

УДК 004.08

**С.П. Риппа**, д-р екон. наук, проф., нац. ун-т гос. налог. службы Украины, г. Ирпень,  
**О.Н. Ляшенко**, д-р екон. наук, Тернопольский нац. экон. ун-т, г. Тернополь

## СЕМАНТИЧЕСКИЕ ПЛАТФОРМЫ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ БАЗ ЗНАНИЙ В ИНФОРМАТИКЕ

*С.П. Риппа, О.М. Ляшенко. Семантичні платформи програмного забезпечення баз знань в інформатиці.* В даній статті розглядається проблематика програмного забезпечення систем баз знань, орієнтованих на сегмент Семантичного Вебу. Аналізується динаміка розвитку програмних інструментів і архітектура технологій програмної підтримки семантики Інтернету. Також дається коротка характеристика платформи програмного забезпечення для оперування базами знань Семантичного Вебу.

*Ключові слова:* база знань, Семантичний Веб, програмне забезпечення, XML, RDFa, OWL, RDFaCE, CMS, Word Press.

*С.П. Риппа, О.Н. Ляшенко. Семантические платформы программного обеспечения баз знаний в информатике.* В данной статье рассматривается проблематика программного обеспечения систем баз знаний, ориентированных на сегмент Семантического Веба. Анализируется динамика развития программных инструментов и архитектура технологий программной поддержки семантики Интернета. Дается также короткая характеристика платформы программного обеспечения для оперирования базами знаний Семантического Веба.

*Ключевые слова:* база знаний, Семантический Веб, программное обеспечение, XML, RDFa, OWL, RDFaCE, CMS, Word Press.

*S.P. Rippa, O.M. Lyashenko. Semantic platforms of knowledge bases software in informatics.* Given paper discusses the problems of knowledge bases software systems in a Semantic Web segment. The dynamics of the development of software tools and technology of Semantic Web software support architecture. There is given also a short description of the software platform for operation of Semantic Web knowledge bases.

*Keywords:* knowledge base, Semantic Web, software, XML, RDFa, OWL, RDFaCE, CMS, Word Press.

Научная актуальность тематики баз знаний является достаточно обширной областью разносторонних исследований уже более 40 лет. В начальные периоды развития компьютерных наук и сетевых технологий достаточно мощный фундамент в развитие теории баз знаний заложили такие разделы информатики как теории искусственного интеллекта, логического программирования, фреймов, семантических сетей, фундаментальной математики и многих других научных направлений. Особую роль сыграло также в этот период появление и распространение Интернета, который сегодня может рассматриваться как частичная архитектура в виде всемирной сетевой интеллектуальной среды для формирования, поддержки и развития глобальной базы знаний человечества. Понятно, что информационные ресурсы Интернета, задействованные в научных исследованиях и образовании та интенсивно интегрируемые в процессах своего развития, являются мощнейшим инструментом цивилизационного познания окружающей нас реальности, в том числе и законодательного регулирования [2]. В данном контексте необходимо назвать современные научные направления информатики, появление которых обеспечило опережающее развитие технической (аппаратной) составляющей компьютерной техники на десятки, а может, и сотни лет в будущем – это фотонные и квантовые компьютеры, нанотехнологии для вычислительной техники, вероятностные чипы, био-компьютерные разработки и многие другие. В вышеназванной формулировке Интернета всемирная интеллектуальная сеть не случайно определяется как «частичная». Следует признать, что наряду с почти фантастическими успехами развития

компьютерной техники, возникли и проявились негативные факторы и угрозы поголовной информатизации, которые значительно ограничили функциональность и целостность интеллектуальных составляющих сети. Сюда можно отнести киберпреступность и кибертерроризм, различные другие информационные угрозы, такие как неуправляемый рост «загрязненного информационным мусором» та «инфицированного компьютерными вирусами» Интернета, начавшиеся недавно и претендующие на расширение «кибер-войны» и «кибер-конфликты» и многие др. негативизмы.

В противоположность всестороннему интенсивному развитию технической информатики, на наш взгляд, наметился определенный кризис ее гуманитарной составляющей, в которой доминирующим понятием является информационная культура и соответствующая ей инфраструктура, которую, в свою очередь, формирует весь гуманитарный комплекс информационных наук и образования. В данной области за последние годы нелегко назвать выдающиеся успехи та имена ученых. Однако два направления в этой сфере, все же, следует отметить — это «Семантический Веб» и «облачные вычисления», которые можно отнести к научным прорывам в теоретической информатике, касающейся компьютерных баз знаний и перспектив их формирования и применения в Интернете. Авторство Семантического Веба принадлежит Тиму Бернерсу-Ли [5], одному из основателей и руководителей Веб консорциума (W3C), а авторами «облачной вычислительной идеологии» можно считать несколько авторских коллективов, распределением ролей в которых еще занимаются историки информатики. Еще одним знаменательным событием глобальной информатики можно считать появление феномена «социальных сетей», оценить влияние которого на формирование мирового информационного общества нам еще предстоит. Однако, среди множества различных и разнородных социальных сетей, таких как Facebook, Twitter, YouTube, ВКонтакте, Одноклассники, пользователями которых являются не просто миллионы посетителей Интернета, а сотни миллионов и миллиарды социально активных граждан всего Интернет-сообщества, необходимо отметить несколько сетевых проектов, типа, так называемых, «Интернет-энциклопедий». Важной характеристикой подобной сетевой энциклопедии является ее открытый формат и возможность заполнения и доопределения понятий, то есть коллективное Интернет-творчество всех участников сетевого общения, естественно, при наличии определенных, согласованных сообществом, этических коммуникационных норм общения и пополнения сетевой энциклопедии.

Модель Семантического Веба можно представить как совместное развитие двух направлений, первое из которых охватывает языки представления данных. На сегодняшний день основными такими языками являются XML (eXtensible Markup Language), языковые средства описания ресурсов RDF (Resource Description Framework), имеющих статус рекомендаций W3C, и язык онтологий OWL, который позволяет определить понятия и отношения между ними.

Второе концептуальное направление несет в себе теоретическое понимание о модели предметных областей. Такие модели в терминологии Семантического Веба называются онтологиями. В феврале 2004 года консорциумом W3C утвердил и опубликовал спецификацию языка сетевых онтологий OWL (Web Ontology Language).

Таким образом, в Семантическом Вебе используются три ключевых языка и соответствующие им технологии программного обеспечения:

- спецификация XML, позволяющая определить синтаксис и структуру документов (так называемые таксономии);
- языковые средства описания ресурсов RDF, обеспечивающие модель кодирования базовых понятий для описания знаний, определенных в будущей онтологии;
- язык онтологий OWL, позволяющий расширенно определить понятия знаний, отношения между ними, методы их обработки.

Язык онтологий OWL предназначен для описания классов и отношений между ними, которые присущи сетевым документам (данным) и программным приложениям.

Кардинальным направлением в процессах создания нового Интернет, а точнее трансформации его инфраструктуры в сторону повышения содержательности и полноты использования информационных ресурсов является формирование соответствующего Интернет-сегмента Семантического Веба, который позволяет поднять на принципиально новый уровень возможности оперирования в Интернете именно базами знаний. Переход от сверхбольших баз данных несовместимой структуры, разнородных форматов и функционально ограниченного программного обеспечения для манипулирования информацией в Сети к знания-ориентированным концепциям современного информационного общества базируется преимущественно на использовании XML-стандартов форматирования и архитектурного построения Интернета. Такой подход можно охарактеризовать двумя принципиальными особенностями:

— во-первых, это многоуровневая архитектура Семантического Веба, которая состоит из различных инструментов и технологий универсально-языковых средств, таких как таксономии данных, RDF-базис распределенных информационных ресурсов Интернета, XSLT-шаблоны описания трансформационных схем баз знаний, а также уровень доверительных компонентов сети для выполнения доказательств и логических построений в пределах баз знаний, который представлен мультиагентными, веб-сервисными компонентами, а также другими проектами онтологических систем, охарактеризованными ниже;

— во-вторых, верхний уровень Семантического Веба, который формирует доверительный его сегмент, базируется на онтологических концепциях и системах, обеспечивает функции оперирования метазнаниями, координацию и управление взаимодействием между составляющими Семантического Веба, а также администрирование на межсистемном уровне информационными ресурсами баз знаний в общем Интернет-пространстве.

Первые годы развития и распространения идеологии Семантического Веба отразились в истории Интернета рядом проектов и программных исследований, большинство из которых продолжают активно развиваться в настоящее время (стандарт XML — официально действует начиная с 1994 г., а с 2000 года по настоящее время – начало реализации, развития и внедрения соответствующих проектов и версий программных продуктов на его основе – DAML+OIL, XBRL, RDF, OWL, SKOS, SPARQL, SHOE, Protege, Jena, DC, FOAF, RSS, OPEN LINK Data Explorer и т.д. [1, 4, 6]). Оценка результативности указанных проектов открывает широкие перспективы понимания Семантического Веба как нового сегмента существующего Интернета через наполнение его обновленной, семантически ориентированной информацией, позволяющей пользователям и компьютерным системам успешно взаимодействовать между собой. Характерным для этого подхода является то, что новый содержательный уровень взаимодействия понимается в широком смысле, т.е. учитывая возможность совместного функционирования приспособленных под Семантический Веб мультиагентных и веб-сервисных систем в автономном режиме, в том числе, и без прямого участия пользователей, когда программные средства начинают «договариваться» и взаимодействовать сами. К наиболее распространенным сценариям будущего использования Семантического Веба можно отнести технологии семантического поиска, объединение знаний (интеграции баз данных и знаний), всепроникающих вычислений и облачных вычислительных технологий (Cloud Computing) и т.п.

Таким образом, архитектуру онтологии в обобщенном виде можно представить на основе трех составляющих — универсально-языковый базис (XML-стандарт), методологии и модели описания глобальных информационных ресурсов (RDF-язык) и, собственно, язык описания онтологий (OWL). Представленная на рис. 1 модель базы знаний в виде онтологий, демонстрирует ее архитектуру, взаимосвязанные структурные элементы, их место и динамику их формирования и развития в Интернете. Графически показано, что Семантический Веб является органической частью традиционного Интернета, в котором через трансформационные процессы происходит его расширение в сегменте онтологий и XML-ориентированных систем.

В свою очередь, онтологическая составляющая отображается обобщенно на основе трехуровневой модели баз знаний в виде XML-, RDF-и OWL-компонент, которые и являются приложениями ИИ в различных предметных областях. На рис. 1 представлены также разнообразные сочетания и объединения составляющих Семантического и традиционного Веба, поскольку процесс трансформации и наполнение онтологий по своей сути является эволюционным и охватывает не только собственно Веб, но и другие сегменты Интернета. В качестве примеров в данном контексте могут рассматриваться проявления в группах новостей RSS-систем или различные применения блого-сферы и социальных сетей, которые также используют в своей основе XML-базис.

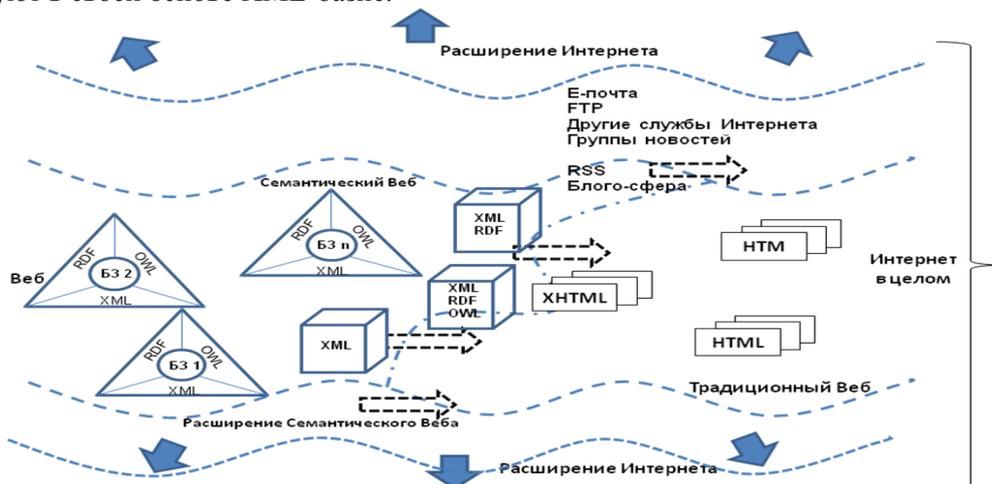


Рис. 1. Обобщенное представление архитектуры и составляющих Семантического Веба

Основная идея Семантического Веба состоит в том, что, будучи надстройкой над существующим Интернетом, он призван сделать размещенную в Сети информацию более удобной для автоматического использования компьютерными средствами и соответствующим программным обеспечением. Подобная семантическая обработка становится возможной в Семантическом Вебе благодаря двум ее важнейшим характеристикам:

— повсеместному использованию унифицированных идентификаторов ресурсов (URI). Традиционная схема использования подобных идентификаторов в пределах современного Интернета сводится к определению ссылок, указывающих на объект адресации. Очевидным свойством такой ссылки является возможность загрузки объекта, на который она указывает.

— повсеместному использованию онтологий и языков описания метаданных (в терминах XML-стандарта). Современные методы автоматической обработки данных, которые доступны в Интернете, как правило, базируются на частотном и лексическом анализе текстов (хотя существуют исключения: например, Swoogle и др.), и, прежде всего, ориентированы на восприятие этой информации исключительно человеком.

Несмотря на все преимущества, которые декларируются относительно Семантического Веба в случае его внедрения, существуют определенные небезосновательные сомнения в возможности его полной реализации. Это означает, что различные комментаторы высказывают разные причины, которые могут быть препятствием к этому, начиная с человеческого фактора (т.е., люди склонны избегать работы по поддержке документов с метаданными, также открытыми остаются проблемы истинности метаданных и т.д.) и, заканчивая косвенной гипотезой Аристотеля об «отсутствии очевидного способа деления мира на концепты», что, в свою очередь, ставит под сомнение возможность существования онтологии верхнего уровня, критической для всей концепции Семантического Интернета.

Следующая группа замечаний касается необходимости избыточного дублирования информации. Необходимость описания метаданных так или иначе приводит к возникновению

дублирования информации. Фактически, каждый документ должен быть создан как минимум в двух экземплярах — размеченным для чтения людьми, а также — в машинно-ориентированном формате. Этот недостаток Семантического Интернета был главным толчком к созданию так называемых микроформатов и языка RDFa. Последний является диалектом языка RDF и отличается от него тем, что не определяет собственного синтаксиса, а предназначается для внедрения (или монтирования) XML-атрибутов в пределах XHTML-страниц.

В то время как совокупность информационных ресурсов и их метаданных можно считать статической частью Семантического Веба, его динамическую составляющую представляют так называемые семантические веб-сервисы, являющиеся законченными элементами программной логики с однозначно описанной семантикой, которые также доступны через Интернет и пригодны для поиска, композиции семантических данных и выполнения различных задач.

Технически, семантический веб-сервис отличается от обычного веб-сервиса наличием не только описания интерфейса (обычно на языке WSDL) в терминах типов данных, передаваемых сервисом, возвращаемых значений и генерируемых ошибок, но и семантическим описанием всех его характеристик. Необходимо отметить, что дублирование данных, которое было упомянуто выше, как недостаток Семантического Веба, в данном случае не происходит: WSDL-описание первично было назначено как раз для машинной обработки.

Потенциальные преимущества от использования семантических Веб-сервисов заключаются в возможностях автоматического поиска (а также композиции) программными агентами соответствующих сервисов для решения поставленных задач. Однако, сложность этой задачи в ее общей формулировке пока еще позволяет получить положительные результаты только в узкоспециализированных областях, которые явным образом выигрывают от внедрения сервисно-ориентированных систем, например, в интеграции корпоративных приложений или отдельных предметных областей. Примерами приложений подобных Веб-сервисов можно считать создание и развитие специфических диалектов XML-языка с соответствующей веб-сервисной средой (например, XBRL — в области международных стандартов финансовой отчетности (МСФО), MathML — в области математики, а также другие подобные примеры в таких сферах как химия, биология, медицина, фармакология, генетика и др.).

Приведенные методические положения по онтологической концепции и технологиям Семантического Веба позволяют начать создание, внедрение, апробацию и поддержку сервисных приложений на основе универсально-языковых средств (XML-RDF-OWL) для отдельных предметных областей, таких как, например, информатизация сфер образования и научных исследований. Дополнительным аргументом, который повышает качество информатизации на основе онтологических средств, является ориентация инструментов Семантического Веба на перспективы формирования и реализации в Интернете баз знаний, позволяющих одновременно решать проблемы будущей интеграции гетерогенных программных приложений. Архитектуру программного обеспечения которых можно уже представить как совокупность разных программных инструментов, сочетающих в себе возможности представления и оперирования семантикой Сети. Один из вариантов архитектуры платформы программного обеспечения, ориентированного на Семантический Веб, изображен на рис. 2.

Важной особенностью платформы программного обеспечения, ориентированного на формирование и поддержку баз знаний, является ориентация на сегмент Семантического Веба и упор на технологии открытых программных систем (open source) и социальных сетей, являющихся одновременно чрезвычайно сильным стимулирующим фактором развития новых семантических сегментов Интернета. Представленная на рис. 2. архитектура реализована при создании совместного веб-портала Национального университета ГНС Украины и Тернопольского национального экономического университета, используемого для подготовки специалистов ИТ-сферы с ориентацией на проектирование и внедрение программных средств Семантического Веба (<http://ndcpo.nusta.com.ua/>).

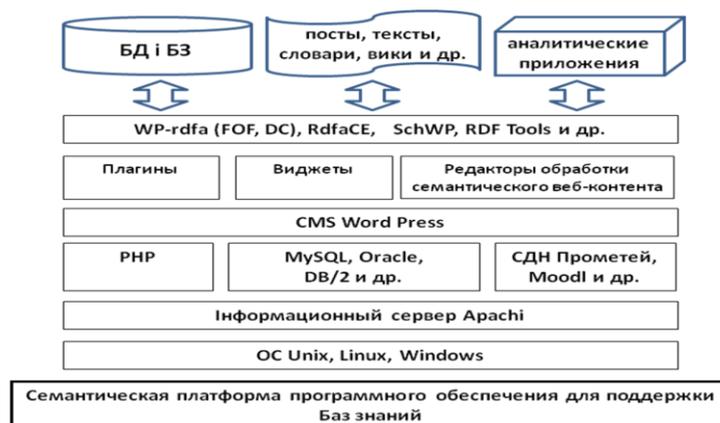


Рис. 2. Представление архитектуры платформы программного обеспечения баз знаний

## Литература

1. Википедия [Электронный ресурс]. — Режим доступа: <http://ru.wikipedia.org/>. — 15.12.12.
2. Закон України «Про основні засади розвитку інформаційного суспільства в Україні на 2007-2015 роки», № 537-V від 09.01.2007 р. // Відомості Верховної Ради України від 23.03.2007. — № 12. — С. 511, ст. 102.
3. Ландэ, Д. Семантический Веб: от идеи – к технологии [Электронный ресурс] / Д. Ландэ // — Режим доступа: <http://poiskbook.kiev.ua/sw.html>. — 20.12.12.
4. Ріппа, С.П. Базы знаний в управлінні ІТ-проектами / С.П. Ріппа // Міжнародна наук.-пр. конф. «Управління проектами: стан і перспективи». — К.: НУК, 2005, — С. 143 — 145.
5. Бернерс-Ли, Т. Семантический Веб [Электронный ресурс] / Т. Бернерс-Ли, Дж. Лендлер, О. Лассила. — Режим доступа: [http://czolin.piscm.net/logic/semantic\\_web\\_rus.html](http://czolin.piscm.net/logic/semantic_web_rus.html). — 23.12.12.
6. Rippa, S. Selection of Alternative Projects Using Data Mining / S. Rippa, T. Lendyuk // Proceedings of IEEE International Workshop, IDAACS' 2007, Dortmund, Germany, September 6 — 9, 2007. — pp. 550 — 554.

## References

1. Vikipedija [Wikipedia] [Electronic resource]. — Access mode: <http://ru.wikipedia.org/>. — 15.12.12.
2. Zakon Ukrainy [Law of Ukraine «About Main Principles of Information Society Development in Ukraine in 2007-2015», No. 537-V dated 09.01.2007] // Bulletin of Verkhovna Rada of Ukraine. — # 12. — P. 511, article 102.
3. Lande, D. Semantical Web: from idea to technology [Electronic resource] / D. Lande // — mode of access: <http://poiskbook.kiev.ua/sw.html>. — 20.12.12.
4. Rippa, S.P. Knowledge Bases in IT-projects management / S.P. Rippa // Proceedings of International Conference «Project Management: state and perspectives». — Kiev, 2005. — pp. 143 — 145. Berners-Lee T. Semantical Web [Electronic resource] / T. Berners-Lee, J. Landler, O. Lassila. — Access mode: [http://czolin.piscm.net/logic/semantic\\_web\\_rus.html](http://czolin.piscm.net/logic/semantic_web_rus.html). — 23.12.12
5. Rippa, S. Selection of Alternative Projects Using Data Mining / S. Rippa, T. Lendyuk // Proceedings of IEEE International Workshop, IDAACS' 2007, Dortmund, Germany, September 6 — 9, 2007. — pp. 550 — 554.

Рецензент д-р техн. наук, проф. Одес. нац. политехн. ун-та Дрозд А.В.

Поступила в редакцию 24 декабря 2012 г.