство | Взаимно-однозначное $L = \{ "link" \} \rightarrow R$ | Взаимно-однозначное $R \rightarrow L = \{ "link" \}$ |
| Семантическое свойство типа «страница» | Объектное свойство онтологии | Взаимно-однозначное $P_{sem_prop_page} \rightarrow \{r_i\}$ | Взаимно-однозначное $\{r_i\} \rightarrow P_{sem_prop_page}$ |
| Семантическое свойство любого другого типа | Свойство данных онтологии | Взаимно-однозначное $P_{sem_prop} \rightarrow \{p_i\}$ | Взаимно-однозначное $\{p_i\} \rightarrow P_{sem_prop}$ |
| Шаблон | Класс онтологии | Однозначное $P_{template} \rightarrow X_{cl}$ | Многозначное $X_{cl} \rightarrow P_{categ} \cup P_{template}$ |

Если для семантической разметки Wiki-ресурса W_s используются элементы онтологии определённой ПрО O_{Sw_0} , то в дальнейшем контент такого семантически размеченного ресурса W_s может стать базой для усовершенствования именно этой онтологии O_{Sw_0} и построения на его основе новой онтологии O_{Sw_r} . Для этого необходимо применять методы автоматизированного сопоставления элементов онтологии и Wiki-ресурса, а также определить те неоднозначности, в которых такое сопоставление не является взаимно-однозначным, а требует участия человека-эксперта.

Семантический Wiki-ресурс рассматривается как множество *семантически размеченных* естественно-языковых текстов, пригодных для автоматического анализа, например, для поиска сведений, которые интересуют пользователя и связаны с определёнными понятиями ПрО, отображённой в онтологии O_{Sw_0} .

На первом этапе на вход в систему подаются:

- начальная онтология ПрО O_{Sw_0} , которая использовалась для разметки Wiki-ресурса W_s , в которой содержатся наиболее очевидные для пользователя понятия и связи между ними;
- W_u – выбранный пользователем набор семантически размеченных Wiki-страниц, которые характеризуют ПрО, интересующую пользователя, $W_u = \{w_i\}, i = \overline{1, n}, W_u \subseteq W_s$.

Заметим, что множества $O_{S_{w_0}}$ и T не должны быть пустыми.

Расширенная онтология ПрО $O_{S_{w_r}}$ строится в процессе анализа страниц из W_u , выбранных пользователем в соответствии с представлениям о своих информационных потребностях (например, это могут быть все Wiki-страницы, созданные позднее определённой даты; Wiki-страницы определённой категории или Wiki-страницы, отобранные по семантическому запросу). В процессе семантико-синтаксического анализа этих страниц определяются классы онтологии, экземпляры классов и отношения между классами и экземплярами.

Семантические свойства определяют семантику связей между страницами в выбранном множестве и соответствуют объектным свойствам экземпляров класса онтологии, которые отвечают Wiki-страницам.

Алгоритм усовершенствования онтологии выполняется в процессе обработки каждой Wiki-страницы w_i из W_u . При анализе w_i необходимо выполнить следующие действия:

- в онтологии $O_{S_{w_0}}$ найти экземпляр e_i , которая соответствует странице w_i ;
- если e_i не найден, то создать его;
- если w_i относится к категории c_j , $w_i \in c_j$ (то есть страница содержит текст [[категория:: c_j]], но в онтологии $O_{S_{w_0}}$ отсутствует такой класс, тогда надо добавить к онтологии этот класс (после подтверждения пользователя относительно того, что такой класс относится к интересующей его ПрО: следует отметить, что страница может быть отмечена различными категориями, которые не важны для рассматриваемой ПрО, и тогда их можно не добавлять в данную онтологию);
- если w_i относится к категории c_j , $w_i \in c_j$ (то есть страница содержит текст [[категория:: c_j]], а в онтологии $O_{S_{w_0}}$ присутствует такой класс, но экземпляр онтологии, соответствующий Wiki странице w_i , не относится к этому классу онтологии, тогда надо добавить в онтологии связь этого экземпляра с этим классом;
- если Wiki-страница w_i имеет семантическое свойство типа «Страница» со значением из $w_k \in W_u$ (т.е. связана с одной из выбранных Wiki-страниц), тогда соответствующий экземпляр онтологии тоже получает соответствующее объектное свойство со значением соответствующего элемента.

Считается, что в процессе усовершенствования Wiki-ресурса могут создаваться новые Wiki-страницы, изменяться категоризация Wiki-страниц и добавляться новые значения семантических свойств, но сам набор категорий и семантических свойств не изменяется и фиксируется с помощью онтологии ПрО. Таким образом, усовершенствованная онтология может пополняться новыми экземплярами и связями между экземплярами, но её классы и набор объектных свойств не зависят от Wiki-ресурса. Это обеспечивает унификацию в семантической разметке и позволяет предотвратить дублирования в названиях свойств и категорий.

Если возникает потребность в усовершенствовании какой-либо другой онтологии O_1 , отличной от $O_{S_{w_0}}$ (например, другой версии этой же онтологии $O_{S_{w_1}}$), то нужно выполнить следующие операции:

- используя описанный выше метод, по Wiki-ресурсу Ws усовершенствовать онтологию $O_{S_{w_0}}$ и построить онтологию $O_{S_{w_r}}$;

- с помощью существующих инструментов онтологического анализа выполнить интеграцию онтологии O_{Sw_0} с онтологией O_1 и построить онтологию O_{1r} .

Частным случаем такой задачи может быть построение онтологии, которая непосредственно касается определённой текущей задачи пользователя (например, связанной с семантическим поиском) и содержит подмножество её классов, экземпляров и свойств. Описанные выше модели Wiki-ресурсов и онтологий ПрО, а также метод их интеграции были апробированы на ряде задач, связанных с научно-исследовательской и учебно-методической деятельностью.

6 Построение типичных информационных объектов e-BUE на основе онтологического анализа

На первом этапе создания Wiki-ресурса целесообразно построить (или выбрать) начальную онтологическую модель ПрО, которая будет использоваться для семантической разметки Wiki-ресурса, то есть сформировать основные типы ИО, которые описываются соответствующими Wiki-страницами, и отношения между ними. ИО – это информационная модель объекта определённой ПрО, которая определяет структуру, атрибуты, ограничение целостности и, возможно, поведение этого объекта. В Wiki-ресурсе ИО – это страница, которая однозначно идентифицируется своим уникальным именем.

Создавая Wiki-ресурсы, целесообразно определять характерные для данной ПрО типичные ИО. Это позволяет унифицировать представления сведений о подобных сущностях, использовать в них одинаковые рубрикаторы и порядок представления информации. Кроме того, это позволяет более эффективно осуществлять поиск и навигацию. В частности, разрабатывая энциклопедии и справочники на основе Wiki, целесообразно создавать категории, которые объединяют ИО с одинаковой структурой и похожими элементами. Для них можно создавать прототип страницы с унифицированным размещением материала – это упрощает восприятие информации пользователями.

Рассмотрим это на примере e-BUE – портальной версии «Большой украинской энциклопедии» («Big Ukrainian Encyclopedia»). Этот Интернет-ресурс работает по технологии Wiki, в частности, на основе Semantic MediaWiki. По жанру e-BUE является универсальной научно-популярной мультимедийной энциклопедией. Большинство статей e-BUE являются авторскими, они написаны ведущими специалистами Украины в соответствующих областях. Наличие у материала высококвалифицированного автора обеспечивает достоверность и качество представляемой информации. Публикация на ресурсе e-BUE работает как на распространение общественно важных знаний и развитие национального энциклопедического проекта, так и на популяризацию научного творчества автора. Энциклопедия является фактором развития и распространения украинского языка, знаний об Украине, её культурного наследия и научных достижений.

В результате обобщения и дальнейшей типологизации ИО энциклопедические статьи в e-BUE распределены по тематическим группам: «Персоналии», «Природа», «Цивилизация», «Области знаний», которые затем подразделяются на подкатегории, например, «Государственные деятели», «Украинские композиторы», «Вид спорта», «Издательства» и т.п. Классификатор «Области знаний» представляет научную классификацию статей, но в нём используются понятия, не всегда понятные пользователям, которые не являются специалистами. Типы ИО могут упорядочиваться иерархически, например, тип ИО «Государственные деятели Украины» является подтипом ИО «Государственные деятели». При этом используется несколько независимых классификаторов, и одна и та же статья может быть отнесена к нескольким различным тематическим группам.

Применительно к e-BUE, *тип ИО* – это группа статей e-BUE, которая имеет общие свойства. С точки зрения поиска целесообразно группировать в один тип статьи одной категории (например, «Государственные деятели»), множества категорий (например, «Физика» и «Научные работники») или статьи с подобными подзаголовками разделов. Чтобы упростить навигацию в БЗ, которой с точки зрения поисковых процедур является e-BUE, целесообразно выделить такие типы ИО, наиболее часто интересующие пользователей, которые не знают точного названия нужной статьи или интересуются несколькими близкими по тематике статьями. Наличие разнообразных типов ИО значительно улучшит качество поиска в e-BUE, но важно зафиксировать иерархию типов ИО, так как это позволяет строить сложные запросы и страницы-интеграторы. Запросы, использующие семантику, позволяют связывать значения определённых семантических свойств с определённым типом ИО. Например, семантическое свойство «Место работы» должно иметь значением статью, которая относится к типу ИО «Организация» или «Учреждение».

Следует отметить, что типы ИО могут пересекаться. Это связано с тем, что одна статья может относиться одновременно к нескольким разным категориям, потому что типизация ИО является не таксономией, а фолксономией, то есть отображает ненаучный подход к классификации ИО, а только фиксирует общеупотребительные представления пользователей. Она, кроме того, не является исчерпывающей, и потому существует много ИО, для которых не построен специфический тип ИО.

При разработке системы категорий такого ресурса, как e-BUE, целесообразно использовать для этого два подхода: «сверху вниз» и «снизу вверх».

Подход «Сверху вниз» – от категорий верхнего уровня к их подкатегориям, которые более детально характеризуют содержимое (например, для категории «Социальные коммуникации» можно создать подкатеорию «Социальная информатика».

Подход «Снизу вверх» – от статей к типам ИО. Создавая новые статьи или перерабатывая существующие, можно пытаться выбирать для них тип ИО с точки зрения пользователей. Например, по статье «Австро-прусская война 1866» можно создать тип ИО «Война», а по статье «Автореферат» – тип ИО «Документ». Потом эти типы можно использовать и для описания других статей e-BUE.

Технология Wiki позволяет создавать и использовать для ИО произвольное количество категорий, которые строятся по разными принципам классификации. Поэтому целесообразно выделять отдельные подкатегории и для таких ИО, которые имеют общий набор семантических свойств. Если семантические свойства являются объектными, то есть ссылками на другие ИО, то целесообразно определить категорию таких ИО. Например, если для персоналий обычно выделяют такие семантические свойства, как «Место рождения» и «Место учебы», то для их определения целесообразно вводить категории «Город» и «Учебное заведение».

Использование онтологического анализа для построения типов ИО e-BUE позволяет явным образом определить семантику этих типов, определить отношения между ними, проанализировать, какие типы ИО оказываются избыточными или дублируют друг друга, а какие необходимо дополнить. Кроме того, онтология позволяет явным образом определить семантику отношений между типами ИО, что отвечает типичным схемам статей e-BUE из разных направлений знаний.

Визуализация онтологии ИО средствами Protégé приведена на рисунке 1.

Создавая онтологию ИО O_{Sw} , нужно выполнить следующие операции.

- Построить и иерархически упорядочить классы онтологии $c_i, i = \overline{1, p}, c_i \in C_{IO}$, которые соответствуют категориям e-BUE, связанным с типичными ИО.

- Построить набор объектных свойств онтологии $\{r_i\}$, которые соответствуют семантическим свойствам Wiki-страниц $P_{sem_prop_page}$ любой из этих категорий, значениями которых является другие Wiki-страницы (возможно, определённой категории или множества категорий).
- Построить набор свойств данных онтологии $\{p_i\}$, которые соответствуют семантическим свойствам Wiki-страниц P_{sem_prop} любой из этих категорий, значениями которых являются конкретные значения определенного типа.
- Создать экземпляры классов онтологии X_{ind} , которые соответствуют определенным Wiki-страницам P_{user} , которые относятся к этим категориям.
- Если нужно, установить отношения между экземплярами и другими классами (для Wiki-страниц, которые относятся к более чем одной категории).

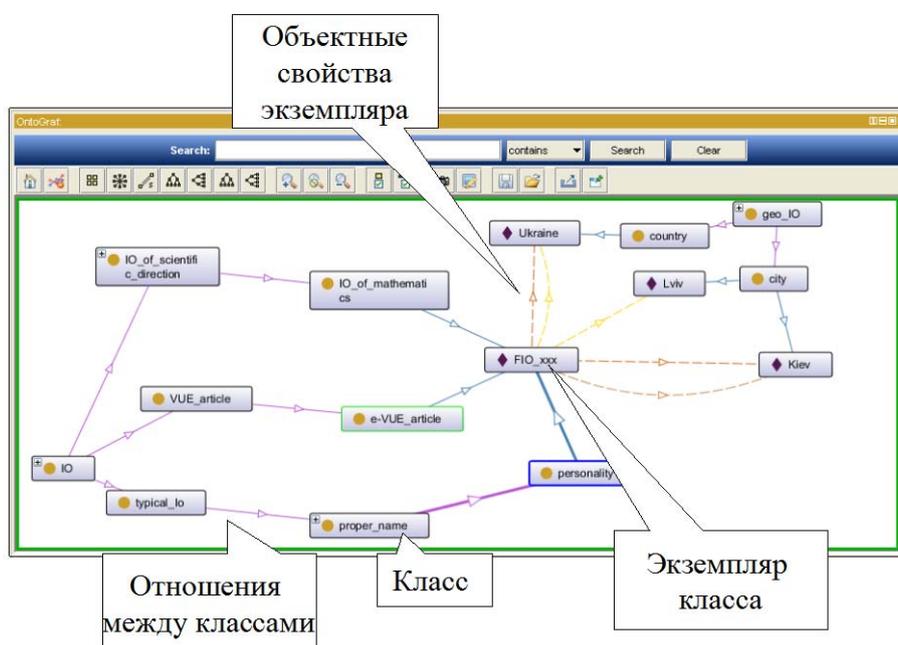


Рисунок 1 – Визуализация онтологии информационного объекта для e-BUE

Такая онтология позволяет более точно определить характеристики семантических свойств, используемых в e-BUE, и формализовать связи между ИО (рисунок 2).

Описанный выше метод позволяет довольно просто превратить такую онтологию в систему категорий Semantic MediaWiki. Кроме того, анализ объектных свойств и свойств данных тех классов онтологии, которые связаны с наиболее часто используемыми ИО, позволяет построить (или усовершенствовать уже существующие) шаблоны, которые должны применяться на Wiki-страницах, которые отнесены к этим типам.

Семантические свойства и типичные ИО позволяют значительно улучшить поиск и навигацию в Wiki-ресурсе.

Средства поиска, реализованные в e-BUE, можно разделить на следующие группы:

- по ключевым словам (для e-BUE – по названию статьи);
- по тематике (для e-BUE – по областям знаний);
- по типу ИО (для e-BUE – поиск персоналий и понятий, статей, связанных с природой и цивилизацией);

- по семантике ИО (для e-BUE – по значениям семантических свойств Wiki-страниц, описанных в онтологии Про).

Для того чтобы формализовать семантические свойства Wiki-страниц типичных ИО, нужно построить их формализованные описания и указать область значений этих свойств. Например, для типичных ИО категории «Персоналия» характерные следующие семантические свойства, приведённые в таблице 2.

Таблица 2 – Семантические свойства страницы категории «Персоналия» в e-BUE (фрагмент)

| Название свойства | Тип данных |
|------------------------------------|--|
| Автор BUE | Ссылка на соответствующую страницу e-BUE |
| Фамилия | Текстовая строка |
| Имя | Текстовая строка |
| Отчество | Текстовая строка |
| Псевдоним | Текстовая строка |
| Дата рождения | Дата ДД.ММ или Текстовая строка |
| Год рождения | Текстовая строка или Число |
| Место рождения | Ссылка на соответствующую страницу e-BUE категории «Административно-территориальная единица» или Текстовая строка |
| Дата смерти | Дата ДД.ММ или Текстовая строка |
| Год смерти | Текстовая строка или Число |
| Место обучения | Ссылка на соответствующую страницу e-BUE категории «Учебное заведение» или категории «Административно-территориальная единица» или Текстовая строка |
| Научная степень | Ссылка на соответствующую страницу e-BUE категории «Научная степень» или Текстовая строка |
| Ученое звание | Ссылка на соответствующую страницу e-BUE категории «Ученое звание» или Текстовая строка |
| Воинское звание | Ссылка на соответствующую страницу e-BUE категории «Воинское звание» или Текстовая строка |
| Титул | Ссылка на соответствующую страницу e-BUE категории «Титул» или Текстовая строка |
| Профессия | Ссылка на соответствующую страницу e-BUE категории «Профессия» или Текстовая строка |
| Школа, направление, группа, партия | Ссылка на соответствующую страницу e-BUE или Текстовая строка |
| Место работы | Ссылка на соответствующую страницу e-BUE категории «Организация или учреждение» или категории «Административно-территориальная единица» или Текстовая строка |
| Сотрудничал с | Ссылка на соответствующую страницу e-BUE категории «Персоналия» или Текстовая строка |
| Повлиял на | Ссылка на соответствующую страницу e-BUE категории «Персоналия» или Текстовая строка |
| Является автором | Ссылка на соответствующую страницу e-BUE категории «Продукт деятельности» или Текстовая строка |
| Является потомком | Ссылка на соответствующую страницу e-BUE категории «Персоналия» или Текстовая строка |
| Является родственником | Ссылка на соответствующую страницу e-BUE категории «Персоналия» или Текстовая строка |
| Веб-сайт | Текстовая строка |

Для отображения этих свойств на Wiki-странице целесообразно использовать соответствующий Wiki-шаблон, который позволяет унифицировать представление этой информации на всех страницах, относящихся к этому типу ИО (рисунок 3).

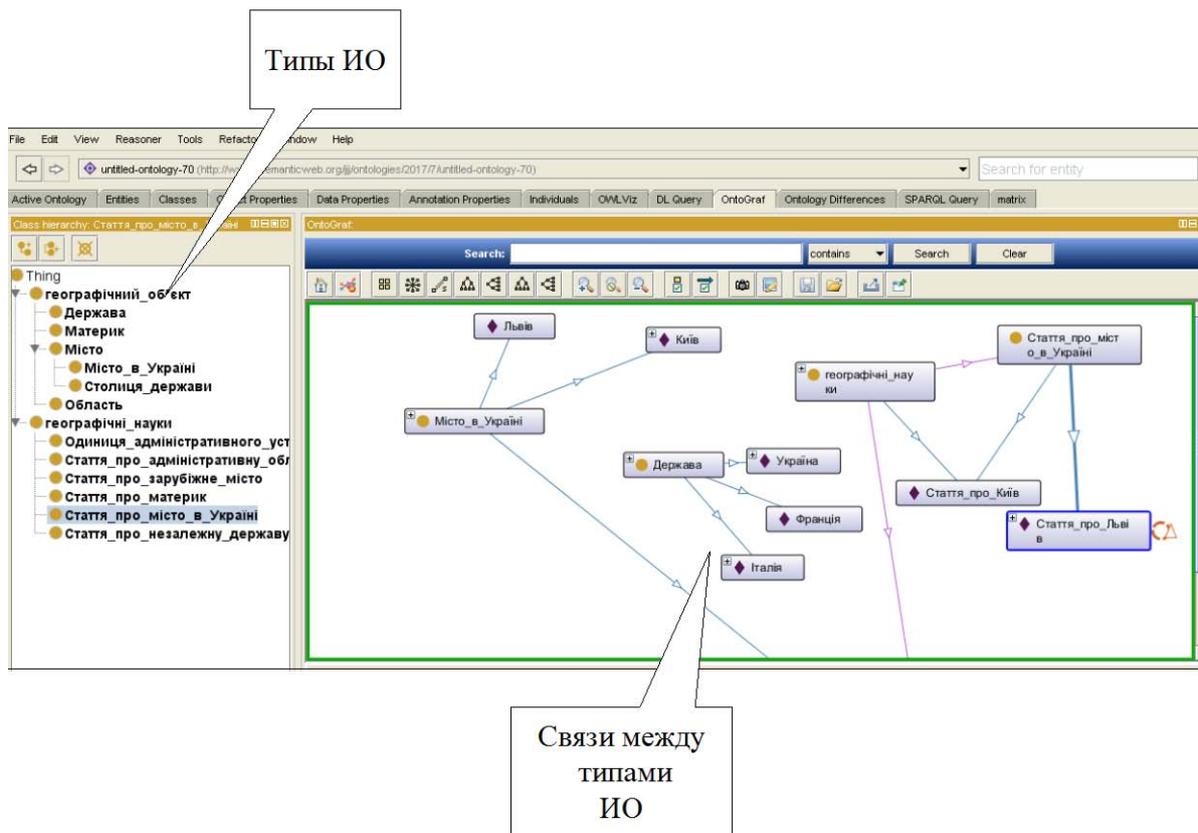


Рисунок 2 – Связи между типами информационных объектов в e-BUE

The screenshot shows a Wiki page for 'Альберт Великий' (Albert the Great). The main text describes him as a medieval philosopher and theologian. A table below provides biographical data. Callouts identify the page as a 'Персоналия' category page, the table as a 'Шаблон типичного ИО' (typical IO template), and the table's content as 'Семантические свойства' (semantic properties) with their 'Значения семантических свойств' (values of semantic properties).

| | |
|------------------|---------------------------------|
| М'я / Псевдонім | Альберт Великий, Святий Альберт |
| Гендер | |
| Дата народження | бл. 1193 |
| Місце народження | м. Лауінген, Німеччина |
| Дата смерті | 15.11.1280 |
| Місце смерті | м. Кельн, Німеччина |
| Alma mater | Падуанський університет |
| Напрями | Богословство |

Рисунок 3 – Wiki-страница категории «Персоналия» в e-BUE

7 Использование онтологического анализа для семантической разметки стандартов информационной безопасности

Для обеспечения пользователям эффективного доступа к нормативно-правовому обеспечению информационной безопасности предлагается использовать семантические Wiki-технологии. Это обуславливается большими объёмами информации, которую надо представить, её динамичностью и мультилингвистичностью. Чтобы интегрировать сведения из разных источников, предлагается разработать онтологию ПрО «Информационная безопасность», элементы которой будут использованы для семантической разметки Wiki-страниц. Использование онтологического подхода обеспечит интероперабельность этих знаний и возможность их применения в других информационных системах.

Онтология «Информационная безопасность» (рисунок 4) базируется на наборе национальных и международных стандартов, связанных с вопросами информационной безопасности. На сегодня она содержит следующие базовые классы – «Предметная область», «Стандарт» и «Термин стандарта», для которых построены соответствующие экземпляры и подклассы. Экземпляры этих основных классов связаны объективными свойствами «Описан в стандарте», «Ссылается на стандарт», «Относится к ПрО» и т.п. Чтобы формализовать семантику отношений между элементами онтологии, которая сгенерирована по терминологическим базисам стандартов, построены специфические для ПрО объективные свойства.

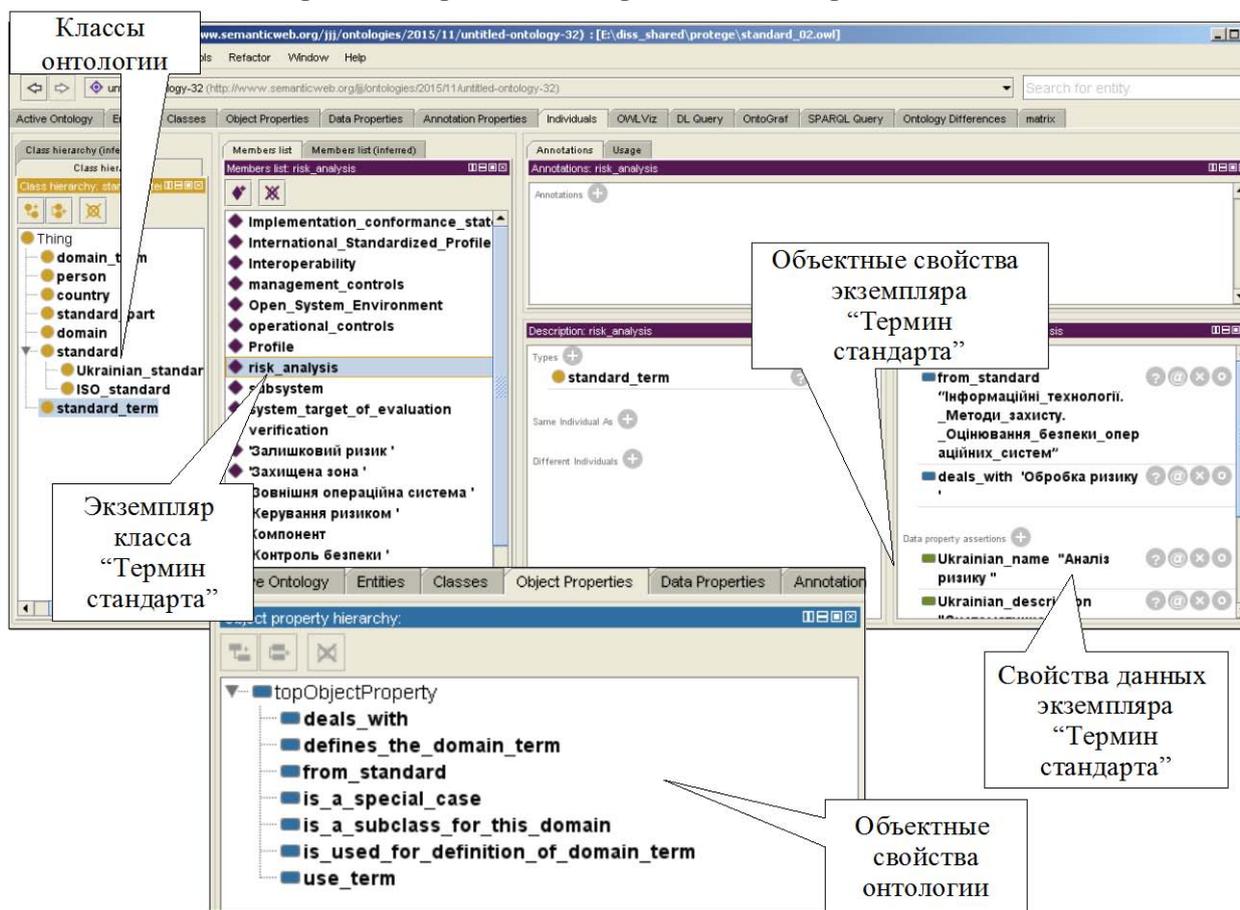


Рисунок 4 – Объективные свойства и экземпляры классов онтологии «Информационная безопасность»

На рисунке 4 приведён набор объективных свойств онтологии «Информационная безопасность», которые позволяют устанавливать связи между типичными ИО этой ПрО.

Protégé позволяет явно указать такие характеристики этих свойств, как транзитивность (например, для свойства «является подклассом для этого домена») и рефлексивность (например, для свойства «связан с»). В дальнейшем эти характеристики могут быть использованы для вывода новых знаний из тех Wiki-ресурсов, которые будут размечены соответствующими семантическими свойствами (следует ещё раз отметить, что непосредственно в Wiki нельзя задавать и обрабатывать такие характеристики свойств).

Отдельные Wiki-страницы целесообразно создать как для каждого стандарта в целом (рисунок 5), так и для отдельных их подразделов и определений. Кроме того, целесообразно использовать гиперссылки на другие Wiki-ресурсы, например, на Википедию или на e-BUE.

Такой подход значительно упрощает поиск и анализ таких документов и обеспечивает возможность их автоматической обработки. При этом пользователям не нужно будет самостоятельно отслеживать изменения в последней редакции стандарта или пересматриваться описание каждой потенциально пригодной системы – сопоставление может выполняться автоматизировано. Перспективы развития этой работы предусматривают создание глобальной семантической сети стандартов, которая свяжет:

- отдельные национальные и международные стандарты;
- объекты, которые используют эти стандарты и ссылаются на них (как материальные, так и ИО);
- специалистов, которые являются экспертами в сфере разработки стандартов, и организации разного уровня, которые поддерживают разные виды деятельности, связанной с разработкой и использованием стандартов.

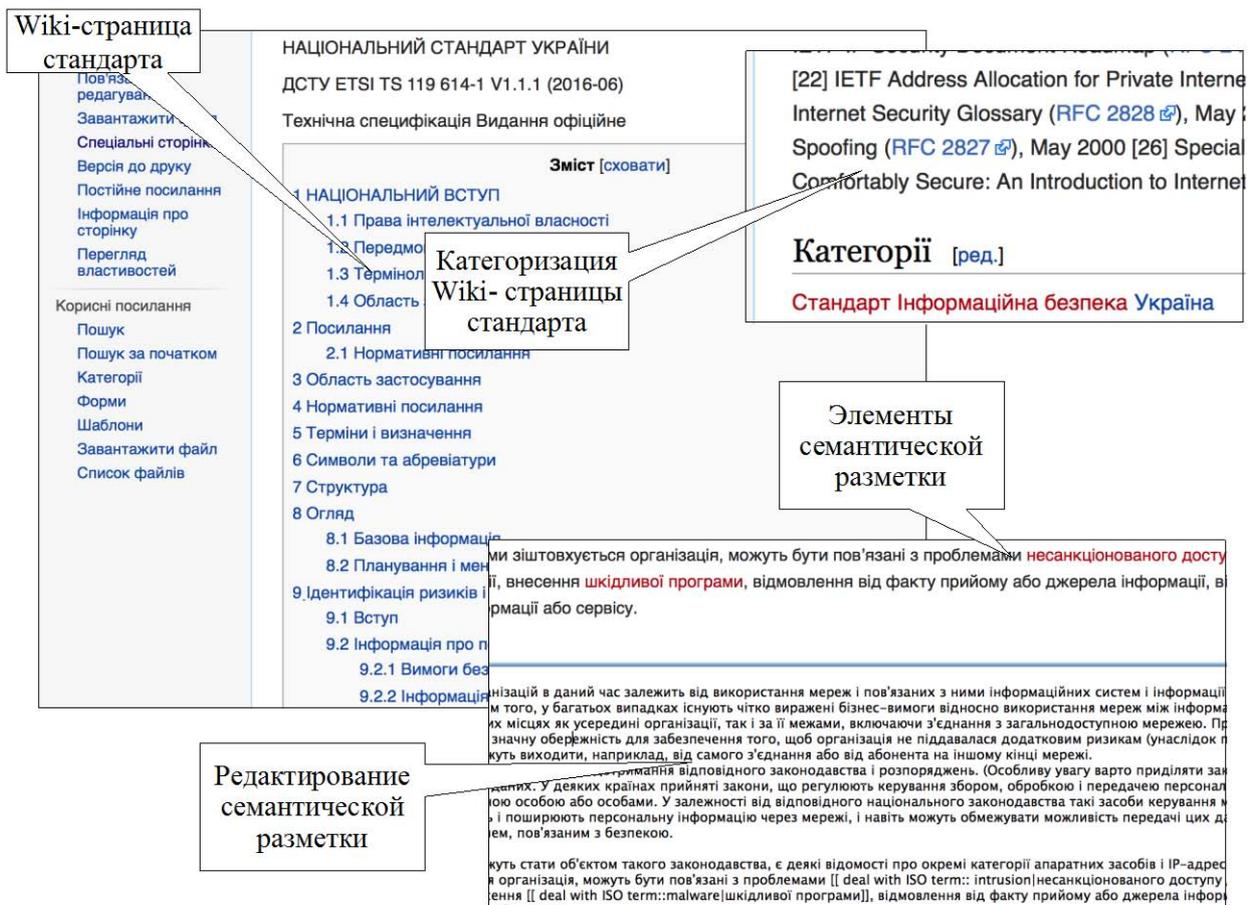


Рисунок 5 – Семантическая разметка Wiki-ресурса с помощью онтологии «Информационная безопасность»

Заключение

Одним из наиболее важных факторов эффективности интеллектуальных систем, ориентированных на функционирование в Web, является способность этих систем находить и использовать знания, содержащиеся в этой среде. Особый интерес вызывают те Web-ресурсы, из которых можно автоматизированно извлекать информацию для пополнения и усовершенствования онтологий, и технологии, обеспечивающие создание и функционирование таких ресурсов. В данной работе показано взаимодействие Web-ресурсов, разработанных на основе семантических Wiki-технологий с использованием онтологического анализа.

Предложенный подход позволяет использовать преимущества онтологического представления информации для управления знаниями в ИИС (например, в различных энциклопедиях и Wiki-справочниках), а также обеспечивает автоматизированный и простой для пользователей способ пополнения и усовершенствования онтологий Про.

Примеры апробации предложенного подхода подтверждают его эффективность для создания распределённых интеллектуальных систем. При разработке электронной версии Большой украинской энциклопедии для создания системы типичных ИО была разработана онтология, в которой были зафиксированы свойства этих объектов и отношения между ними. На основании описанных в работе сопоставлений по этой онтологии были сформированы категории и семантические свойства, которые в дальнейшем использовались при создании Wiki-страниц, соответствующих статьям энциклопедии. Такой подход позволил значительно упростить и ускорить согласование терминологии, используемой при разметке, между авторами, редакторами и техническим персоналом e-BUE.

Использование онтологии позволило формально описать характеристики этих семантических свойств, визуализировать связи между категориями и свойствами, значительно облегчить создание семантических запросов, эффективно и динамично интегрировать информацию, содержащуюся в этом ресурсе. Этот пример продемонстрировал целесообразность интеграции онтологий с Wiki-ресурсами большого объёма и сложной структуры.

На примере разработки Wiki-ресурса, содержащего стандарты, связанные с информационной безопасностью, показана полезность использования онтологий при обработке мультилингвистических документов, связи между которыми позволяют устанавливать термины онтологии.

Список источников

- [1] **Krotzsch M.** Semantic MediaWiki / M. Krotzsch, D. Vrandečić, M. Volkel. – <http://c.unik.no/images/6/6d/SemanticMW.pdf>.
- [2] OWL Web Ontology Language Semantics and Abstract Syntax. Section 2. Abstract Syntax – <http://www.w3.org/TR/owl-semantics/syntax.html>.
- [3] W3C Semantic Web Activity. – <http://www.w3.org/2001/sw/Activity/>.
- [4] SPARQL Query Language for RDF. W3C Recommendation, 2008 – <http://www.w3.org/TR/rdf-sparql-query/>.
- [5] **Gruber T.R.** Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human-Computer Studies, 1995, V. 43, Issues 5-6. - P.907-928.
- [6] **Obr L.** The evaluation of ontologies / L.Obr, W. Ceuster, I. Mani, S. Ra, B.Smith // In Semantic web: Revolutionizing Knowledge Discovery in the Life Sciences, New York: Springer Verlag, 2006, 139-158. – <https://philpapers.org/archive/OBRTEO-6.pdf>.
- [7] Protégé. – <http://protege.stanford.edu/>.
- [8] **Rogushina J.** Semantic Wiki resources and their use for the construction of personalized ontologies // CEUR Workshop Proceedings 1631, 2016, - P.188-195.
- [9] **Rogushina J.** Analysis of Automated Matching of the Semantic Wiki Resources with Elements of Domain Ontologies // International Journal of Mathematical Sciences and Computing (IJMSC), Vol.3, No.3, 2017. – P.50-58. – <http://www.mecs-press.org/ijmsc/ijmsc-v3-n3/IJMSC-V3-N3-5.pdf>

DEVELOPMENT OF DISTRIBUTED INTELLIGENT SYSTEMS ON BASE OF ONTOLOGICAL ANALYSIS AND SEMANTIC WIKI TECHNOLOGIES

J.V. Rogushina

*Institute of Software Systems of the National Academy of Sciences of Ukraine, Kyiv, Ukraine
ladamandraka2010@gmail.com*

Abstract

The necessity of automated knowledge acquisition by intelligent information systems from the distributed Web environment is considered. The feasibility of using ontological analysis to represent such knowledge is assessed. The necessity of replenishment of this knowledge from more dynamic information resources that provide automated processing of information (such as semantic Wiki resources) is grounded. Such information processing requires the development of formal model and methods for mapping of Wiki elements to ontology elements. The expressive possibilities of the Semantic MediaWiki technological environment that can be used for semantic marking of natural language and multimedia information are analyzed. The method of semantic markup of Wiki-resources by elements of the ontological model of the domain is proposed. This markup is based on the establishment of correspondences between the elements of the formal model of such ontology (classes, individuals, object and data properties) and the elements of the Semantic MediaWiki (categories, semantic properties, references). Method for improving the domain ontology is developed on the basis of a Wiki resource semantically marked by elements of this ontology. Approbation of the developed models and methods is resulted on example of development of the electronic version of the Great Ukrainian Encyclopedia on the Semantic MediaWiki platform. The novelty of the work lies in the way of comparing the elements of the ontological model of the domain with the elements of the semantic Wiki-resource, and also in the method of improving the ontology of the domain that is used for semantic markup of the Wiki resource, based on the changes and additions that are made to this resource.

Keywords: *Wiki-resource, ontology, information object, semantic markup.*

Citation: *Rogushina, J.V.* Development of distributed intelligent systems on base of ontological analysis and semantic wiki technologies [In Russian]. *Ontology of designing.* 2017; 7(4): 453-472. DOI: 10.18287/2223-9537-2017-7-4-453-472.

References

- [1] **Krotzsch M.** Semantic MediaWiki / M. Krotzsch, D. Vrandečić, M. Volkel. – <http://c.unik.no/images/6/6d/SemanticMW.pdf>.
- [2] OWL Web Ontology Language Semantics and Abstract Syntax. Section 2. Abstract Syntax – <http://www.w3.org/TR/owl-semantics/syntax.html>.
- [3] W3C Semantic Web Activity. – <http://www.w3.org/2001/sw/Activity/>.
- [4] SPARQL Query Language for RDF. W3C Recommendation, 2008 – <http://www.w3.org/TR/rdf-sparql-query/>.
- [5] **Gruber TR.** Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human-Computer Studies*, 1995, V. 43, Issues 5-6. - P.907-928.
- [6] **Obr L.** The evaluation of ontologies / L.Obr, W. Ceuster, I. Mani, S. Ra, B.Smith // *In Semantic web: Revolutionizing Knowledge Discovery in the Life Sciences*, New York: Springer Verlag, 2006, 139-158. – <https://philpapers.org/archive/OBRTEO-6.pdf>.
- [7] Protégé. – <http://protege.stanford.edu/>.
- [8] **Rogushina J.** Semantic Wiki resources and their use for the construction of personalized ontologies // *CEUR Workshop Proceedings 1631*, 2016, - P.188-195.
- [9] **Rogushina J.** Analysis of Automated Matching of the Semantic Wiki Resources with Elements of Domain Ontologies // *International Journal of Mathematical Sciences and Computing (IJMSC)*, Vol.3, No.3, 2017. – P.50-58. – <http://www.mecspress.org/ijmsc/ijmsc-v3-n3/IJMSC-V3-N3-5.pdf>

Сведения об авторе



Рогущина Юлия Витальевна, 1967 г. рождения. Окончила факультет кибернетики Киевского государственного университета им. Т.Г. Шевченко, кандидат физико-математических наук. Старший научный сотрудник Института программных систем НАН Украины. Автор более 200 научных публикаций, среди которых монографии и учебники в области онтологического анализа, семантического поиска, интеллектуальных агентов и менеджмента знаний.

Rogushina Yulia Vitalyevna (b. 1967). Graduated from the Faculty of Cybernetics of Kiev State University named after TG Shevchenko, Candidate of Physical and Mathematical Sciences. Senior Researcher of the Institute of Software Systems of the National Academy of Sciences of Ukraine. Author of more than 200 scientific publications, including monographs and text-

books in the field of ontological analysis, semantic search, intelligent agents and knowledge management.