УДК 004.853, 004.55

ИСПОЛЬЗОВАНИЕ ОНТОЛОГИЧЕСКОГО АНАЛИЗА ДЛЯ СОЗДАНИЯ СОВРЕМЕННЫХ ЭНЦИКЛОПЕДИЧЕСКИХ ПОРТАЛОВ

Ю.В. Рогушина

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

Аннотация

Обосновывается целесообразность разработки интеллектуальных, семантически структурированных информационных ресурсов Web, предусматривающих как их использование человеком, так и пригодность для автоматизированного анализа. Анализируются преимущества и недостатки технологии Wiki и перспективы её семантического расширения с использованием онтологического анализа. Предложена формальная модель онтологии Wiki-ресурса, на основе которой строится онтология задачи пользователя, предназначенная для использования во внешних интеллектуальных приложениях – например, при семантическом поиске. Описаны основные этапы построения этой онтологии. Рассматривается использование предложенных в работе моделей и методов на примере создания онлайновых энциклопедических изданий, объединяющих свободный доступ к материалам через среду Web с высоким уровнем доверия к контенту, разработанному экспертами. Сформированы базовые требования, предъявляемые к программному обеспечению таких проектов. Обосновывается необходимость семантизации ресурсов, анализируются основные направления развития функционала семантизированных электронных энциклопедий. На основе анализа выразительной мощности средств представления и обработки знаний, которые основываются на Wiki-технологиях, обосновывается необходимость расширения их методами искусственного интеллекта и технологиями Semantic Web. Для разработки портальной версии Большой украинской энциклопедии (е-ВУЕ) построена онтологическая модель энциклопедии. Описаны основные категории и семантические свойства типичных информационных объектов, которые используются в е-ВУЕ и могут применяться для поиска и интеграции информации.

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

Цитирование: *Рогушина, Ю.В.* Использование онтологического анализа для создания современных энциклопедических порталов / Ю.В. Рогушина // Онтология проектирования. – 2019. – Т.9, №1(31). – С. 70-84. – DOI: 10.18287/2223-9537-2019-9-1-70-84.

Введение

Главное направление в борьбе с информационным взрывом – переход от сохранения и обработки данных к накоплению и обработке знаний. Это определяет актуальность разработки инновационных технологий для поддержки функционирования распределённых интеллектуальных информационных систем (ИИС), обеспечивающих приобретение, хранение и использование знаний различных предметных областей (ПрО). При этом возникает проблема, связанная с наличием открытых и надёжных источников знаний, к которым могут обращаться ИИС.

Преобладающая часть современных ресурсов Web не является семантизированными, но количество информационных ресурсов (ИР), содержащих семантическую разметку и разнообразные метаописания, постоянно возрастает [1]. Пользователям сложно находить такие ресурсы в общей массе ИР, несмотря на то, что существует много поисковых систем, ориентированных именно на поиск структурированных ИР, например, для выявления, индексации

и запросов документов в формате RDF [2] или OWL [3]. Используя такие системы (например, Corese [4], ONTOSEARCH2 [5]), можно найти конкретный документ или онтологию, но сложно построить множество документов, которые соответствуют какой-либо конкретной задаче, так как непосредственные ссылки между такими ИР, как правило, отсутствуют или не являются наглядными.

Поэтому удобнее использовать базы знаний, построенные на Wiki-платформе [6]. Такие системы применяют стандартизированные средства для представления семантической разметки (с помощью системы категорий и свойств). Эти элементы можно легко распознавать даже в тех случаях, если разные информационные объекты (ИО) получены из разных ресурсов. В семантических Wiki-ресурсах всегда есть разнообразные средства для внутренней навигации и поиска, и это позволяет пользователю довольно быстро определить набор Wiki-страниц, связанных с его задачей. Важный фактор – наличие разнообразных семантических Wiki-ресурсов, количество, объём и качество которых постоянно увеличиваются. В случае, если информации в семантических Wiki недостаточно, их довольно легко дополнить сведениями с несемантизированных Wiki (например, из Википедий или Wiki-справочников). Из таких ресурсов можно получить меньше семантической информации, но в сочетании с семантизированными они позволяют довольно корректно описать произвольную проблему [7].

Следует отметить, что автоматизированное приобретение знаний значительно более эффективно осуществляется для тех ИР, которые имеют формализованную структуру и используют семантическую разметку контента (в отличие от естественноязыковых или мультимедийных ИР, для которых извлечение знаний требует большего участия человека). В то же время тенденции развития ресурсов Web показывают, что в поиске источников знаний целесообразно ориентироваться на распространённые и понятные для пользователей формы представления информации.

Этим условиям удовлетворяют семантические Wiki-ресурсы, которые довольно легко интегрировать с разнообразными ИИС. Wiki представляет собой технологию коллективного создания и использования распределённых ресурсов. Она всё чаще воспринимается как новый тип коллаборативной технологии, которая может повлиять на управление знаниями, а также поддерживать их создание и совместное использование. Из различных программных средств для разработки Wiki-ресурсов наиболее широко используется MediaWiki. Именно на этом свободном программном обеспечении базируются многие всемирно известные проекты энциклопедий и справочников, такие как Wikipedia, Wikibooks, Wiktionary и Wikidata. Поэтому при создании Большой украинской энциклопедии e-ВУЕ было принято решение также ориентироваться на Wiki-технологию.

1 Семантические Wiki-ресурсы

Чтобы при разработке Wiki-ресурсов перейти от обработки данных к обработке и поиску знаний, используют семантические расширения. Сформированные на их основе ИР могут динамично обновляться всем сообществом пользователей, которые обеспечивают актуальность информации, имеют удобную и простую для понимания структуру, обеспечивают обработку информации на семантическому уровне, предоставляя технологическую платформу для группового управления знаниями. Одним из наиболее известных инструментов является *Semantic MediaWiki* (SMW) [8]. Эта надстройка над MediaWiki имеет высокую выразительную мощность, надёжную реализацию и удобный интерфейс пользователей. С её использованием реализован ряд успешных проектов. Следует отметить, что это были относительно небольшие тематические энциклопедии (например, энциклопедия Первой мировой войны) и порталы научных учреждений с однородным контентом. Однако знания, представленные в

Большой украинской энциклопедии, имеют значительно более сложную структуру, и поэтому возникает необходимость расширить базовые возможности SMW современными технологиями Semantic Web и средствами управления знаниями. Для этого вначале необходимо определить что именно можно представить встроенными средствами SMW.

SMW – это надстройка над инструментальным средством построения Wiki-сайта MediaWiki, которая позволяет интегрировать информацию из разных Wiki-страниц, осуществляя поиск на уровне знаний, и генерировать из Wiki-страниц онтологические структуры, которые могут использовать другие ИИС. Для организации знаний [3] в SMW используются такие механизмы, как категории, *семантические свойства* и *семантические запросы*.

Семантические свойства используются для привязывания данных к Wiki-страницам. Каждое свойство имеет тип, название и значение, а также собственную Wiki-страницу в специальном пространстве имен. Эта страница используется для того, чтобы задавать тип свойства, определять его место в иерархии свойств, а также документировать то, как это свойство необходимо использовать.

Семантические запросы позволяют интегрировать сведения из разных Wiki-страниц, осуществляя поиск на уровне знаний. В качестве параметров запросов используются не только категории, но и семантические свойства и их значения. Это значительно расширяет возможности таких запросов и обеспечивает целостность и актуальность информации.

В частности, в обычных, не семантизированных Wiki-ресурсах поиск ограничен только названиями и категориями страниц. Например, для того, чтобы найти людей, родившихся в 1800 году, необходимо создать отдельную категорию «Родившиеся в 1800 году» и присвоить её соответствующим страницам (в частности, такой подход реализован в Википедии). Такое решение является достаточно громоздким, и, что более важно, не позволяет использовать условия, например, найти людей, родившихся в интервале между 1750 и 1800 годами.

Шаблоны – это специальные Wiki-страницы, содержимое которых предназначено для встраивания в другие страницы. Использование шаблонов позволяет упростить и ускорить создание новых Wiki-страниц, а также обеспечить однотипное представление информации для пользователей. Значительный интерес представляет следующее: если в текст шаблона поместить семантическое свойство, то это свойство будут иметь все страницы, использующие шаблон.

Для любой ПрО, в том числе для энциклопедии, можно выделить *типичные* ИО, – объекты с подобной структурой и одинаковым набором семантических свойств. Создавая Wiki-ресурс, целесообразно разработать специальные шаблоны для таких ИО. Эти шаблоны упрощают построение Wiki-страниц и унифицируют визуализацию контента. Обычно такие шаблоны соответствуют одной или нескольким категориям или входят в состав страниц этих категорий. Шаблоны подкатегорий могут конкретизировать шаблоны для категорий более высокого уровня путём добавления семантических свойств, характерных только для этих подкатегорий.

Кроме того, такие шаблоны позволяют более эффективно и правильно вводить на страницах значения семантических свойств. Однако в использовании шаблонов ИО возникают определённые проблемы, связанные с их универсальностью: в различных экземплярах ИО могут быть определены не все значения семантических свойств, присущие этому типу ИО, а некоторые свойства могут иметь несколько различных значений. Разрабатывая шаблоны ИО, надо заранее предусмотреть такие ситуации. Это усложняет код шаблона, но обеспечивает корректное представление и обработку неполных данных.

В шаблонах SMW ситуации относительно неполноты и многозначности данных обрабатываются в отдельности, и потому необходимо заранее проанализировать семантику таких данных. Кроме того, если шаблон предусматривает визуализацию информации, связанной с

неполными и многозначными семантическими свойствами, то надо предусмотреть, чтобы в случае отсутствия такой информации не выводилось не только само значение, но и сопутствующая информация.

Использование в шаблонах семантических свойств с неполными и множественными значениями позволяет значительно уменьшить количество самих шаблонов, которые используются для описания типичных ИО. Например, в е-ВУЕ используется единый шаблон «Организация» для разных типов организаций, который содержит параметры, релевантные только для отдельных подтипов организаций. В этом шаблоне значение параметра «Вид медпомощи» может вводиться только для медицинских или ветеринарных учреждений, а значение параметра «Целевая аудитория» – только для издательств, СМИ и т.д. При меньшем количестве шаблонов при создании Wiki-страницы значительно проще выбрать соответствующий шаблон для создания статьи, а относительно небольшое количество самих шаблонов типичных ИО позволяет уделять больше внимания проверке и тестированию каждого из них.

Семантические значения, которые вводятся в шаблонах, могут использоваться в семантических запросах, которые позволяют находить Wiki-страницы по определённым требованиям, предъявляемым к этим значениям. Именно благодаря использованию шаблонов для типичных ИО можно достичь унификации в именах этих свойств, которая является значительной проблемой в разработке Wiki-ресурсов большого объёма со сложной и неоднородной структурой.

Чтобы преобразовать семантический Wiki-ресурс со сложной структурой и разнообразным гетерогенным контентом в распределённую базу знаний, к которой могут обращаться внешние ИИС, необходимо разработать средства интероперабельного представления его семантики. Предлагается использовать для этого онтологическое описание структуры Wiki-ресурса. Для этого необходимо сформировать онтологическую модель самого Wiki-ресурса, а также разработать методы её пополнения и использования для извлечения из Wiki-ресурса тех знаний, которые пертинентны той или иной задаче.

2 Онтологии и Semantic MediaWiki

Для разработки и поддержания сложной системы понятий семантического Wiki-ресурса, а также их свойств и отношений целесообразно применять онтологии и связанные с ними средства управления знаниями [9, 10]. Ряд таких возможностей предусмотрен непосредственно в SMW.

С точки зрения онтологического анализа, каждая Wiki-страница представляет собой онтологический элемент одного из RDF-классов – Thing, Class, ObjectProperty, DatatypeProperty, AnnotationProperty. Кроме того, каждая статья имеет собственный URI, что позволяет избежать путаницы между понятиями и HTML- страницами. Обычно статьи являются экземплярами классов онтологии OWL, категории – классами, а отношения – объектными свойствами онтологии.

Исходя из этого, с помощью специальной страницы ExportRDF для любой Wiki-страницы или набора страниц по запросу может генерироваться соответствующий OWL/RDF-файл [1, 11]. К сожалению, эта функция реализована в SMW неудачно и поддерживает ограниченный набор опций. Поскольку SMW совместима с моделью OWL DL [12], то существует возможность использования в Wiki внешних онтологий. Это возможно осуществить двумя путями: импорт онтологии позволяет создавать и модифицировать страницы в Wiki для представления отношений, заданных в некотором OWL DL-документе; а повторное использование словаря позволяет пользователям отображать Wiki-страницы на элементы существующих онтологий.

Функция импорта онтологии для чтения RDF-документов использует инструментарий RAP toolkit. Он извлекает RDF-утверждения, которые могут быть представлены в Wiki. Наименования статей импортированных элементов извлекаются с их меток (labels), или, в случае отсутствия метки, из идентификатора раздела их URI. Основной целью импорта является инициализация (первичная автоматическая загрузка) основы-шаблона для заполнения Wiki. Кроме того, импорт онтологии добавляет специальные аннотации, которые генерируют эквивалентные утверждения в экспорт OWL (owl:sameAs, owl:equivalentClass или owl:equivalentProperty). Импорт онтологий разрешён только администраторам сайта.

Импорт словаря позволяет пользователям идентифицировать элементы Wiki, указывая связь с элементами существующих онтологий. Например, Category:Person может непосредственно экспортироваться в класс foaf:Person словаря Friend-Of-A-Friend. Wiki-пользователи могут решать, какие страницы Wiki должны иметь внешнюю семантику, тем не менее набор имеющихся внешних элементов управляется только администраторами. Вводя в словарь Wiki некоторый новый элемент, они должны удостовериться в том, что повторное использование словаря соотносится с типами ограничений OWL DL. Например, внешние классы, такие, как foaf:Person, не могут быть импортированы в отношения.

Экспорт в OWL/RDF является средством обеспечения внешнего повторного использования данных из Wiki, но только практическое применение этой функции может показать качество сгенерированного RDF. Кроме того, SMW предоставляет сервис для поддержки запросов SPARQL. Система базируется на автономном RDF-сервере Joseki, синхронизированном с семантическим контентом Wiki.

3 Построение онтологии задачи на основе Wiki-ресурса

В ряде случаев для решения задачи пользователю нужна онтология, которая содержит знания о ПрО. Если пользователя не удовлетворяют встроенные в SMW средства построения онтологий, то он может использовать более сложный способ, при котором часть работы не может быть автоматизирована и требует его участия. Такая ситуация может возникнуть в тех случаях, если пользователю сложно построить формализованное описание ПрО, но он достаточно чётко представляет, какие именно сведения важны для его задачи. Такая онтология, в частности, может быть применена для персонифицированного поиска информации в Web, в рекомендующих системах, в задачах машинного обучения.

Для описания онтологий будем использовать формальную модель $O = \langle X, R, F, T \rangle$, более подробно описанную в [13], которая состоит из следующих элементов:

- $X = X_{cl} \cup X_{ind}$ – множество концептов онтологии, где X_{cl} – множество классов, X_{ind} – множество экземпляров классов;
- $R = r_{ier\_cl} \cup \{r_i\} \cup \{p_j\}$ – множество отношений между элементами онтологии, где r_{ier_cl} – иерархические отношения между классами онтологии и свойствами классов; $\{r_i\}$ – множество объектных свойств, которые устанавливают отношения между экземплярами классов; $\{p_j\}$ – множество свойств данных, которые устанавливают отношения между экземплярами классов и значениями;
- 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).

Формально построение онтологии задачи пользователя состоит в следующем: по онтологии ПрО $O_{domain} = \langle X_{domain}, R_{domain}, F_{domain}, T_{domain} \rangle$ и набору Wiki-страниц W_{user}, семантиче-

ская разметка которых базируется на O_{domain}, строится «лёгкая» онтология задачи пользователя O_{user}, знания которой являются подмножеством знаний из O_{domain}. Онтология ПрО может иметь произвольную структуру, высокую выразительную мощность и быть сформирована как экспертами ПрО, так и с помощью средств получения онтологических знаний.

$O_{user} = \langle X_{user}, R_{user}, F_{user}, T_{user} \rangle$, такая, что:

- $X_{user} \subseteq X_{domain}$, то есть $X_{cl_{user}} \subseteq X_{cl_{domain}}$, $X_{ind_{user}} \subseteq X_{ind_{domain}}$;
- $R_{user} \subseteq R_{domain}$, то является $r_{ier\_cl_{user}} = r_{ier\_cl_{domain}}$, $\{r_{user_j}\} \subseteq \{r_{domain_i}\}, i = \overline{0,n}, j = \overline{0,m}, m \le n$;
- $F_{user} = \varnothing$;
- $T_{user} \subseteq T_{domain}$.

Такую работу целесообразно выполнять в том случае, если пользователь начинает работать над сложной проблемой, решение которой будет требовать информации на протяжении довольно значительного времени (значительно большего, чем время на построение собственной онтологии). Например, планируя исследования на несколько лет, целесообразно израсходовать несколько часов на то, чтобы в дальнейшем получать семантически отфильтрованные сведения.

3.1 Основные этапы построения онтологии задачи

Этап 1. Найти семантический Wiki-ресурс W, который по тематике соотносится с задачей пользователя или перекрывает более широкую ПрО. Проще всего использовать неспециализированные энциклопедии и справочники (такие, как e-BYE), но, если пользователь располагает сведениями о более специализированных ресурсах, то их применение может увеличить эффективность работы.

Этап 2. Отобрать в этом Wiki-ресурсе множество страниц W_{user}, которые непосредственно связаны с задачей пользователя, $W_{user} \subseteq W$. Начать этот отбор можно с поиска Wiki-страниц, названия которых совпадают с ключевыми словами из описания задачи, а в дальнейшем воспользоваться одним или несколькими из следующих способов:

- с помощью встроенных средств навигации по Wiki-ресурсу переходить к страницам, соединённым с этими страницами семантическими свойствами (всеми или только теми, которые интересуют пользователя);
- воспользоваться семантическим поиском по выбранным свойствам;
- найти страницы тех же категорий.

На этом этапе пользователь может выполнить определённое количество работы самостоятельно, чтобы охарактеризовать ту информацию, которая ему нужна, и отвергнуть ту, которая не касается его текущей задачи (это может быть ценная информация, важная для ПрО в целом, но не нужная именно для решения текущей проблемы). От того, насколько точно будет выполнен отбор, зависит эффективность использования построенной онтологии в задаче пользователя: отсутствие нужной информации не разрешит находить соответствующие ресурсы, а наличие лишних страниц увеличит время обработки.

Этап 3. Проанализировать информацию из $W_{user} = \{w_{user_i}\}, i = \overline{1,s}$ для каждой страницы:

- информация о классах страницы (все или отобранные пользователем) позволяет пополнить $X_{cl_{user}}$, иерархические отношения между этими классами можно определить с помощью страниц этих категорий;
- имя самой страницы заносится в $X_{ind_{user}}$;

- имена тех семантических свойств страниц, которые использованы на данной странице и область определения которых относится к типу «Страница» (все или отобранные пользователем), заносятся в $\{r_{user_j}\}$;
- имена страниц, на которые данная страница ссылается с помощью семантических свойств типа «Страница» (все или отобранные пользователем), также заносятся в $X_{ind_{user}}$;
- имена страниц, на которые данная страница ссылается с помощью гиперссылок (все или отобранные пользователем), также заносятся к $X_{ind_{user}}$;
- если данная страница отсылает на другую страницу, то имя такой страницы рассматривается как синоним текущей страницы, заносится в $X_{ind_{user}}$ и связывается отношением синонимии с именем текущей страницы.

  При обработке несемантизированных Wiki-страниц алгоритм значительно сокращается.

3.2 Семантический поиск

Использование семантического поиска [14, 15] позволяет сократить время обработки, т.к. в этом случае надо обрабатывать не каждую страницу отдельно, а только результат запроса, который содержит важные для пользователя сведения в упорядоченном виде. В таком поиске в качестве условий могут быть заданы категории и условия, налагаемые на значения семантических свойств страниц.

{{#ask:
[[Категорія:персоналії]]
[[Рік народження::>1900]]
[[Рік народження::<1950]]
[[Місце народження::Україна]]
|?Рік народження
|?Місце народження
|?Alma mater
|?Напрями діяльності
|format=broadtable
|link=all
|headers=show
|searchlabel=... подальші результати
|class=sortable wikitable smwtable }}

Пользователь может вводить эти условия явным образом и определять, в каком формате получить результат – список, таблица (рисунок 1), облако тэгов (рисунок 2) и т.п.

Использование семантического поиска очень удобно как при генерации, так и при обновлении онтологий ПрО. Например, если надо составить онтологию рек и населённых пунктов определённого *Региона А*, то без семантического поиска надо обработать все страницы категорий *Города, Реки* и *Регион А*, а при использовании поиска – страницу, сгенерированную запросом с условиями, выделяющими на страницах категории *Реки* со значением свойства *Регион* информацию о *Городах* на берегу таких рек:

{{#ask: [[Категорія:Річки]]
[[Регіон::Україна]]
|?Міста на березі }}

Сам код запроса генерируется средствами SMW автоматически (следует отметить, что SMW позволяет вручную создавать значительно более сложные запросы, но это требует от

пользователя соответствующих навыков). Такие запросы можно встроить потом в Wiki-страницу, скопировав их код, и информация будет обновляться автоматически при изменении контента удовлетворяющих запросу страниц.

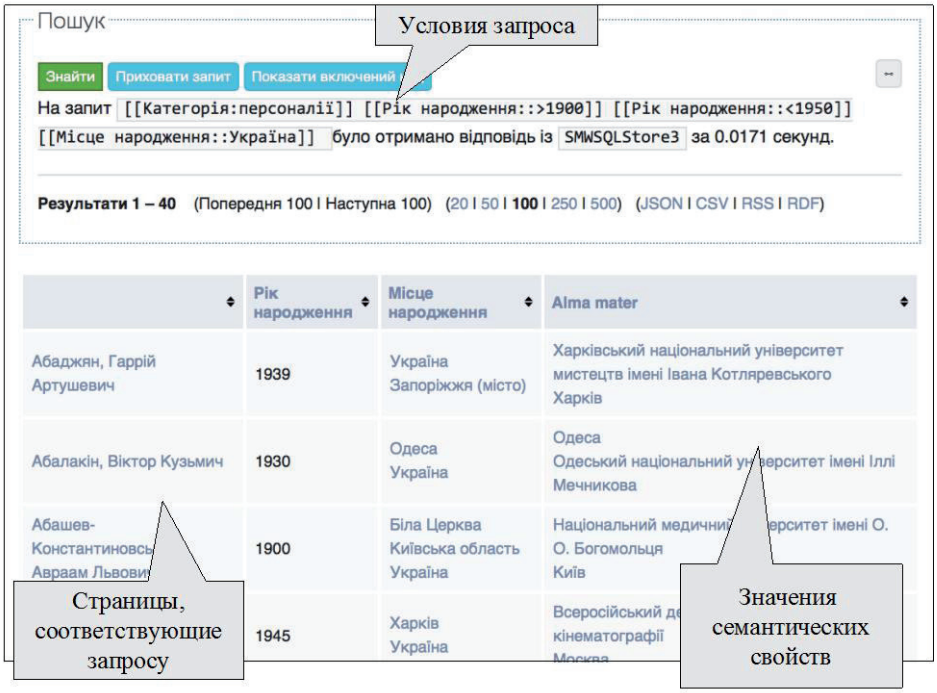

Рисунок 1 - Результаты семантического поиска в табличном виде

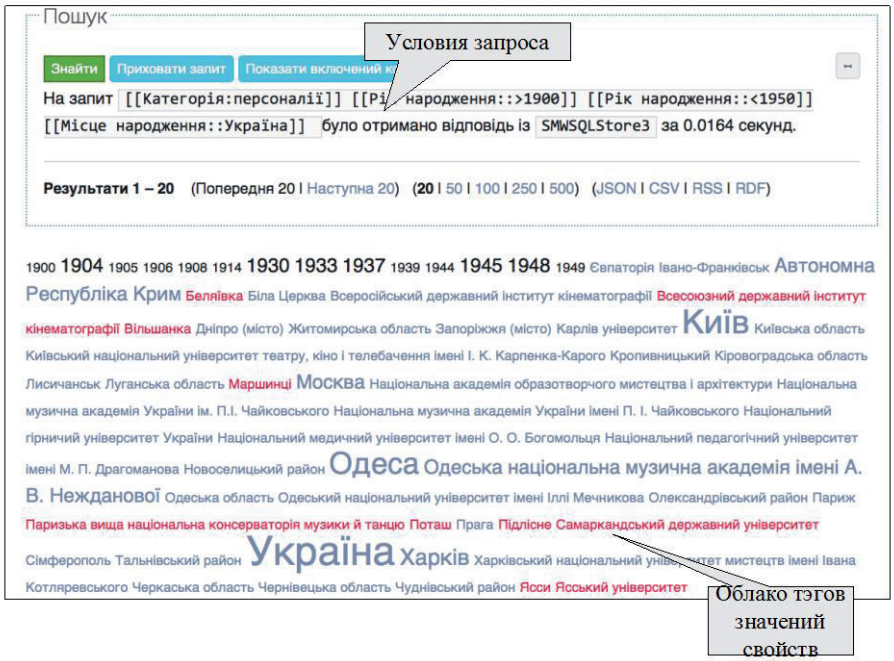

Рисунок 2 - Результаты семантического поиска в виде облака тэгов

Рассмотренный выше алгоритм позволяет построить онтологию задачи пользователя. Хотя существует возможность различной программной реализации, из-за того, что большинство операций нуждается во вмешательстве пользователя и принятии решения относительно

каждого понятия и отношения, на практике проще выполнять эти действия непосредственно с помощью редактора онтологий (например, Protégé [16]). Таким образом формируется онтология на языке OWL.

Основная проблема для пользователя при создании таких запросов заключается в необходимости правильно указать имена категорий и семантических свойств. Эту информацию можно получить, проанализировав код шаблонов типичных ИО, используемых на соответствующих Wiki-страницах. Но такой путь довольно сложен. Поэтому целесообразно при разработке семантических Wiki-ресурсов создавать специальную справочную страницу, описывающую вводимые в шаблонах типичных ИО семантические свойства и примеры их значений (рисунок 3).

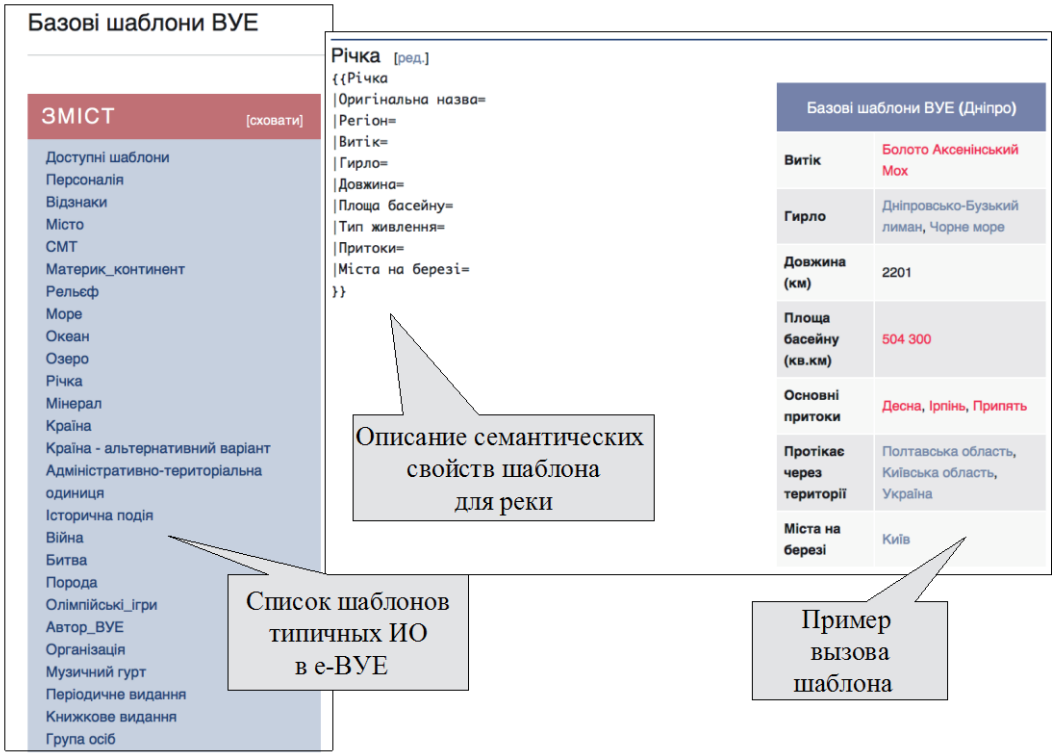

Рисунок 3 - Использование шаблонов типичных информационных объектов для построения онтологии ПрО.

3.3 Портальная версия Большой украинской энциклопедии (е-ВУЕ) - семантический Wiki-ресурс

Сейчас в Украине активно ведутся работы по созданию Большой украинской энциклопедии. Качество информации, представленной в энциклопедии, обеспечивается ориентацией на рецензированные авторские статьи с оригинальным контентом, подготовленные специалистами в соответствующих областях, в которых представлены проверенные факты и признанные научным сообществам теории. е-ВУЕ (vue.gov.ua) – портальная версия Большой украинской энциклопедии. Кроме текста, страницы е-ВУЕ могут содержать другие типы контента (изображение, карты, видео, аудио и т.п.) и ссылки на доверенные источники. Для создания этого инновационного ИР, на базе современных знания-ориентированных технологий и оригинальных разработок проводится исследование соответствующих моделей и методов представления и обработки информации. При необходимости разрабатываются оригинальные программные решения, которые базируются на современных методах представления распределённых знаний (в частности, на технологиях Semantic Web).

Можно выделить следующие преимущества е-ВУЕ по сравнению с другими электронными справочниками и энциклопедиями:

- явное установление содержательных связей между страницами статей и их элементами;
- поиск информации по смыслу – по категориям и значениями семантических свойств страниц;
- возможность интегрировать информацию из разных статей и автоматизированно генерировать целостный контент;
- возможность экспорта знаний в форматах современных Web-технологий.

е-ВУЕ использует современные технологии и научные достижения в области управления знаниями, искусственного интеллекта, онтологического анализа, интеллектуального поиска и Semantic Web. Предполагается, что энциклопедия станет источником знаний не только для людей, но и для ИИС, которые смогут использовать сведения, экспортированные из е-ВУЕ в общепринятых форматах представления. Портальная версия Большой украинской энциклопедии использует свободное программное обеспечение MediaWiki версии 1.29.1. и его семантическое расширение SMW версии 2.5.5.

Каждая статья е-ВУЕ может быть отнесена к произвольному количеству категорий. Средства Wiki-среды позволяют явным образом указывать иерархические связи между такими категориями, которые могут отображать разные аспекты классификации статьи энциклопедии, учитывать специфику ПрО, условия публикации, использование материала и т.д. [17]. С точки зрения пользователей на верхнем уровне статьи подразделяются на три не пересекающиеся категории: «Персоналии», «Цивилизация» и «Природа». К каждой из этих категорий можно перейти непосредственно с главной страницы е-ВУЕ (рисунок 4).

Рисунок 4 - Категории верхнего уровня на главной странице е-ВУЕ

Кроме того, на этой же странице предусмотрен переход к иерархическому набору категорий, упорядоченному по областям знаний. В каждой из этих категорий есть соответствующие подкатегории. Например, для категории «Персоналии» – это «Учёные», «Лауреаты

Нобелевской премии» и т.д., для категории «Природа» – это «Географические объекты», «Гидронимы» и т.д. (рисунок 5).

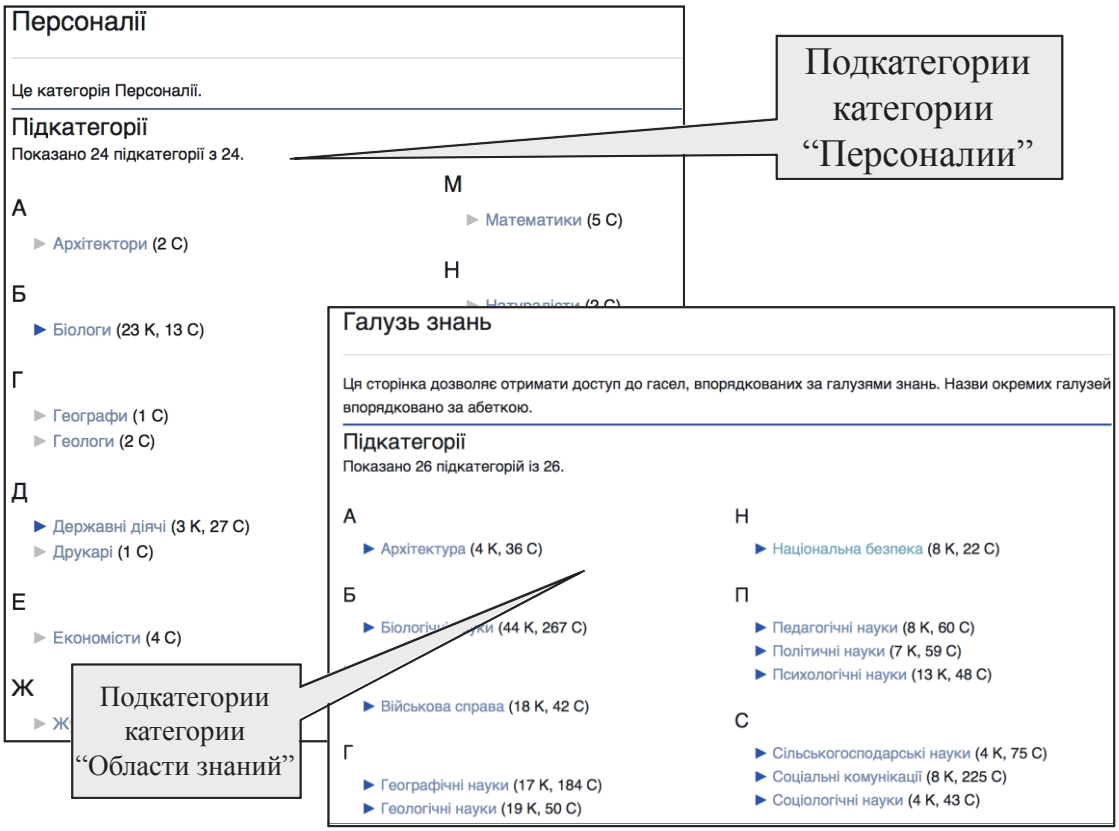

Рисунок 5 - Подкатегории областей знаний в е-ВУЕ

4 Онтологическая модель е-ВУЕ

Для расширения функциональных возможностей портальной версии Большой украинской энциклопедии на основе её семантизации необходимо разработать методы построения полной, формализованной и однозначно интерпретируемой системы категорий и семантических свойств Wiki-страниц. Чтобы е-ВУЕ была способная функционировать как распределённая база знаний, являющаяся источником полезной и проверенной информации как для людей, так и для внешних ИИС, необходимо создать и практически применить онтологическую модель [18] знаний энциклопедического издания. Онтологическая модель е-ВУЕ формализует отношения между её основными объектами, их типами и свойствами. Эта модель должна удовлетворять требованиям со стороны средств анализа знаний и соответствовать специфическим ограничениям ПрО, корректно отображая её базовые закономерности. Использование этой модели как основы семантической разметки обеспечивает формирование и программную реализацию соответствующего набора иерархически связанных категорий, шаблонов типичных ИО, их семантических свойств и запросов, которые их используют. Наличие формальной модели позволяет предотвратить неоднозначную интерпретацию знаний разными разработчиками и пользователями портала.

Чтобы система категорий и семантических свойств была полной, непротиворечивой и пертинентной ПрО, целесообразно использовать существующие технологии представления и анализа знаний, ориентированные на Web-применение. Широко применяемый сегодня онто-

логический подход обеспечивает возможность визуализации знаний и их анализа специализированными инструментами. Кроме того, наличие онтологии ресурса для метаописания е-ВУЕ значительно упрощает доступ к контенту внешних ИИС.

С помощью онтологий можно явным образом определить семантику типичных ИО Wiki-ресурса – их семантические свойства и отношение с другими ИО. Важно, что такое онтологическое представление позволяет проявлять и разрешать неоднозначные интерпретации и некорректное использование терминов, связанных с описанием ИО. Кроме того, онтология позволяет решить проблему унификации названий семантических свойств и категорий, которые используют разработчики. Онтология е-ВУЕ (рисунок 6) определяет:

- иерархические отношения между категориями е-ВУЕ;
- объектные свойства, связанные с семантическими свойствами страниц, которые отображают содержательные отношения между разными страницами энциклопедии (например, семантическое свойство «Место рождения» страниц категории «Персоналии» связывает их со страницами категорий «Страна» и «Город»);
- связи между категориями и шаблонами типичных ИО (одной категории может соответствовать несколько шаблонов типичных ИО (например, для категории «Персоналии» это шаблоны «Персоналия» и «Награды»), а один и тот же шаблон может использоваться на Wiki-страницах нескольких категорий (например, шаблон «Награды» используется на страницах категорий «Персоналии» и «Организации»;
- характеристики самих семантических свойств (например, симметричность, транзитивность и т.д.).

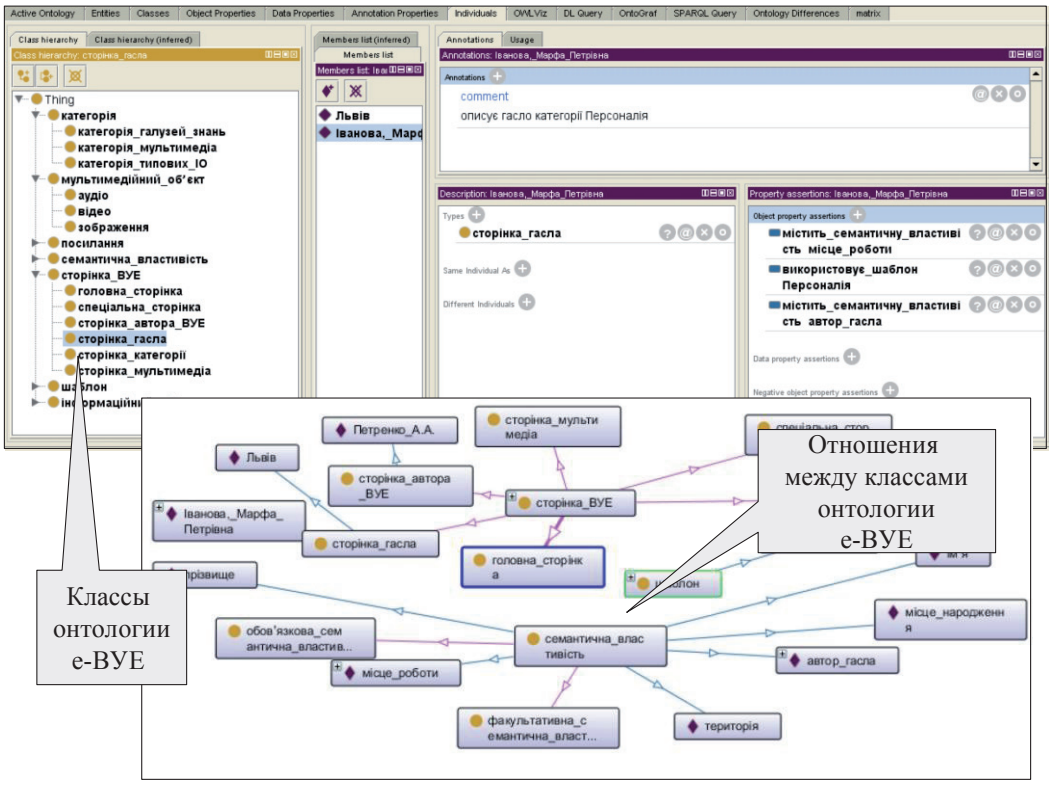

Рисунок 6 - Онтологическая модель е-ВУЕ

Такая онтология в значительной мере зависит не только от специфики ПрО самого Wiki-ресурса, но и от особенностей его реализации. Поэтому для каждого Wiki-ресурса нужно сотрудничество экспертов ПрО с инженерами знаний, чтобы создавать оригинальную онтоло-

гию [19]. Наличие таких онтологий значительно облегчает установление семантических соответствий между различными энциклопедическими ресурсами.

Онтология e-BYE является источником знаний о структуре Wiki-ресурса и помогает пользователям строить семантические запросы, обеспечивая сведения не только для правильного написания названий категорий и семантических свойств типичных ИО, но и для однозначного понимания связей между этими ИО.

Выводы

Предложенные в работе модели и методы были апробированы при разработке портальной версии Большой украинской энциклопедии. Онтологический анализ обеспечил формализацию знаний о структуре знаний, содержащихся в энциклопедии. Семантизация e-BYE позволяет легко создавать и экспортировать базы знаний, которые описывают определённые ПрО в общепринятых форматах (RDF). Интероперабельное представление знаний в онтологической модели обеспечивает корректное выполнение семантических запросов для формирования набора сведений, на основе которых может быть сгенерирована онтология ПрО, интересующая пользователя.

Благодарности

Работа выполнена при проведении исследований по теме «Развитие информационного, функционального и программного обеспечения электронного варианта энциклопедических изданий» в рамках Программы информатизации Национальной академии наук Украины на 2018 год в Институте программных систем НАН Украины в сотрудничестве с Государственным научным учреждением «Энциклопедическое издательство».

Список источников

[1] ***Lassila, O.*** Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation / O. Lassila, R. Swick. — http://www.w3.org/TR/REC-rdf-syntax.
[2] OWL 2 Web Ontology Language Document Overview. W3C. 2009. — http://www.w3.org/TR/owl2-overview/.
[3] ***Soumen, C.*** Mining the Web: Discovering knowledge from hypertext data. Morgan Kaufmann, 2003. — 345 p.
[4] ***Corby, O.*** Querying the Semantic Web with Corese search engine / O. Corby, R. Dieng-Kuntz, C. Faron-Zucker // Proc. ECAI-2004, IOS Press, 2004. – P. 705-709.
[5] ***Thomas, E.*** ONTOSEARCH2: Searching ontologies semantically / E. Thomas, J.Z. Pan, D.H. Sleeman // Proc. OWLED-2007, CEUR Workshop Proceedings 258. CEUR-WS.org, 2007.
[6] ***Leuf B.*** The Wiki way: collaboration and sharing on the Internet, 2001 / B. Leuf, W. Cunningham. — http://www.citeulike.org/group/13847/article/7659081.
[7] ***Krotzsch, M.*** Semantic MediaWiki / M. Krotzsch, D. Vrandecic, M.Volkel. – http://c.unik.no/images/6/6d/SemanticMW.pdf.
[8] Semantic MediaWiki. – https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki.
[9] ***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.
[10] ***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.
[11] OWL 2 Web Ontology Language Document Overview. W3C. 2009. — http://www.w3.org/TR/owl2-overview/.
[12] ***Calvanese, D.*** Tractable reasoning and efficient query answering in description logics: The DL-Lite family / Calvanese D., De Giacomo G., Lembo D., Lenzerini M., Rosati R. // JAR, 39(3), 2007. — P. 385-429.
[13] ***Rogushina, J.*** Analysis of Automated Matching of the Semantic Wiki Resources with Elements of Domain Ontologies / J. Rogushina // 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.

[14] ***Hendler, J.*** Web 3.0: The dawn of semantic search / J. Hendler // Computer, 2010, 43(l), P.77-80.
[15] ***Baeza-Yates, R.*** Next generation Web search / R. Baeza-Yates, A. Raghavan R. // S. Ceri and M. Brambilla, editors, Search Computing, Springer, 2010, P.11-23.
[16] Protégé. – http:// protege. stanford.edu/.
[17] ***Rogushina, J.V.*** The Use of Ontological Knowledge for Semantic Search of Complex Information Objects / J.V. Rogushina // Open semantic technologies for intelligent systems, OSTIS-2017, Minsk, 2017, P.127-132.
[18] ***Rogushina, J.*** Semantic Wiki resources and their use for the construction of personalized ontologies / J. Rogushina // CEUR Workshop Proceedings 1631, 2016, P.188-195.
[19] ***Missikoff, M.*** The usable ontology: An environment for building and assessing a domain ontology / M. Missikoff, R. Navigli, P. Velardi // International semantic web conference, 2002, P. 39-53.

USE OF ONTOLOGICAL ANALYSIS FOR CREATION OF ENCYCLOPEDIC PORTALS

J.V. Rogushina

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

Abstract

We analyze the expediency of the development of the intelligent, semantically structured information Web resources that are oriented both on their use by humans and their suitability for automated analysis. The advantages and disadvantages of the Wiki technology and the perspectives of its semantic expansion with the use of ontological analysis are considered. Formal ontological model of the Wiki-resource that underlies the development of the user task ontology oriented for utilization by external intelligent applications (for example, for semantic search) is proposed. The main stages of construction of this ontology are described. The implementation of the proposed ontology-based approach is demonstrated on example of the development of on-line encyclopaedias combining free access to materials through the Web with a high level confidence in content developed by experts, is considered. We analyze the basic requirements for software that provides such projects. The necessity of semantization of such resources is substantiated and the main directions of development of the functional of semantic electronic encyclopedias are formulated. The article presents an analysis of the expressive power of the Wiki means for knowledge representation and processing. The need to expand them with artificial intelligence methods and Semantic Web technologies is stated. An ontological model of the encyclopedia serving for development of the portal version of the Great Ukrainian Encyclopedia is built. The main categories and semantic properties of typical information objects that are used in this Encyclopedia that can be applied for retrieval and integration of information are described.

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

Citation: *Rogushina JV.* Use of ontological analysis for creation of encyclopedic portals [In Russian]. *Ontology of designing*. 2019. 9(1): 70-84. – DOI: 10.18287/2223-9537-2019-9-1-70-84.

Acknowledgment

The work was carried out as a part of a research “Development of information, functional and software of the electronic version of encyclopedic issues” as a part of the Informatization Program of the National Academy of Sciences of Ukraine for 2018 at the Institute of Software Systems of the NAS of Ukraine in collaboration with the State Scientific Institution “Encyclopedic Publishing House”.

References

[1] ***Lassila O, Swick R.*** Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation. — http://www.w3.org/TR/REC-rdf-syntax.
[2] OWL 2 Web Ontology Language Document Overview. W3C. 2009. — http://www.w3.org/TR/owl2-overview/.

[3] Soumen C. Mining the Web: Discovering knowledge from hypertext data. Morgan Kaufmann, 2003. — 345 p.
[4] ***Corby O, Dieng-Kuntz R, Faron-Zucker C***. Querying the Semantic Web with Corese search engine // Proc. ECAI-2004, IOS Press, 2004. – P. 705-709.
[5] ***Thomas E, Pan JZ, Sleeman DH***. ONTOSEARCH2: Searching ontologies semantically // Proc. OWLED-2007, CEUR Workshop Proceedings 258. CEUR-WS.org, 2007.
[6] ***Leuf B, Cunningham W.*** The Wiki way: collaboration and sharing on the Internet, 2001. — http://www.citeulike.org/group/13847/article/7659081.
[7] ***Krotzsch M, Vrandecic D, Volkel M.*** Semantic MediaWiki, – http://c.unik.no/images/6/6d/SemanticMW.pdf.
[8] Semantic MediaWiki. – https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki
[9] ***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.
[10] ***Obr L, Ceuster W, Mani I, Ra S, Smith B.*** The evaluation of ontologies // In Semantic web: Revolutionizing Knowledge Discovery in the Life Sciences, New York: Springer Verlag, 2006, 139-158. – https://philpapers.org/archive/OBRTEO-6.pdf.
[11] OWL 2 Web Ontology Language Document Overview. W3C. 2009. — http://www.w3.org/TR/owl2-overview/.
[12] ***Calvanese D, De Giacomo G, Lembo D, Lenzerini M, Rosati R.*** Tractable reasoning and efficient query answering in description logics: The DL-Lite family // JAR, 2007; 39(3): 385-429.
[13] ***Rogushina J.*** Analysis of Automated Matching of the Semantic Wiki Resources with Elements of Domain Ontologies // International Journal of Mathematical Sciences and Computing (IJMSC), 2017; 3(3): 50-58. – http://www.mecs-press.org/ijmsc/ijmsc-v3-n3/IJMSC-V3-N3-5.pdf.
[14] ***Hendler J.*** Web 3.0: The dawn of semantic search // Computer, 2010, 43(l): 77-80.
[15] ***Baeza-Yates R, A. Raghavan R***. Next generation Web search // S. Ceri and M. Brambilla, editors, Search Computing, Springer, 2010, P.11-23.
[16] Protégé. – http://protege.stanford.edu/.
[17] ***Rogushina JV.*** The Use of Ontological Knowledge for Semantic Search of Complex Information Objects // Open semantic technologies for intelligent systems, OSTIS-2017, Minsk, 2017, P.127-132.
[18] ***Rogushina J.*** Semantic Wiki resources and their use for the construction of personalized ontologies // CEUR Workshop Proceedings 1631, 2016, P.188-195.
[20] ***Missikoff M, Navigli R, Velardi P.*** The usable ontology: An environment for building and assessing a domain ontology // International semantic web conference, 2002, P. 39-53.

Сведения об авторе

Рогушина 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. - http://orcid.org/0000-0001-7958-2557.

Yulia Vitalyevna Rogushina, born in 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 textbooks in the field of ontological analysis, semantic search, intelligent agents and knowledge management. - http://orcid.org/0000-0001-7958-2557.