

## ПРЕИМУЩЕСТВА СЕМАНТИЧЕСКИХ ТЕХНОЛОГИЙ: ПРАКТИЧЕСКИЙ АСПЕКТ

Слесарев Евгений Владимирович  
ФГБОУ ВПО «Мордовский государственный университет им. Н. П. Огарёва»,  
Российская Федерация, г. Саранск  
E-mail: [slesarev@gmail.com](mailto:slesarev@gmail.com) тел. +7 (902) 6686457,  
430005, г.Саранск, ул. Богдана Хмельницкого, д. 39, ком. 503.

**Аннотация.** В статье рассматриваются понятия семантических технологий, существующие проблемы данной отрасли и преимущества практического применения данных технологий на предприятиях.

**Ключевые слова:** семантическая паутина; семантические технологии; RDF; концепция связанных данных; адаптивная онтология; веб-ориентированная архитектура.

### I. ВВЕДЕНИЕ

За последние тридцать лет развитие информационных технологий претерпело ряд технических революций, каждая из которых ставила под вопрос прежние методы и подходы и предлагала новые, приносящие заметные преимущества по сравнению с принятыми подходами. Сегодня начинающееся внедрение спецификаций Semantic Web (Семантического Web), предлагаемых World Wide Web консорциумом, приносит новый кардинальный сдвиг парадигмы в области интеграции приложений и управления данными.

### II. СЕМАНТИЧЕСКАЯ ПАУТИНА

Семантическая паутина (Semantic Web) – это направление развития Всемирной паутины, целью которого является представление информации в виде, пригодном для машинной обработки [1]. В обычной Паутине, основанной на HTML-страницах, информация заложена в тексте страниц и извлекается человеком с помощью браузера. Семантическая же паутина предполагает запись информации в виде семантической сети с помощью онтологий. Таким образом программа-клиент может непосредственно извлекать из паутины факты и делать из них логические заключения. Семантическая паутина – это надстройка над существующей Всемирной паутиной, которая призвана сделать размещённую в ней информацию более понятной для компьютеров. Машинная обработка возможна в семантической паутине благодаря двум её важнейшим характеристикам.

*Повсеместное использование унифицированных идентификаторов ресурсов (URI)*, широко известных как адреса. Традиционно в Интернете эти идентификаторы используются для установки ссылок на адресуемый объект (например, веб-страницу, файл или ящик электронной почты). В семантической паутине URI используются также для именования объектов, то есть каждый URI однозначно называет некоторый объект. Свои URI в семантической паутине есть не только у страниц, но и у объектов реального мира, и даже у абстрактных понятий. Поскольку URI глобально уникальны, они позволяют называть одни и те же предметы в разных местах в семантической паутине.

*Использование семантических сетей и онтологий.* Современные методы автоматической обработки данных, доступных в Интернете, как правило, основаны на частотном и лексическом анализе текстового содержимого, которое прежде всего предназначено для восприятия человеком. В семантической паутине вместо этого используется стандарт RDF, описывающий семантические сети (графы), в которых узлы и дуги имеют URI. Утверждения, кодируемые с помощью RDF, в дальнейшем можно интерпретировать с помощью онтологий, созданных по стандартам RDF Schema и OWL, чтобы получать из них логические заключения. В основе онтологий лежат математические формализмы, называемые дескрипционными логиками.

Без сомнений, мечта Тима Бернерса-Ли, основоположника принципов Семантического Web, достойна воплощения. Он мечтал о том, как хороша и легка была бы

жизнь, если бы все данные в сети были пригодны компьютерам для анализа, они бы понимали, чего хотят от них люди, и все взаимодействия в областях торговли, бюрократии и других повседневных аспектов жизни сводились бы к взаимодействию интеллектуальных агентов, подобно тому, как происходит во многих фантастических фильмах.

Мы же будем рассматривать применимость семантических технологий не в романтическом ключе «новых свежих технологий» и чистого теоретизирования, как бы всё было хорошо, если бы всё это претворилось в жизнь, а именно в практическом аспекте, в частности использования уже сейчас на предприятиях.

### III. ПРИНЦИПЫ ИСПОЛЬЗОВАНИЯ СЕМАНТИЧЕСКИХ ТЕХНОЛОГИЙ НА ПРЕДПРИЯТИИ

Рассмотрим основополагающие принципы использования семантических технологий на предприятии. Эти методы не требуют замены существующих средств и систем, они могут быть протестированы и развернуты постепенно, с низким риском и затратами.

#### Модель данных RDF.

Resource Description Framework (RDF) – это разработанная консорциумом Всемирной паутины модель для представления данных, в особенности – метаданных. RDF представляет утверждения о ресурсах в виде, пригодном для машинной обработки.

Ресурсом в RDF может быть любая сущность – как информационная (например, веб-сайт или изображение), так и неинформационная (например, человек, город или некое абстрактное понятие). Утверждение, высказываемое о ресурсе, имеет вид «субъект – предикат – объект» и называется триплетом. Для обозначения субъектов, предикатов и объектов в RDF используются URI. Множество RDF-утверждений образует ориентированный граф, в котором вершинами являются субъекты и объекты, а рёбра помечены предикатами.

#### RDF (N3):

```
@prefix : <http://www.example.org/> .
:john a :Person .
:john :hasMother :susan .
:john :hasFather :richard .
:richard :hasBrother :luke .
```

#### RDF (XML):

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ns="http://www.example.org/#">
  <ns:Person rdf:about="http://www.example.org/#john">
    <ns:hasMother rdf:resource="http://www.example.org/#susan" />
    <ns:hasFather
      <rdf:Description rdf:about="http://www.example.org/#richard">
        <ns:hasBrother rdf:resource="http://www.example.org/#luke" />
      </rdf:Description>
    </ns:hasFather>
  </ns:Person>
</rdf:RDF>
```

Рис. 1. Пример синтаксиса RDF (N3) и RDF (XML)

RDF является наиболее важной основой для семантической системы предприятия. RDF может применяться одинаково ко всем структурированным (например, стандартные базы данных), частично структурированным (например, HTML-документы) и неструктурированным данным (например, текст). Определяя новые типы и предикаты, можно создавать более выразительные словари в RDF.

Из-за этой универсальности, в настоящее время существует более 150 продуктов для конвертирования различных не-RDF представлений в RDF. В общем представление RDF легко внедрять новые наборы данных или новые атрибуты. Это позволяет с лёгкостью проводить объединение данных из различных приложений, независимо от их формата.

На практике это означает, что на RDF может быть основан некий уровень интеграции, но все исходные данные и схемы все еще могут находиться в родных форматах. Если проще и удобнее передавать или представлять данные в не-RDF форме – нет проблем. RDF необходим только в точке объединения данных, совершенно необязательно всё переводить на него.

### Методы Linked Data.

Linked Data (концепция связанных данных) представляет собой набор лучших практик для публикации и развертывания данных об экземплярах и классах. Две основных заключаются в следующем: называть объекты данных, используя универсальный идентификатор ресурса (URI); предоставлять доступ к данным через протокол HTTP. Обе эти практики позволяют глобальной сети стать распределенной базой данных.

Другие лучшие практики, которые используются достаточно редко, связаны с тем, как характеризовать и классифицировать данные, в особенности в части использования предикатов с правильной семантикой для установления степени родства для связанных элементов данных из разрозненных источников.

### Адаптивные онтологии.

Онтология – это попытка всеобъемлющей и детальной формализации некоторой области знаний с помощью концептуальной схемы. Обычно такая схема состоит из структуры данных, содержащей все релевантные классы объектов, их связи и правила (ограничения), принятые в этой области.

В стандартном виде, семантические онтологии могут варьироваться от небольших и простых до больших и сложных, и могут выполнять следующие основные роли: определение отношений между понятиями, интеграция экземпляров данных, ориентирование на другие области знаний, сопоставление с другими схемами данных.

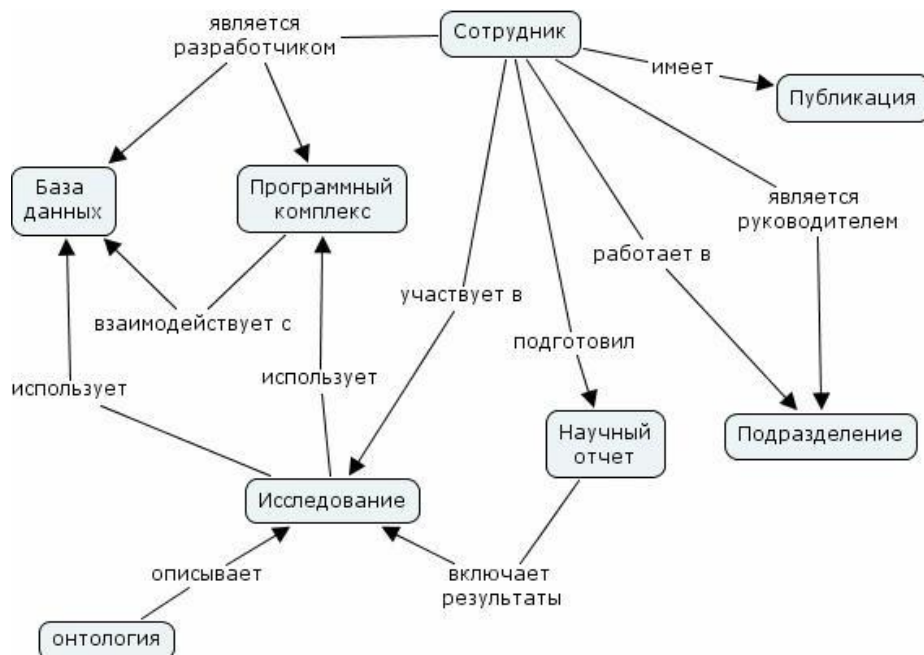


Рис. 2. Пример онтологии в графическом виде

Но, помимо стандартных ролей, онтологии могут взять на себя ещё одну роль – наполнение информацией пользовательских интерфейсов, в дополнение к стандартной роли по интеграции информации.

В этой связи такие структуры называются адаптивными онтологиями. Некоторые из дополнительных применений адаптивных онтологий: названия атрибутов и подсказки, навигация и просмотр структур и деревьев, структуры меню, автоматическое завершение при вводе данных, контекстные выпадающие списки выбора, проверка правописания, и т.д. Иными словами, то, что делает онтологию адаптивной, это добавление к стандартной машиноориентированной цели онтологии дополнительных средств для человеческого понимания - названий, синонимов, определений и прочего.

Усилия по управлению знаниями теперь могут перейти к фактическому описанию характера и отношений информационной среды. Другими словами, онтологии становятся основным объектом усилий и развития. Навыки, необходимые для создания таких адаптивных онтологий: логика, последовательное мышление и знания в предметной области. То есть, любой значимый эксперт в предметной области или в области управления знаниями скорее всего, имеет необходимые навыки, чтобы способствовать развитию и улучшению онтологий. Таким образом, мы определяем положительную тенденцию – смещение акцента от ИТ к самим знаниям, к предметной области.

### **Основанные на онтологиях приложения.**

Приложения основанные на онтологиях – это модульные приложения общего плана, предназначенные для работы в соответствии со спецификациями, содержащимися в адаптивной онтологии.

Приложения основанные на онтологиях предназначены для выполнения конкретных общих задач: импорт и экспорт в различные форматы, создание и управление наборами данных, отчетность, просмотр, поиск, визуализации данных, распределение прав доступа пользователей, и тому подобное. Эти приложения определяют свои специфические функции в соответствии с содержащимися в онтологии спецификациями.

Такие приложения более похожи на виджеты или API-ориентированные фреймворки, в отличие от специализированного программного обеспечения прошлого, хотя функциональность, очевидно, схожа. Главное изменение – подстройка под довольно общий уровень абстракции, который отвечает структуре и конвенциям руководящей онтологии. Основным преимуществом является то, что одно общее приложение может реализовать некоторые общие функции, основанные на любой правильно построенной адаптивной онтологии.

Такая конструкция ограничивает так называемую хрупкость ПО и увеличивает возможности его повторного применения. Кроме того, как отмечалось выше, это сдвигает фокус усилий от разработки и сопровождения программного обеспечения к созданию и изменению структур знаний.

### **Веб-ориентированная архитектура.**

Веб-ориентированная архитектура (WOA) является подмножеством сервис-ориентированной архитектуры (SOA). Сервис-ориентированная архитектура (SOA, service-oriented architecture) – модульный подход к разработке программного обеспечения, основанный на использовании распределённых, слабо связанных заменяемых компонентов, оснащённых стандартизированными интерфейсами для взаимодействия по стандартизированным протоколам.

Веб-ориентированная архитектура отличается тем, что использует REST-подход. REST (Representational State Transfer, «передача состояния представления») – подход к архитектуре сетевых протоколов, обеспечивающих доступ к информационным ресурсам. Самой известной системой, построенной в значительной степени по архитектуре REST, является современная Всемирная паутина. Данные должны передаваться в виде небольшого

количества стандартных форматов (например HTML, XML, JSON). Сетевой протокол должен поддерживать кэширование, не должен зависеть от сетевого слоя, не должен сохранять информацию о состоянии между парами «запрос-ответ». Такой подход обеспечивает масштабируемость системы и позволяет ей легко эволюционировать с новыми требованиями.

### Инкрементный, многоуровневый подход.

Семантические технологии не меняют того факта, что большинство видов деятельности предприятия являются транзакционными, коммуникативными или документального характера. Структурированные, реляционные информационные системы – проверены, производительны и понятны. Семантические технологии не являются исходной базой, но это способ выражения и взаимосвязи результатов действий. Очевидную пользу семантические технологии дают предприятию в объединении всех существующих источников данных. Это должно стать целью первого семантического «слоя», и нужно добиться того, чтобы по максимуму использовалась уже имеющаяся информация. Такой подход является по своей сути инкрементным, и если все сделано правильно, то он обладает низкой стоимостью и низкими рисками.

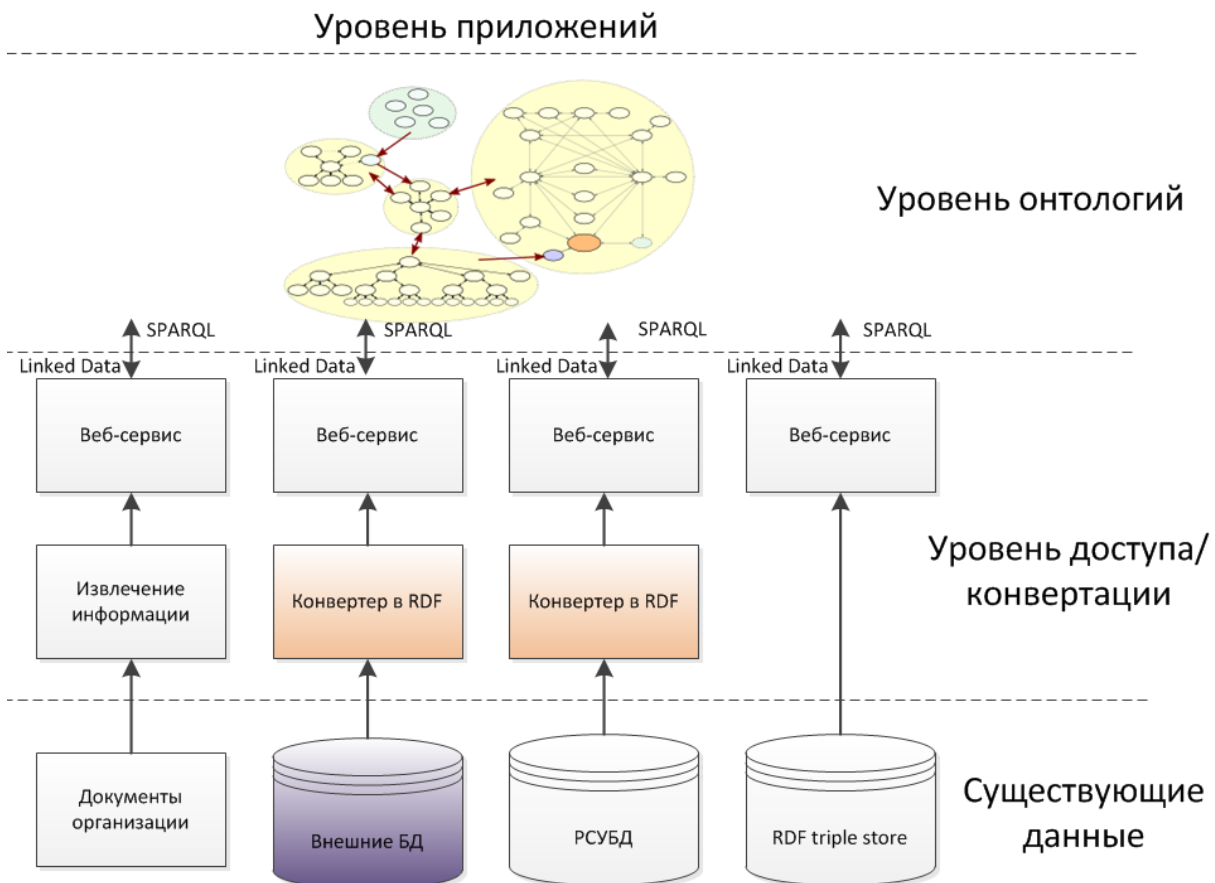


Рис. 3. Схема инкрементного подхода

Мы начинаем с существующих активов, данных, как внутренних, так и внешних, в самых различных форматах. Эти данные преобразуются в RDF-форму. Затем эти различные источники публикуются как веб-сервисы. Интегрируем и объединяем эту информацию с помощью адаптивных онтологий, после чего можно искать, проверять и управлять данными с помощью приложений, основанных на онтологиях.

#### IV. ПРЕИМУЩЕСТВА

##### **Интеграция разнородных данных.**

Компании часто сталкиваются с такой проблемой, когда, несмотря на фактическое наличие всех необходимых данных для принятия критических решений, воспользоваться этими данными нельзя, так как информация разбросана по различным специализированным системам, построенным в разное время и для разных целей. Если бы только они смогли повторно использовать эти осколки информации, полученные за счет проведенных капиталовложений, соединить их вместе в новом качестве, получив свежее достоверное знание, и затем суметь автоматизировать процесс быстрого принятия решений в зависимости от динамики ситуации.

Существующим техникам интеграции данных для полноценного решения подобной задачи недостаёт как масштабируемости, так и функционального охвата. Эти подходы в основном концентрируются на простом перемещении данных из одной системы в другую и применении трансформации и агрегации к элементам данных.

Расширяя техники интеграции информации и контента предприятия (EII и ECI) понятием семантики данных, семантическая интеграция претендует на то, чтобы предоставить полноценное решение проблемы. Семантика позволяет комбинировать и коррелировать элементы данных для установления взаимосвязи, определения тенденций и получения новой информации из новых комбинаций - информации, которая затем может использоваться для принятия решения.

Семантическая интеграция – это дополнение, а не замена стандартных методов, таких как извлечение данных, трансформация и загрузка (ETL) и техник интеграции информации предприятия (EII), но это дополнение заполняет критичный пробел в обеспечении необходимой практичности данных и их взаимосвязи. Потребность в объединении данных не ограничивается самим предприятием: государственные и правительственные источники данных становятся все более важными. Такие инициативы, как Linked Open Data [11] способствуют доступности государственных наборов данных с впечатляющими темпами роста за последние 5 лет. Главная цель открытых протоколов данных, таких как Linked Data, OData (Microsoft) и SAP Data Protocol (SAP) - избежать сильно изолированных систем хранения данных (т. н. data silos) и сделать данные доступными через Интернет.

##### **Гибкое изменение схем данных.**

Схемы данных и информации, как правило, не очень стабильны: они могут изменяться с течением времени, представления о сущностях и связях изменяются или расширяются, добавляются новые типы сущностей или отношений, в то время как другие типы или отношения могут стать устаревшими и, следовательно, должны быть исключены из информационной модели предприятия.

Предприятия должны справиться со следующей проблемой: прежде всего, слишком большая часть моделей данных и бизнес-логики по-прежнему жестко зашита в приложениях. По этой причине, вносить изменения в модель или бизнес-логику чревато ошибками и дорогостояще, как в плане финансовых, так и в плане трудозатрат.

Как происходит, например с базами данных: сначала разрабатывается концептуальная схема, потом она транслируется в реляционную модель, и только после того как реляционная модель готова, можно наполнять базу данными. Если даже модель данных можно в каких-то пределах менять, то это достаточно трудоёмко.

Теперь рассмотрим семантические технологии. Во-первых, благодаря высокой выразительности, можно большую часть бизнес-логики и информационной модели предприятия внести в семантическую модель, отсюда более четкое разделение приложения и данных/знаний. Во-вторых, нет необходимости сначала разрабатывать схему данных, можно всё хранить например в RDF-triple store, и после добавлять нужную информацию о схеме. Более того, изменения модели данных проводятся намного легче (по сравнению с реляционными БД).

Другая проблема состоит в том, что с распространением взаимодействия через Интернет кардинально изменилась динамика и скорость развития бизнеса. В динамичном бизнесе терминология, структура и специфика информации компании постоянно меняется. Чтобы справиться с частыми и неоднородными изменениями терминологии, таксономий и справочных данных, происходящими в отдельных информационных системах инфраструктуры компании, нужна поддержка адаптивных онтологий и быстрая и надежная синхронизация нормативно-справочных данных с помощью автоматизированных интеграционных процессов.

#### **Возможности семантического поиска.**

Когда встает вопрос взаимодействия пользователя и приложения, семантические технологии могут тоже предложить значительные преимущества. Упомянутые выше адаптивные онтологии, которые предназначены именно для человеческого понимания, покрывают бизнес-терминологию конечных пользователей, включая синонимы и омонимы. Это позволяет внедрять продвинутые технологии поиска данных, такие как фасетный поиск. Фасетный поиск – поиск с использованием фасетной классификации (совокупность нескольких независимых классификаций, осуществляемых одновременно по различным основаниям), многочисленных фильтров, основанных на классификациях информации по различным признакам. Естественным образом реализуется на основе онтологий (фасеты генерируются на основе иерархии концептов в онтологии), плюс учитываются отношения между сущностями в онтологии.

В конце концов, с помощью семантических возможностей немного приближается к реальности высокая цель поиска, когда ответ на поисковый запрос становится не «что пользователь сказал», а «что пользователь подразумевал».

#### **Совместная выработка знаний.**

Информация, как уже упоминалось выше, является одним из важнейших активов предприятия. Нужно принимать во внимание не только существующие в различных внутренних источниках и системах данные, но и знания самих работников предприятия. Перспективным преимуществом при внедрении семантических технологий является получение удобной интегрированной системы для совместной выработки знаний, которая включает в себя функции существующих в настоящее время платформ для коллаборации: социальных сетей, вики-систем, блогов.

## **V. ЗАКЛЮЧЕНИЕ**

Конечно, не всё так радужно в области семантических технологий, у них, как у каждого решения, существуют как преимущества, так и недостатки. Как уже говорилось в статье, семантические технологии действительно могут выполнить некоторые из своих обещаний, хотя и при переходе от научных исследований к реальному применению на предприятиях возникает новый набор проблем. Наиболее важные из них: масштабируемость и производительность семантических технологий, наличие квалифицированных кадров в области управления знаниями (особенно на постсоветском пространстве) и сложность количественной оценки практических преимуществ. В части производительности в последнее время есть значительные подвижки [3, 5]. По зрелому рассуждению, остальные проблемы не столь нерешаемы, чтобы хоть сколько-нибудь значительно откладывать практическое внедрение семантических технологий.

## **VI. ЛИТЕРАТУРА**

1. Семантическая паутина – Википедия [Электронный ресурс]. – Режим доступа: [http://ru.wikipedia.org/wiki/Семантическая\\_паутина](http://ru.wikipedia.org/wiki/Семантическая_паутина) (дата обращения: 13.06.12).

2. Bergman M.K. Seven Pillars of the Open Semantic Enterprise [Электронный ресурс] // AI3::Adaptive Innovation blog. – URL: <http://www.mkbergman.com/859/seven-pillars-of-the-open-semantic-enterprise/> (дата обращения: 15.05.12).

3. Хорошевский В.Ф. Семантические технологии: ожидания и тренды // Открытые семантические технологии проектирования интеллектуальных систем – Open Semantic Technologies for Intelligent Systems (OSTIS-2012): материалы II Междунар. научн.-техн. конф. (Минск, 16-18 февраля 2012 г.). – Минск : БГУИР, 2012. – с. 143-158.

4. Dau F. Semantic Technologies for Enterprises // Conceptual Structures for Discovering Knowledge – Proceedings of the 19th International Conference on Conceptual Structures (ICCS) 2011. – Berlin: Springer, 2011. – pp. 50-62.

5. Moulton L. Semantic Software Technologies: Landscape of High Value Applications for the Enterprise. Gilbane Group Report (2010) [Электронный ресурс]. – URL: [http://www.expertsystem.net/documenti/pdf\\_eng/technology/semanticsoftwaretechnologies\\_gilbane2010.pdf](http://www.expertsystem.net/documenti/pdf_eng/technology/semanticsoftwaretechnologies_gilbane2010.pdf) (дата обращения: 13.06.12).



## ADVANTAGES OF SEMANTIC TECHNOLOGIES: PRACTICAL ASPECTS

Evgeny V. Slesarev  
N. P. Ogarev's Mordovian State University,  
Russian Federation, Saransk city  
E-mail: slesarev@gmail.com, tel +7 (902) 6686457  
39, B. Hmel'nitskogo str., room 503, 430005, Saransk, Mordovia, Russia

**Annotation.** The article describes what semantic technologies are and points challenges and practical advantages in the semantic enterprise applications.

**Keywords:** Semantic Web; Semantic Enterprise; RDF; Linked Data; adaptive ontology; web-oriented architecture.