



OSTIS-2012

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

СЕМАНТИЧЕСКИЕ ТЕХНОЛОГИИ: ОЖИДАНИЯ И ТРЕНДЫ

Хорошевский В.Ф.

*Федеральное государственное бюджетное учреждение науки,
Вычислительный центр им. А.А. Дородницына РАН
г. Москва, Россия*

khor@ccas.ru

В статье обсуждается состояние исследований и разработок в области нового направления «Семантические технологии». Основное внимание уделяется анализу новых технологических трендов в данной области и экспертной оценке их значимости и сроков появления реальных прикладных продуктов на основе использования методов и средств семантической обработки информации.

Ключевые слова: научно-технический тренд, прикладная интеллектуальная система, семантические технологии, технологический триггер.

ВВЕДЕНИЕ

Исследования мировых лидеров в прогнозных исследованиях [IDC, 2012; Tofler, 2006; GARTNER, 2012] показывают, что «...до 2020 года количество информации и потребности в ней будут расти экспоненциально... Без умения создавать и обрабатывать такие объемы информации ЛПР будущего будут введены в состояние, которое можно назвать "аналитический паралич"...». Таким образом, одной из самых больших проблем современного общества является информационное переполнение. И в настоящее время уже осознано направление главного «удара» в борьбе с информационным взрывом – переход от хранения и обработки данных к накоплению и обработке знаний, что, в свою очередь, формирует новую семантическую волну (Рис. 1), которая, по оценке руководителя проекта Project 10X Миллса Дэвиса [Mills, 2006], существенным образом изменит характер работы с информацией.

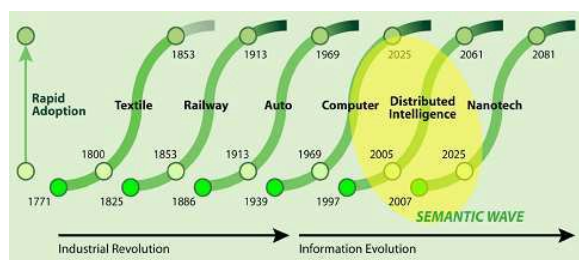


Рисунок 1 – Семантическая волна

Понятно, что в рамках такой семантической волны формируются и развиваются семантические

информационные технологии представления и обработки знаний, анализ зарождения и развития которых представляет значительный интерес.

Учитывая вышесказанное, целью настоящей статьи является обсуждение основных трендов в области семантических технологий и экспертная оценка тех ожиданий, которые связываются с их внедрением в практику. Организовано изложение следующим образом: сначала кратко представлены базисные семантические технологии, без которых, по нашему мнению, невозможно появление практически значимых семантических приложений, а затем, более подробно, обсуждаются некоторые из основных трендов в области семантических технологий.

1. Базовые тренды в тематической области «Семантические технологии»

Как представляется, базовые тренды в области семантических технологий в значительной мере связаны с концепцией Semantic Web (SW), которую в 2000 г. выдвинул Тим Бернерс-Ли (Tim Berners-Lee) – один из основоположников WWW и нынешний председатель консорциума W3C [Berners-Lee, 2003].

С момента появления этой концепции прошло уже более 10 лет, но пока SW-эра, в отличие от эпохи Интернет, еще только приближается и на этом пути существует значительное число научных, технических, технологических и человеческих проблем, основными из которых являются [Benjamins et al., 2002] доступность семантического контента, доступность онтологий и средств их

разработки, а также эволюция онтологий, масштабируемость, мультязыковость, визуализация и стабильность.

Доступность семантического контента является основной проблемой на пути формирования и использования пространств знаний, так как сейчас основная масса информации не представлена в «семантических» форматах и нет надежды, что эта работа может быть выполнена вручную. Онтологии, по мнению практически всех специалистов, являются ключевым компонентом в решении проблемы семантизации контента. В связи с этим особое значение приобретают проблемы онтологического инжиниринга, а также доступность уже существующих онтологий. Значительные усилия должны быть предприняты для хранения, обработки и поиска семантического контента, причем решения в этой области должны обеспечивать эффективную работу с огромными объемами знаний. Проблема мультязыковости контента существует давно, но для пространств знаний, где, по сути, должен поддерживаться эффективный доступ к информации независимо от того, на каком языке она представлена изначально, данная проблема является одной из основных. Представление информации для пользователей (визуализация контента) также должно претерпеть существенные изменения и обеспечить свободную ориентацию в огромном количестве фактов, которые отвечают его потребностям. Последняя по счету, но не по важности, проблема связана с обеспечением стабильности пространств знаний, а это, в свою очередь, предполагает, что существенные усилия должны быть предприняты в области стандартизации, обеспечивающей создание технологий, необходимых для формирования и сопровождения пространств знаний.

Понятно, что подробное рассмотрение и обсуждение даже только перечисленных выше базовых проблем и трендов семантических технологий в рамках одной статьи является практически безнадежным делом. Поэтому ниже мы сосредоточимся на нескольких трендах прикладных семантических технологий, а читателей, интересующихся базисными семантическими технологиями отсылаем к аналитическим обзорам, представленным в работах [Хорошевский, 2008; Хорошевский, 2009; Хорошевский, 2012].

2. Прикладные тренды семантических технологий

2.1. Предварительные замечания

Как известно, «тренд» (от англ. Trend) это долговременная общая тенденция изменения исследуемого временного ряда. В данном разделе обсуждаются технологические тренды – активно развивающиеся в последние 5 лет технологические направления в области семантических технологий, которые, как ожидается, продолжат свое развитие в ближайшие 10 лет. Обсуждение основано на

результатах анализа новых технологических трендов в области семантических технологий в рамках Программы фундаментальных исследований НИУ ВШЭ в 2011 году, в которых принимал участие автор настоящей работы. В процессе этих исследований был проведен опрос экспертов, в результате чего для более детального анализа был сформирован перечень основных технологических трендов в области «Семантические технологии», куда вошли:

- Открытые связанные данные (Linked Open Data);
- Социальный семантический Вэб (Social Semantic Web);
- Бизнес-анализ на базе семантических технологий (Semantic Business Intelligence);
- Семантические цифровые библиотеки (Semantic Digital Libraries) и
- Семантические технологии в приложениях (Semantic-based e-Apps).

Учитывая ограничения на объем настоящей работы, ниже обсуждаются подробнее три из них – открытые связанные данные, социальный семантический Вэб и семантические цифровые библиотеки.

2.2. Открытые связанные данные

Следует сразу отметить, что термин Linked Open Data является не вполне удачной альтернативой для термина Linked Data, поскольку последний представляет действительно технологию, а первый – название проекта, который задал направление исследований и разработок в области связывания данных в Web-пространстве. Вместе с тем, в настоящее время термин Linked Open Data часто используется как альтернативный и/или даже синонимичный к термину Linked Data.

2.2.1. Описание тренда

Термин Linked Data соответствует одной из центральных идей Семантического Вэба, в основе которой лежит то, что первичными объектами здесь являются описания сущностей с явным указанием их семантики и семантики ссылок (отношений) между ними. Технологически это обеспечивается представлением данных в виде триплетов «субъект-предикат-объект» на языке RDF, идентификацией данных с помощью URI, механизмом доступа по протоколу HTTP, а также спецификацией контролируемых словарей на языках RDFS и OWL. Таким образом, понятие Linked Data относится к установлению наилучших способов публикации и связывания данных в Web-пространстве [Bizer et al., 2009], а характерной чертой связанных данных является их ориентация на Вэб и Семантический Вэб, позволяющая производить разделение данных на глобальном уровне между множественными разнородными источниками.

В представленной выше трактовке Linked Open Data является технологическим трендом появление которого достаточно четко связывается со статьей

Тима Бернерса-Ли [Berners-Lee, 2006], где были сформулированы следующие правила (Linked Data principles) публикации данных на Вэбе:

- Использование URI как имен сущностей.
- Использование HTTP URI, чтобы люди могли «видеть» и использовать эти имена.
- Использование стандартов RDF и SPARQL для получения полезной информации через посредство URI.
- Включение в публикации связей с другими URI и, таким образом, обеспечение возможностей отражения новых связанных сущностей.

Сформулированные Тимом Бернерсом-Ли принципы технологии Linked Data, по существу, определили подходы и методы их реализации, которые в значительной мере использовали результаты предшествующих исследований и разработок в области объединения и совместного использования распределенных баз данных, а также в области представления и обработки знаний в среде Интернет.

Вместе с тем, становление и развитие LD-технологии поставило перед исследователями и разработчиками, работающими в данной области, новые научно-технические проблемы. К основным из них, которые не решены полностью до настоящего времени, относятся:

- Проблема функционально полных и то же время эффективных методов и средств спецификации моделей пространств знаний (онтологическое моделирование предметных областей).
- Проблема объединения и выравнивания (Merging & Alignment) онтологических моделей, разработанных разными коллективами и используемых в разных приложениях, а также валидация и верификация онтологических моделей.
- Проблема оценивания связываемых данных и формализации методов определения доверия к используемым данным.

Отметим, что решение первой из указанных проблем в настоящее время связывается, в первую очередь, с использованием уже разработанных стандартов W3C, среди которых основную роль играют языки спецификации онтологий семейства OWL [OWL, 2009] и язык запросов к RDF-хранилищам SPARQL [SPARQL, 2008].

Одновременно с научно-техническими проблемами развитие LD-технологии потребовало решения ряда технологических задач, среди которых, в первую очередь необходимо выделить

- эффективную реализацию RDF-хранилищ,
- масштабирование связанных данных и
- эффективность доступа к связанным данным в условиях экспоненциально увеличивающихся объемов числа запросов на данные.

Решение этих проблем в настоящее время идет по нескольким направлениям. С одной стороны, разрабатываются новые методы и средства

эффективной реализации RDF-хранилищ и эффективного хранения данных в RDF-хранилищах [Лапшин, 2010], а с другой – предпринимаются усилия по реализации RDF-хранилищ в среде облачных вычислений.

В целом же можно констатировать, что в настоящее время технология связанных данных активно развивается в проектах открытых связанных данных (Linked Open Data), формируя тем самым новый тренд – «Открытые связанные данные».

При этом интересно отметить, что научно-технические проблемы LD-технологии практически непонятны ни бизнес-сообществу, участвующему в финансировании работ в данной области, ни потребителям полученных здесь решений. И более того, потребители результатов LD-технологии и LOD-проектов могут даже не подозревать, что в основе конечных продуктов и сервисов, которыми они пользуются, лежат инновационные решения из области семантических технологий. Указанная ситуация связана с тем, что LD-технология является базовой и обеспечивает разработку новых информационных пространств, но не имеет конечных потребительских свойств, очевидных для пользователей.

2.2.2. Проекты, продукты и сервисы тренда

С учетом вышесказанного, в основе данного технологического тренда в настоящее время лежат не столько продукты, сколько проекты в области открытых связанных данных (LOD-проекты) и те сервисы, которые постепенно формируются на их основе.

Как известно, первым проектом нового поколения в области связанных данных был проект LOD [LOD, 2012], выполняемый в рамках SWEO (Semantic Web Education and Outreach) Interest Group, миссия которой в расширении взаимодействия в Вэб-сообществе и в обучении необходимым для этого решениям и технологиям. Целью общего движения за открытые данные было и остается желание сделать данные свободно доступными для всех заинтересованных организаций и частных лиц. При этом уже до начала собственно проекта W3C SWEO существовали разные интересные наборы открытых данных, доступные на Вэбе, среди которых наиболее известными были Wikipedia, Wikibooks, Geonames, MusicBrainz, WordNet, the DBLP bibliography и другие, опубликованные под лицензиями Creative Commons или Talis.

Учитывая вышесказанное, целью проекта LOD было расширение Вэба различными уже опубликованными открытыми данными за счет использования RDF-представлений и установления RDF-связей между «атомами» данных из различных источников данных, что должно обеспечить навигацию от атомарных данных одних источников данных к соответствующим атомарным данным других источников с использованием браузеров

Semantic Web и, кроме того, индексацию RDF-связей краулерами семантических поисковиков. Такой подход, по мнению авторов проекта, должен обеспечить интеллектуальный поиск и возможности запросов на данные, результатами которых были бы не просто ссылки на HTML-страницы, а структурированные данные, которые бы могли использоваться другими приложениями.

В настоящее время (сентябрь 2011 г.) в рамках данного проекта уже опубликовано 295 наборов данных, содержащих более 31 блн. RDF-триплов, которые «перевязаны» около 504 млн. RDF-связей.

При этом аргументы «ЗА» и «ПРОТИВ» открытых данных хорошо известны и сводятся, в частности, к следующему:

- Данные принадлежат всему человечеству.
- Для выполнения работ использованы деньги общества и потому данные должны быть в свободном доступе.
- Данные создавались правительственными организациями (обычно это US National Laboratories и другие государственные агентства).
- Факты не могут быть копирайтными легально.
- Спонсоры исследований не получают полной «отдачи» пока результирующие данные находятся в свободном доступе.
- и т.д.

С учетом вышесказанного, новые технологии открытых связанных данных могут стать препятствием для их внедрения в повседневную практику использования, хотя полезность и целесообразность их несомненны.

Кроме рассмотренного выше проекта LOD наиболее известными в данной области в настоящее время являются:

- Проект европейского сообщества Linked open data around the clock (LATC) [LATC, 2012], который выполняется в рамках 7-й Рамочной Программы и направлен на поддержку публикации и использования связанных открытых данных, на совершенствование инфраструктуры для мониторинга использования и повышение качества связанных открытых данных; снижение барьеров доступа к данным для издателей и потребителей; разработку библиотеки инструментов обработки открытых источников данных; эксперименты по обработке связанных данных в комбинации с данными Европейского Союза и поддержку сообщества учебной литературой и лучшими примерами (best practices).
- Проект европейского сообщества PlanetData [PlanetData, 2012], целью которого является формирование сети достижений, где были бы объединены европейские исследователи в области управления данными большой размерности, в которой бы были представлены RDF-данные, опубликованные в соответствии с

принципами Linked Data principles. Проект Planet Data уникален тем, что открыт для новых партнеров в процессе его выполнения.

- Европейский проект Linking Open Data 2 [LOD2, 2012] – часть 7-й Рамочной Программы Европейского Сообщества с бюджетом в 6.5 млн. Евро (2010-2014 г.г.), который продолжает работы по проекту LOD и ставит своей целью создание пространств знаний через взаимосвязанные данные путем разработки промышленных инструментов и методологий для экспозиции и управления очень большими объемами структурированной информации из Вэба данных (Data Web); испытательного стенда и расширяемой сети высококачественных мультимедийных и многоязыковых онтологий из таких источников, как Wikipedia и OpenStreetMap; алгоритмов на основе машинного обучения для автоматического связывания и объединения данных, представленных на Вэбе; стандартов и методов для надежного отслеживания «происхождения», обеспечения приватности и безопасности данных, а также достижения высокого качества информации; адаптивных средств поиска, навигации и авторства связанных данных.

Приведенные аннотации LOD-проектов еще раз подтверждают представленное выше утверждение о том, что внедрение технологии L(O)D только предстоит, а пока идет работа по ее реализации, поскольку, по определению, технология это совокупность методов и средств, направленных на достижение определенных целей.

Из продуктов и сервисов в области охвата технологического тренда «Открытые связанные данные» в настоящее время можно отметить браузеры и навигаторы связанных данных, визуализаторы связанных данных и редакторы запросов.

2.2.3. Новые потребительские свойства и ожидаемые эффекты тренда

Новые потребительские свойства, присущие технологическому обсуждаемому тренду определяются тем, что результаты развития L(O)D-технологии несут в себе новые возможности в формировании пространств взаимосвязанных данных в сети Интернет, приближая тем самым переход от Вэба данных к Вэбу знаний. Такой переход будет характеризоваться принципиально новыми возможностями поиска и навигации по данным в сети Интернет, с одной стороны, и переходом от информационных систем к информационно-аналитическим системам нового поколения практически во всех предметных областях.

При этом наиболее востребованными, на наш взгляд, зонтичными областями, где использование L(O)D-технологий может дать максимальный эффект являются

- информационные системы органов государственной власти всех уровней;
- социальные сети;
- информационно-аналитические сети научных и профессиональных сообществ;
- системы дистанционного обучения и
- системы экспертных консультаций.

Среди важнейших ожидаемых эффектов технологического тренда «Открытые связанные данные», в первую очередь, целесообразно выделить: социальный (существенное увеличение открытости данных и информационной прозрачности органов власти, вовлечение новых слоев гражданского общества в использование открытых данных и расширение социальных сетей); экономический (сокращение сроков поиска нужной информации, облегчение поиска бизнес-партнеров и необходимых для производства материальных, производственных и людских ресурсов, сокращение сроков и повышение качества маркетинга; политический (повышение доверия граждан к власти и демократизация общества за счет вовлечения в процессы принятия решений значительных слоев населения).

2.2.4. Оценка сроков внедрения приложений и возможные области применения тренда

Как показывает анализ, основные работы в рамках технологического тренда «Открытые связанные данные» ведутся на стадиях фундаментальных и прикладных исследований, результаты которых начинают использоваться при создании опытных образцов.

Учитывая вышесказанное, вероятный срок появления на рынке важнейших продуктов и услуг в рамках технологического тренда «Открытые связанные данные» определяется сроками завершения основных фундаментальных и прикладных исследований после чего должно пройти еще 1-3 года. Таким образом, если считать, что основные фундаментальные и прикладные исследования дадут необходимые для практики результаты высокого качества в течение 5-ти ближайших лет, то реального появления на рынке важнейших продуктов/услуг в рамках обсуждаемого технологического тренда можно ожидать не ранее 2017-2020 г.г.

Следует также отметить, что прогнозы сообщества R&D значительно отличаются от прогнозов пользователей и даже от прогнозов бизнес-сообщества (Рис. 2).

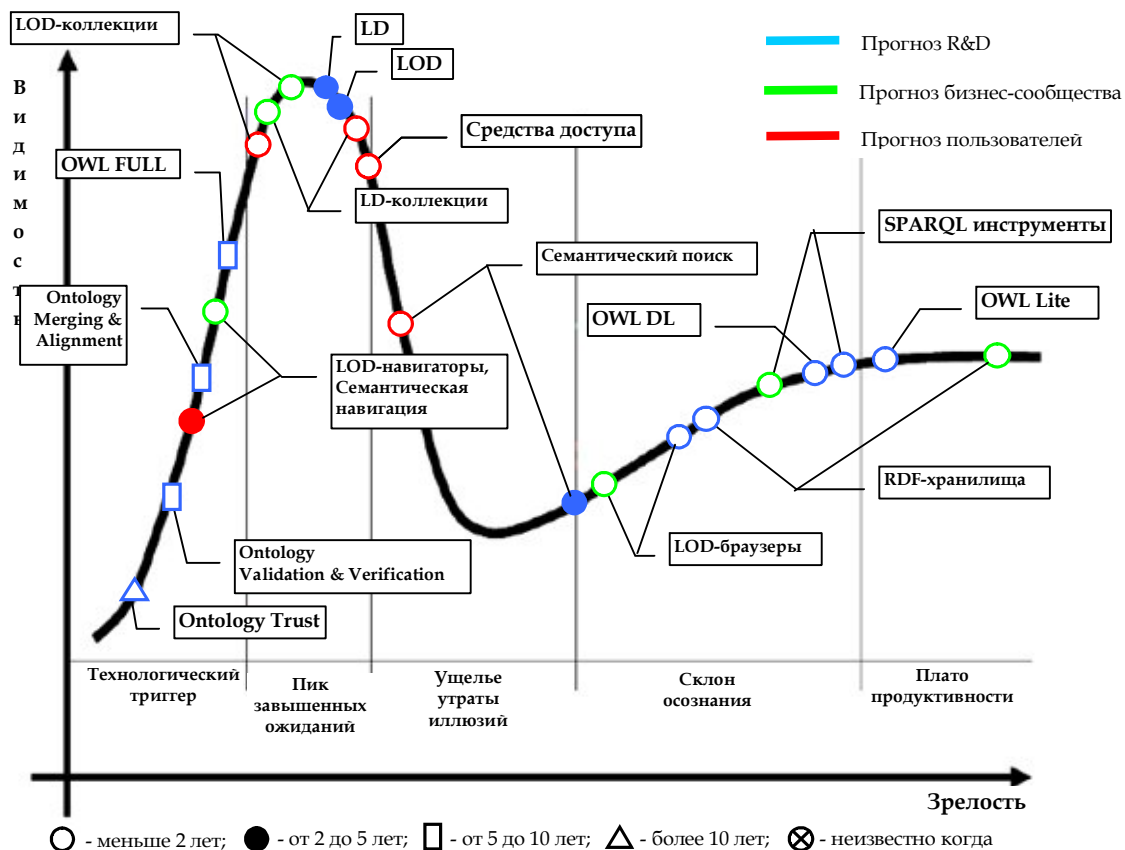


Рисунок 2 – Прогнозная кривая для технологического тренда «Открытые Связанные Данные»

Как отмечалось выше, технологический тренд «Открытые связанные данные» является базисным для всех областей применения, где уже имеются представительные множества коллекций востребованных данных. По нашему мнению, в первую очередь, это

- области фундаментальных и прикладных исследований, для которых необходимо формирование и активное использование общих пространств знаний;
- государственное управление всех уровней;
- новые государственные и коммерческие информационные системы.

2.2.1. Технологии-конкуренты и технологии, от которых зависит развитие тренда

Как представляется, в качестве основного конкурирующего технологического направления по отношению к тренду «Открытые связанные данные» можно указать тренд «Интеграция распределенных баз данных», основными преимуществами которого можно считать существенно более длительное время развития данного направления; наличие центров компетенции по данному технологическому направлению во всем мире; многолетнюю работу национальных и международных комитетов и организаций по стандартизации способов представления и обработки информации в базах данных; сформировавшиеся рынки производителей программного обеспечения СУБД и поставщиков контента, а также опыт потребителей в использовании продуктов этого технологического направления во всем мире.

Среди недостатков технологии-конкурента можно отметить неспособность в рамках базисной парадигмы легкой интеграции в среду Интернет; медленную адаптацию основных вендоров СУБД к потребностям потребителей в эффективных и легко масштабируемых RDF-хранилищах, а также отсутствие общезначимых и признанных результатов в области формирования и использования пространств знаний.

Развитие технологического тренда «Открытые связанные данные», по нашему мнению, зависит от таких технологий из смежных областей, как представление и обработка знаний; онтологический инжиниринг; семантический поиск информации и аналитика на знаниях.

2.2.5. Разрушающий потенциал, барьеры и риски внедрения тренда

Разрушающий потенциал технологии связанных данных состоит, прежде всего, в том, что продукты и сервисы этой технологии, будучи введенными в массовое использование, существенно изменят рынок поставки информации потребителям. Изменится и технологический ландшафт информационных технологий, в котором, по нашему мнению, ведущее положение займут поставщики

семантизированного контента. Нарушится существующий баланс сил в отраслевом сегменте за счет того, что наиболее успешными будут те технологические компании, которые первыми выйдут на рынок программных продуктов, поддерживающих создание и обработку больших и сверхбольших пространств открытых связанных данных.

При этом барьеры и риски внедрения тренда, по нашему мнению, определяются следующими факторами:

- Завышенные ожидания потребителей продуктов технологического тренда «Открытые связанные данные» относительно реальных возможностей производителей программных средств для эффективной поддержки процессов формирования и обработки сверхбольших объемов связанных данных.
- Слабость законодательной базы, регламентирующей условия и способы свободного использования открытых связанных данных.
- Относительно небольшие объемы коллекций открытых связанных данных, доступных потребителям.

2.2.6. Заключительные замечания по тренду

В настоящем разделе рассмотрены некоторые из основных аспектов технологического тренда «Открытые связанные данные», дано описание тренда, кратко представлены проекты, продукты и сервисы, а также новые потребительские свойства и ожидаемые эффекты тренда и дана оценка сроков внедрения приложений и возможные области его применения. Идентифицированы технологии-конкуренты и технологии, от которых зависит развитие тренда, фиксирован разрушающий потенциал, барьеры и риски внедрения тренда.

2.3. Социальный семантический Вэб

Одной из самых известных диаграмм развития Интернет является диаграмма Миллза Дэвиса, представленная на Рис. 3.

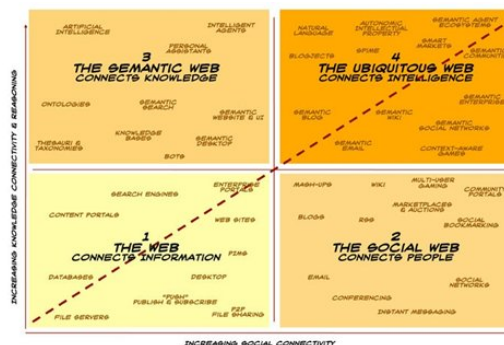


Рисунок 3 – Эволюция Интернет до 2020 года

Как следует из диаграммы, технологический тренд «Социальный Вэб» является одной из 4-х важнейших составляющих эволюции Интернет,

который сформировался в начале XXI века и устойчиво развивается уже более 5 лет. При этом основные социальные сети демонстрируют впечатляющие характеристики: 800 блн. мин. сообщений на площадке FaceBook; 90 млн. «переговоров»/день на площадке Twitter; 24 часа видео/мин на площадке YouTube; более 200 языков в среде Wikipedia и т.д., и т.п., что показывает высокую востребованность социальных сетей.

Вместе с тем, в последние несколько лет появился новый технологический тренд «Социальный Семантический Вэб», обсуждению которого посвящен настоящий подраздел.

2.3.1. Описание тренда

Основными составляющими технологического тренда «Социальный Семантический Вэб» являются интеграция уже активно используемых в таких продуктах социальных сетей, как блоги, форумы, энциклопедии и т.п., средств коммуникации и совместной работы, со средствами и технологиями Семантического Вэба. При этом используются как общие средства уровня RDF(S), OWL, SPARQL, так и более простые модели внесения семантики, например, микроформаты.

По сути дела, этот тренд опирается на концепцию Open Distributed Semantic (Открытость-Распределенность-Семантичность), в рамках которой составляющая Open предполагает возможности разделения персональных данных в любой момент и в любой точке по желанию пользователя, составляющая Distributed – на пресуппозицию хранения персональных данных в персональном месте, а составляющая Semantic предполагает обогащение персональных данных таким образом, чтобы наполнить их смыслом.

Основными технологиями, обеспечивающими данную концепцию, являются Семантический Вэб (Semantic Web) и Связанные Открытые Данные (Linked Open Data), которые интегрируются в рамках технологии Социальных Сетей.

В новом технологическом тренде «Социальный Семантический Вэб» эти технологии используются на уровне формализмов RDF(S) и RDFa, SPARQL и описания онтологических моделей с помощью языков семейства OWL, а также применения средств семантического тэгирования.

В рамках нового тренда объединение людей происходит не на уровне общей площадки для коммуникации, а на уровне объектно-центрированной социализации (Object-centred sociality). Согласно данной концепции люди не просто объединяются друг с другом в социальной сети, но объединяются через общие разделяемые объекты. Поэтому полезные социальные продукты и сервисы должны обеспечивать создание социальных объектов, объединяющих людей. При этом формируются социальные сети по интересам – работа, учебные заведения, хобби и т.п. Так, например, площадка Flickr объединяет ее

посетителей по объекту «фото» (photos), площадка del.icio.us ориентирована на обмен «закладками» (bookmarks) и т.д. Вместе с тем, в настоящее время все большее внимание привлекают площадки, на которых возможна интеграция интересов, что обеспечивает развитие нового тренда «Социальный Семантический Вэб».

Таким образом, обсуждаемый технологический тренд не является результатом собственных фундаментальных и прикладных исследований, а интегрирует результаты других технологических трендов – «Семантический Вэб», «Связанные Открытые Данные» и «Социальный Вэб». При этом новый технологический тренд активно формирует собственные продукты и сервисы, кратко представленные в следующем подразделе.

2.3.2. Продукты, сервисы и проекты тренда

Продукты и услуги социального Вэба (Web 2.0) хорошо известны – это, в первую очередь, площадки Facebook, Flickr, LinkedIn, Twitter, Slideshare, Wikipedia и многие др., активно используемые во всем мире, а также отечественные социальные сети «ВКонтакте», «Мой мир», «Одноклассники», «Ютуб», «Википедия», «ЛайвДжорнал» и др.

Менее известны продукты и услуги технологического тренда «Социальный Семантический Вэб», поскольку здесь наблюдается 2 «потока»: вовлечение семантических технологий в уже известные продукты давно работающих на данном рынке производителей и в новые проекты, а также продукты/услуги новых компаний, которые появились совсем недавно.

Среди «ветеранов» социального Вэба, активно использующих семантические технологии, по нашему мнению, лидируют Freebase и Facebook, продукты которых входят в Top 10 Semantic Web Products в 2009-2010 г.г. по версии авторитетного издания ReadWriteWeb [ReadWriteWeb, 2010].

Как известно, Freebase открытая база данных, объекты в которой проиндексированы семантическими тэгами. В этом смысле база Freebase похожа на базу Wikipedia, но в базе Freebase представлена информация о структурированных данных и о том, что с ними можно делать. При этом Freebase является одним из явных адептов Semantic Web и Linked Data. Интересно, что база Freebase недавно была приобретена Google.

Новый «рывок» Facebook в сегменте социальных сетей, на наш взгляд, можно связать с анонсированной новой платформой Open Graph, целью которой является предоставление издателям возможностей интеграции их Вэб-страниц в социальные графы. При этом каждая страница становится объектом социального графа Facebook, через который пользователи могут связываться друг с другом в своих сетях. И более того, это означает, что страницы могут связываться в рамках социальной сети через профили пользователей,

сообщения в блогах, результаты поиска, новостные ленты Facebook и т.п.

Серьезные инструментальные средства для семантических социальных сетей выведены на рынок сервисом Zemanta [Zemanta, 2012]. По существу это инструменты блоггеров (blogging tool), которые базируются на семантических технологиях для включения релевантного контента в сообщения (заметки) пользователей. Основные возможности инструментария Zemanta связаны с легким поиском тематических фотографий к постам, быстрым поиском релевантных постов, автоматическим формированием ссылок на авторитетные сайты и подбором наиболее подходящих тэгов для SEO. Программное обеспечение Zemanta распространяется как Open Source и отвечает стандартам Semantic Web.

Активную позицию в области социальных семантических сервисов демонстрирует в последнее время и Google, который «запустил» новые сервисы Google Squared и Google Search Options and Rich Snippets.

Google Squared поддерживает обработку простых вопросов на естественном языке, дает простые ответы и вводит новую возможность в рамках Google's sidebar, которая называется "Something different", где представляется список результатов поиска, которые могут находиться в области интересов пользователя. Технология использует анализ текущих поисковых терминов пользователя. Сервис Google Search Options and Rich Snippets является расширением «ядерного» поисковика Google за счет обогащения найденных фрагментов структурным контентом. При этом используются открытые стандарты структурированных данных (микроформаты и RDFa).

Серьезной заявкой на социальный семантический Вэб является инициатива SIOC [SIOC, 2012] (Semantically-Interlinked Online Communities), цели которой – следующие:

- Семантически связать существующие online-сообщества.
- Полностью описать контент/структуру социальных Вэб-сайтов.
- Создать новые связи между online-дискуссиями, форумами и контейнерами.
- Обеспечить интеграцию информации online-сообществ.
- Поддержать инновационный просмотр связанных элементов Social Web.
- Разрешить «chicken-and-egg» проблему с Semantic Web.

Основным подходом к решению поставленных задач инициатива SIOC считает онтологический инжиниринг. Уже в настоящее время SIOC-онтология, опубликованная как W3C Member Submission, поддержана 16 организациями.

Другими интересными проектами в области

охвата технологического тренда «Социальный Семантический Вэб» являются:

- SCOT (Social Semantic Cloud of Tags) [SCOT, 2012] - проект, целями которого является создание модели для описания tagclouds (тэги и совместное их появление), обеспечение возможностей перемещения собственных tagcloud от одного сервиса к другому, разделение tagclouds между сервисами и между пользователями, а также обеспечение портативности тэгов (Tag portability).
- MOAT (Meaning Of A Tag) [MOAT, 2012] - проект с целями создания модели для описания смысла тэгов, «перевязки» тэгов, управляемой пользователем, интеграции тэгированного контента с Вэбом связанных данных, а также коллаборативный подход к «разделению» смыслов в сообществе.
- CommonTag [CommonTag, 2012] - проект, в котором объединяются усилия AdaptiveBlue, DERI at NUI Galway, Faviki, Freebase, Yahoo!, Zemanta и Zigtag для разработки формата связи значимых ресурсов необходимыми тэгами.

Общая схема взаимосвязи рассмотренных выше проектов и объектов социального Вэба, что, собственно, и знаменует собой переход к социальному семантическому Вэбу, представлена на Рис. 4.

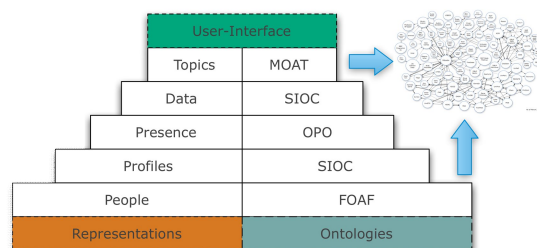


Рисунок 4 – Общая схема взаимосвязи проектов и объектов социального семантического Вэба

2.3.3. Новые потребительские свойства и ожидаемые эффекты тренда

Новые потребительские свойства, присущие технологическому тренду «Социальный Семантический Вэб», определяются, по нашему мнению, в значительной мере результатами развития семантических технологий в целом. При этом основную роль должны сыграть новые методы и средства извлечения информации из ЕЯ-текстов, методы семантического индексирования контента на основе онтологических моделей, методы семантического поиска информации, а также новые методы и средства семантического связывания данных в социальных сетях.

Такой переход будет характеризоваться принципиально новыми возможностями в социальных сетях за счет

- внесения семантики в данные, представленные в социальных сетях в настоящее время;

- явной спецификации смысловых связей между данными различных социальных сетей и за счет
- формирования новых сообществ, в которые будут вовлечены разные люди из уже существующих сообществ, и динамического изменения состава и размера существующих сообществ.

Представляется интересным и такое новое свойство социальных семантических сетей, как появление в них разделов многопользовательских игр, в которых участвуют члены социальной сети, а также зарождение нового технологического тренда Enterprise 2.0, в рамках которого формируются социальные бизнес-сети.

Среди важнейших ожидаемых эффектов нового технологического тренда, в первую очередь, целесообразно выделить: социальный (существенное увеличение «осмысленности» данных, циркулирующих в сообществах, представленных на той или иной площадке социальных сетей, вовлечение новых слоев общества в активное использование социальных сетей в повседневной жизни, интеграция социальных сетей и, как следствие, с увеличением взаимодействия между разными слоями общества, появление нового класса социальных сетей – сетей экспертных сообществ, предоставляющих независимые консультации в области своей компетенции); экономический (появление социальных бизнес-сетей, где производители товаров и услуг смогут быстро получать реальный отклик на свои новые решения, а потребители получат возможности разносторонней оценки предоставляемых товаров и услуг, а также быстрый доступ к приобретению нужных товаров и услуг); политический (возможность создания социальных сетей, где формируется общественное мнение по поводу политики властей).

2.3.4. Оценка сроков внедрения приложений и возможные области применения тренда

Основные работы в рамках технологического тренда «Социальный Семантический Вэб» ведутся, как это не странно на первый взгляд, на стадиях создания опытных образцов и запуска их в массовое использование. В этом смысле технологический тренд «Социальный Семантический Вэб» имеет гораздо более благоприятную почву для своего развития в виде востребованного обществом социального Вэба и уже существующими продуктами и сервисами для его массового использования, чем другие технологические тренды, обсуждаемые в настоящей работе.

Прогнозная кривая технологического тренда «Социальный Семантический Вэб» представлена на Рис. 5 и показывает, что в данном случае имеется хорошее соответствие между мнением бизнес-сообщества и потребителями (хотя последние, по нашему мнению, дают более оптимистические прогнозы) и понятное расхождение во мнениях между бизнес-сообществом и сообществом R&D, поскольку первое считает, что переход к

семантическому социальному Вэбу будет естественным продолжением работ по Вэбу 2.0, а второе – что на пути к реально функционирующему семантическому социальному Вэбу предстоит решить много наукоемких и технически сложных задач.

2.3.5. Технологии-конкуренты и технологии, от которых зависит развитие тренда

Конкурирующими технологиями по отношению к семантическим социальным сетям, по нашему мнению, являются интерактивные медийные технологии и, в частности, средства коллективной мобильной связи, обеспечивающие целевую доставку контента потребителям, и интерактивное телевидение. При этом надо отметить, что социальные семантические сети очень быстро интегрируют конкурентные продукты в качестве своих сервисов и используют конкурирующие технологии для развития собственных продуктов и сервисов. Хорошим примером интеграции конкурентных продуктов являются социальные игры, приобретающие все большую популярность в социальных сетях.

Преимущества технологий-конкурентов, по-видимому, связаны с их большей «понятностью» для старших поколений населения, а недостатки определяются ограничениями технических средств поддержки этих технологий.

По нашему мнению, развитие технологического тренда «Социальные Семантические Сети» опирается на результаты «родительского» технологического тренда «Социальные Сети» и существенно зависит от таких технологий из смежных областей, как представление и обработка знаний; онтологический инжиниринг; семантизация контента; семантический поиск информации и экспертные системы.

Специфика обсуждаемого технологического тренда, по нашему мнению, состоит в том, что массовое производство продуктов здесь не требуется, а массовое использование существующих продуктов и их новых версий будет происходить одновременно с появлением первых важнейших продуктов и/или услуг. При этом, по нашему мнению, целесообразно эти продукты/услуги рассматривать в 2-х аспектах: продукты и сервисы первого поколения, к которым мы относим модификации существующих платформ путем частичного внедрения в них семантики, и продукты и сервисы второго поколения, где будут в полном объеме использоваться разработанные и реализованные к тому времени решения в области семантических технологий, в первую очередь, методов и средств семантизации контента, Semantic Web, L(O)D-продуктов и программно-аппаратных средств передачи информации в сетях. Для первого поколения наша экспертная оценка 2013 – 2015 г.г., а для второго –: 2014 – 2018 г.г.

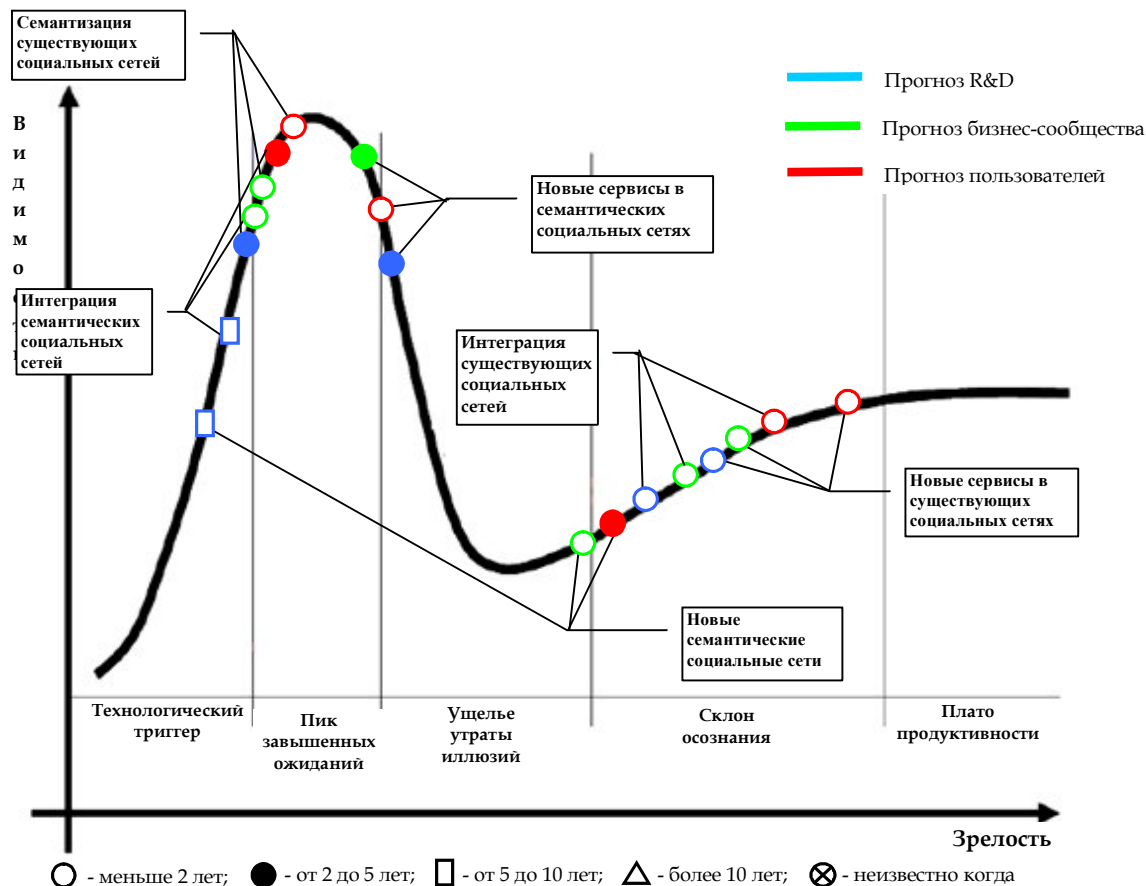


Рисунок 5 – Прогнозная кривая для технологического тренда «Социальный Семантический Вэб»

2.3.6. Разрушающий потенциал, барьеры и риски внедрения тренда

По нашему мнению, разрушающего потенциала у технологического тренда «Социальный Семантический Вэб» нет, поскольку его продукты и сервисы являются естественным развитием существующих продуктов и сервисов социального Вэба. Будучи введенными в массовое использование, продукты и услуги этого технологического тренда просто расширят рынок поставки информации потребителям, а технологический ландшафт при этом, по нашему мнению, практически не изменится. Не нарушится существенно и баланс сил в отраслевом сегменте, поскольку основные игроки сами активно продвигают данный технологический тренд.

Барьеры, риски внедрения технологии и факторы неопределенности развития технологии связаны, прежде всего, с неразвитостью законодательной базы для использования данных, циркулирующих в социальных сетях и в Интернет в целом, а также с научно-техническими проблемами семантизации контента и семантических сервисов.

2.3.7. Выводы по тренду

Как и в предыдущем случае, выше рассмотрены

некоторые из основных аспектов технологического тренда «Социальный Семантический Вэб», дано описание тренда, кратко представлены проекты, продукты и сервисы, а также новые потребительские свойства и ожидаемые эффекты тренда и дана оценка сроков внедрения приложений и возможные области его применения. Идентифицированы технологии-конкуренты и технологии, от которых зависит развитие тренда, фиксирован разрушающий потенциал, барьеры и риски внедрения тренда.

2.4. Семантические электронные библиотеки

С ростом компьютерной грамотности всё большее количество людей начинает пользоваться электронными книгами, а число читателей обычных библиотек снижается. Так за период 1997—2002 годы в Университете Айдахо число посетителей библиотеки снизилось более чем на 20 %, а число пользователей электронных версий только за период 1999—2002 увеличилось на 350 %. В связи с этим многие библиотеки начали создавать электронные версии хранящихся в их фондах книг и уже начиная с 70-х годов прошлого века стартовали проекты по созданию электронных библиотек.

Как известно, первым проектом по созданию электронной библиотеки был Проект «Гутенберг» (1971 г.). В Рунете первой электронной библиотекой стала библиотека Максима Мошкова. В 1990 году библиотекой конгресса США начат проект «Память Америки», в рамках которого предоставляется свободный и бесплатный доступ к электронным материалам по истории США. В 2002 году Google начал собственный проект по оцифровке книг, а в декабре 2004 года было объявлено о начале работы библиотечного проекта «Google Print», который в 2005 году был переименован в «Поиск книг Google». 20 ноября 2008 года начала функционировать общеевропейская цифровая библиотека Europeana, а 21 апреля 2009 года состоялось официальное открытие Всемирной цифровой библиотеки. Активно развиваются электронные библиотеки и в России. 27 мая 2009 года в Санкт-Петербурге была открыта президентская библиотека имени Бориса Ельцина, в задачи которой входит предоставления электронных материалов по истории России. Серьезные научно-практические проекты по тематике электронных библиотек ведутся и в таких организациях, как Российская Государственная Библиотека (РГБ), Национальная Электронная Библиотека (НЭБ), Государственная Публичная Научно-Техническая Библиотека (ГПНТБ), Центре Информационных Технологий и Систем (ЦИТиС), Библиотеке Естественных Наук РАН (БЕН) и в других организациях. Таким образом, технологический тренд «Электронные библиотеки» имеет достаточно долгую историю и впечатляющие результаты во всем мире и в нашей стране.

2.4.1. Описание тренда

Вместе с тем, в последние годы наблюдается формирование нового технологического тренда – «Семантические Электронные Библиотеки» (СЭБ), которые являются хранилищами семантических моделей библиотечных объектов. И, если в обычных электронных библиотеках «читатели» пользуются проиндексированными материалами на базе средств контекстного поиска, в семантической электронной библиотеке поиск материалов осуществляется на базе предварительной семантической разметки и индексации единиц хранения. Для этого используются семантические модели библиотечных объектов, где представлены модели библиотечных материалов (к-во стр., автор, и т.п.) и модели их содержания. При этом для описания таких моделей в СЭБ используются методы и средства онтологического моделирования пространств библиотечных знаний.

В настоящее время наблюдаются и процессы интеграции контента электронных библиотек, что ведет к появлению информационных пространств, основанных на федеративных принципах, а новым качеством, возникающим в результате этих изменений, является предоставление пользователям единых интерфейсов доступа к интегрированным разнородным данным на основе стандартизации

метаданных, а также унификация способов доступа к информации и способов ее извлечения для обработки и использования.

Для поддержки этого направления уже много лет ведутся исследования в области анализа семантики связей между информационными (в первую очередь, научными) материалами. Системным обобщением этих результатов стало появление комплекса онтологий SPAR [SPAR, 2012], обеспечивающих достаточно детальную категоризацию отношений, которые могут возникать между научными материалами в электронном виде, а также появление семантического раздела в модели научных данных CERIF [CERIF, 2010]. Объединение этих двух результатов – создание средств и сервисов информационных научных ресурсов, представляющих интегральный контент ЭБ, и разработка классификаторов отношений и семантических словарей, позволяющих выражать наличие определенных связей и отношений между объектами научного конкретного информационного ресурса, порождает важное новое качество – возможность разработки технологий семантического структурирования контента ЭБ. Серьезная работа в этом направлении ведется и в рамках консорциума W3C, где в проекте SKOS (Simple Knowledge Organization System) предлагается модель связывания научных данных, адаптированная для компьютерной обработки [SKOS, 2009]. В частности, SKOS включает контролируемые структурные словари семантических значений для связывания научных данных.

В дополнение к этому в различных научных дисциплинах (в первую очередь в биологии и медицине) были предприняты попытки разработать более подробную категоризацию отношений между научными текстами. Наиболее известными результатами этих попыток являются онтология SWAN (Semantic Web Applications in Neuromedicine), CiTO (Citation Typing Ontology), DoCo (Document Components Ontology) [Shotton, 2010] и др. В дальнейшем эти отдельные разработки были систематизированы, дополнены и объединены в единый комплекс под названием SPAR, включающий взаимосвязанную совокупность онтологий различного назначения. Независимо от этого, в рамках разработки концептуальной модели научных данных CERIF (Common European Research Information Format), ведутся работы по развитию стандартизированной формальной семантики для отображения отношений между объектами научных информационных систем (CRIS).

Форматы метаданных для описания электронных ресурсов, коллекций и авторитетных данных (таких как библиографические данные об электронных Интернет-ресурсах, записях о коллекциях архивных материалов и др.) разрабатываются в течение многих лет и в рамках проектов Dublin Core; Metadata Object Description Schema (MODS) и Metadata Encoding and Transmission Standard,

которые поддерживаются Библиотекой Конгресса США, Encoded Archival Description (EAD), Encoded Archival Context (EAC) и др. Форматы для описания логических структур электронных изданий и отображения непосредственно их содержания (полнотекстовые и полнотекстовые статьи, электронные книги) предлагаются в проекте Text Encoding Initiative (TEI) и других проектах.

Таким образом, в настоящее время уже заложена научно-техническая база технологического тренда «Семантические Электронные Библиотеки» и можно ожидать новых продуктов и услуг, разработанных на основе использования семантических технологий.

2.4.2. Проекты, продукты и сервисы тренда

Следует сразу отметить, что в настоящее время продуктов, которые бы четко позиционировались как СЭБ практически нет, а в области охвата данного технологического тренда наблюдается, в основном, активность проектной деятельности.

Среди проектов по семантическим электронным библиотекам явно выделяется проект СЭБ JeromeDL (eLibrary with Semantics) [JeromeDL, 2009] – совместный проект Main Library of Gdansk University of Technology [http://www.bg.pg.gda.pl/] и DERI International [http://www.deri.org/]. Основными требованиями (от библиотечных работников и «читателей»), которые должны быть реализованы в рамках данного проекта, являются:

- Поддержка легальности классических библиотек.
- Обеспечение навигации по библиотечным фондам, ориентированной на пользователей (user-oriented browsing).
- Поддержка эффективного поиска нужных библиотечных единиц.
- Обеспечение защиты от несанкционированного доступа.
- Поддержка множественных форматов ресурсов.
- Возможность коммуникации с другими электронными библиотеками.
- Использование последних результатов, полученных в области Semantic Web и в области коммуникации и управления информационными ресурсами.

По сути дела, в проекте JeromeDL последовательно решаются проблемы, которые характерны для эволюции электронных библиотек (Рис. 6).

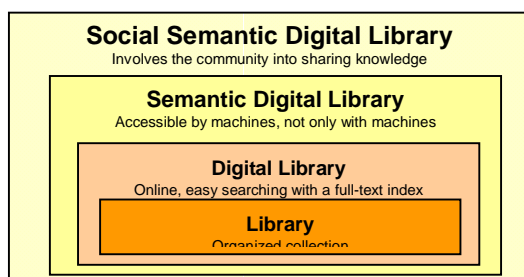


Рисунок 6 – Эволюция электронных библиотек

Проект Muruca – совместные усилия исследователей и индустрии библиотечного дела в области разработки набора Open Source приложений для создания семантических электронных библиотек и менеджирования их ресурсов [Muruca, 2012]. Инструментарий Muruca опирается на концепции Linked Open Data, Open Access content и технологии Semantic Web. Консорциум исполнителей по проекту Muruca включает Net7 – Internet Open Solutions – a european SME и Semedia Lab at Università Politecnica delle Marche, Ancona (Италия). Кроме того, проект поддерживается A-32 Open Scholarly Communities COST Action, где задействованы многие европейские центры исследований.

В проекте "Milan ICT Wired City - Open Digital Library of the city of Milan" [Occelli, 2006] выполняется при поддержке Lombard Interuniversity Consortium for Automatic Computation (CILEA) и Department of Research and Innovation of the City of Milan. В рамках проекта была создана первая семантическая электронная библиотека Италии.

Проект «A Semantic Digital Library of Urban Models and Resources» [COST, 2009] выполняется в университете Женевы (Université de Genève) в рамках инициативы COST. В проекте участвуют практически все европейские страны, а его целью является создание семантической электронной библиотеки для хранения урбанистических моделей и ресурсов. В данном проекте термин «семантика» подразумевает, что каждая модель будет семантически специфицирована связями с соответствующими концептами концептуального слоя из нескольких онтологий, ядром которого будет ontology of 3D city objects, построенная на основе CityGML. Предполагается, что библиотека будет «оборудована» Вэб-интерфейсами, с помощью которых пользователи будут импортировать модели и ресурсы, искать и сравнивать их, а также создавать и запоминать новые взаимосвязанные ресурсы.

Совместным проектом в области семантических электронных библиотек является проект SemLib [SemLib, 2012]. Консорциум исполнителей по этому проекту включает Net7 и Liberologico (Италия), IN2 (Великобритания), Knowledge Hives (Польша), National University of Ireland Galway (NUIG) - Digital Enterprise Research Institute (DERI, Ирландия), а также SeMedia group, A3Labs, Università Politecnica delle Marche, Ancona (Италия).

Мощным исследовательским проектом в области технологического тренда «Семантические Электронные Библиотеки» является проект Europeana [Europeana, 2012], поддерживаемый The Europeana Foundation, целью которого является кооперация музеев, архивов, коллекций аудио и видео материалов и библиотек таким образом, чтобы пользователи могли получить к ним доступ через сервисы Europeana.

Кроме рассмотренных выше проектов СЭБ можно отметить следующие более мелкие проекты в этой области: SWickyNotes (инструментарий для коллективного семантического аннотирования Вэб-страниц и, в частности, Digital Libraries. Первая версия инструментария (Philospace) реализована в рамках европейского проекта Discovery EU и ориентирована на изучающих философию); Talia (фреймворк на базе языка ruby для создания и кастомизации Semantic Digital Libraries); MPEG7AudioENC (активность по созданию в кооперации с Holger Crysandt (Aachen University) экстрактора метаданных MPEG-7, который бы покрывал практически полностью аудио спецификации) и др.

Таким образом, как показывает проведенный выше анализ, в области семантических электронных библиотек пока наблюдаются, в основном, проектная деятельность на уровне R&D.

2.4.3. Новые потребительские свойства и ожидаемые эффекты тренда

Как показывает анализ литературы и проектов по семантизации электронных библиотек, новые потребительские свойства, присущие технологическому тренду «Семантические Электронные Библиотеки», - следующие:

- Повышение эффективности использования многочисленных лицензионных баз данных в коллекциях библиотек, которое должно определяться
 - созданием средств получения статистики использования баз данных (Тратя значительные средства на платные базы данных, библиотеки столкнулись с проблемой оценки реального использования читателями как отдельных электронных изданий, так и баз данных в целом. Многочисленные вендоры баз данных или вообще не предоставляют такую статистику использования или предоставляют ее в разных форматах, что затрудняет ее анализ. Библиотекари настойчиво требуют включения пункта отчетности в формате COUNTER в лицензионные соглашения с владельцами баз данных);
 - обеспечением связи (linking) между различными представлениями одного и того же документа/статьи в разных базах данных. (Библиотеки, приобретая себе в коллекции тысячи электронных изданий и библиографические базы данных у разных фирм, сталкиваются с тем, что одна и та же статья может присутствовать одновременно в нескольких базах данных. При этом одна база данных, например, может содержать эту статью только в виде текста, другая в виде PDF, а третья—ее библиографическое описание);
 - одновременным поиском по различным базам данных (Multi-database или Federated

search). (Сегодня читатели вынуждены производить один и тот же поиск по интересующим их словам во все большем количестве баз данных. Последние отличаются друг от друга интерфейсами, навигацией по экранам и даже правилами запросов к системе. Федеративный поиск (Federated search) призван облегчить эту задачу путем осуществления поиска одновременно по многим базам данных и группировки/представления результатов поиска на одном экране для удобства читателей).

- Улучшение электронных каталогов (OPACs) и Вэб-сайтов библиотек, поскольку, например, американские библиотекари с беспокойством наблюдают процесс отчуждения читателей от библиотек и их электронных каталогов в пользу простых Интернет-поисковиков, которые индексируют Вэб сайты вне зависимости от их качества и авторитетности. Среди возможных способов повышения привлекательности и полезности библиотечных ресурсов и Вэб сайтов библиотеки видят, в частности, следующие:
 - Улучшение и повышение привлекательности для читателей электронных каталогов библиотек путем добавления (или линкования) дополнительной информации к традиционным кратким библиографическим описаниям изданий в формате MARC, интеграции с соответствующими метаданными, изображениями и полными текстами в других информационных системах, а также предоставления расширенных комплектов метаданных для интеграции с электронными библиографическими записями.
 - Создание и предоставление доступа к более обширным и авторитетным коллекциям электронных ресурсов путем сбора (harvesting) метаданных с использованием протокола Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) и перехода от устаревающих форматов представления записей к стандарту XML для описания коллекций.
 - Оптимизация электронных каталогов и Вэб-сайтов библиотек для облегчения их индексирования бесплатными поисковыми системами Интернета.
 - Внедрение круглосуточных справочно-библиографических услуг по электронной почте и через Вэб-сайты библиотек (electronic, online или red-eye reference service).
- Внедрение новых стандартов для электронных библиографических записей и документов, осуществления взаимосвязи между ними в базах данных и обмена информацией, что

должно обеспечить использование единого языка для обмена электронной информацией между библиотеками, архивами и другими организациями, внести единообразие в описания и структуры электронных ресурсов, способствовать их миграции с технологически устаревающих носителей и баз данных на новые, сохранению для потомков.

Интеграция метаданных отдельных ЭБ и новые средства семантического индексирования библиотечных ресурсов обеспечивает множество полезных возможностей как разработчикам, так и пользователям. При этом главным положительным моментом формирования информационного библиотечного пространства на основе контента отдельных ЭБ является предоставление пользователям единых интерфейсов доступа к интегрированным разнородным информационным ресурсам на основе стандартизации метаданных, а также унификация способов доступа к информации и способов ее извлечения для обработки и использования.

Среди важнейших ожидаемых эффектов технологического тренда СЭБ, в первую очередь, целесообразно выделить социальный (связан с существенным увеличением качества библиотечных ресурсов, представленных в электронных библиотеках, и, как следствие, повышением роли библиотек в образовательных процессах, с повышением эффективности доступа к библиотечным ресурсам и, как следствие, увеличением числа «читателей» в электронных библиотеках); экономический (связан с существенным увеличением количества «читателей» в семантических электронных библиотеках, и, как следствие, повышением отдачи от таких библиотек для государства и коммерческих организаций, поддерживающих библиотечные ресурсы) и политический (связан с формированием законодательной базы электронных библиотек и, как следствие, с повышением легитимности библиотечных ресурсов, представленных в семантических электронных библиотеках).

2.4.4. Оценка сроков внедрения приложений тренда

Работы в рамках технологического тренда «Социальный Семантический Вэб» ведутся, в основном, на стадиях фундаментальных и прикладных исследований и, редко, на стадии создания опытных образцов и запуска их в «опытную эксплуатацию». Учитывая это, и прогнозы по трендам «Открытые Связанные Данные» и «Семантический Социальный Вэб», представленные выше, реального появления на рынке первых важнейших продуктов и/или услуг в рамках технологического тренда СЭБ можно ожидать не ранее 2015 – 2017 г.г. При этом необходимо отметить, что первые семантические сервисы в существующих электронных библиотеках, которые, собственно, и сформируют сам технологический тренд «Семантические Электронные Библиотеки», появятся раньше. По

нашей оценке это произойдет в 2013 – 2015 г.г., а массового производства продуктов и услуг, предоставляемых семантическими электронными библиотеками «читателям», можно ожидать в 2016-2018 г.г.

Прогнозная кривая технологического тренда СЭБ, представленная на Рис. 7, показывает, что между бизнес-сообществом и потребителями нет согласованного мнения по поводу продуктов и сервисов, а сообщество R&D смотрит на научно-технические проблемы пессимистичнее, чем разработчики, поскольку считает, что переход к семантическим электронным библиотекам потребует серьезных исследований и разработок по семантической индексации информационных ресурсов существующих цифровых библиотек и практически значимых результатов в области семантизации контента и семантического поиска.

2.4.5. Технологии-конкуренты и технологии, от которых зависит развитие тренда

Технологиями-конкурентами к СЭБ, по нашему мнению, являются семантические социальные сети и открытые связанные данные. При этом надо отметить, что семантические электронные библиотеки, по существу, базируются на результатах указанных выше трендов и используют их для развития собственных продуктов и сервисов.

Преимущества конкурирующих направлений, по-видимому, связаны с их большей «популярностью» у молодого поколения населения, а недостатки определяются слишком широким охватом потребностей аудитории в случае социальных семантических сетей и ограничениями на открытые связанные данные, вытекающими из несовершенства законодательной базы в этой области.

По нашему мнению, развитие технологического тренда СЭБ опирается на результаты «родительского» технологического тренда ЭБ и существенно зависит от таких технологий из смежных областей, как семантические социальные сети, открытые связанные данные, онтологический инжиниринг, а также технологий семантизации контента и семантического поиска.

2.4.6. Разрушающий потенциал, барьеры и риски внедрения тренда

По нашему мнению, разрушающего потенциала у технологического тренда СЭБ нет, поскольку его продукты и сервисы являются естественным развитием существующих продуктов и сервисов технологического тренда ЭБ. Будучи введенными в массовое использование, продукты и услуги этого нового технологического тренда просто расширят рынок поставки информации потребителям, а технологический ландшафт при этом, по нашему мнению, практически не изменится. Не нарушится существенно и баланс сил в отраслевом сегменте, поскольку основные игроки сами активно продвигают данный технологический тренд.

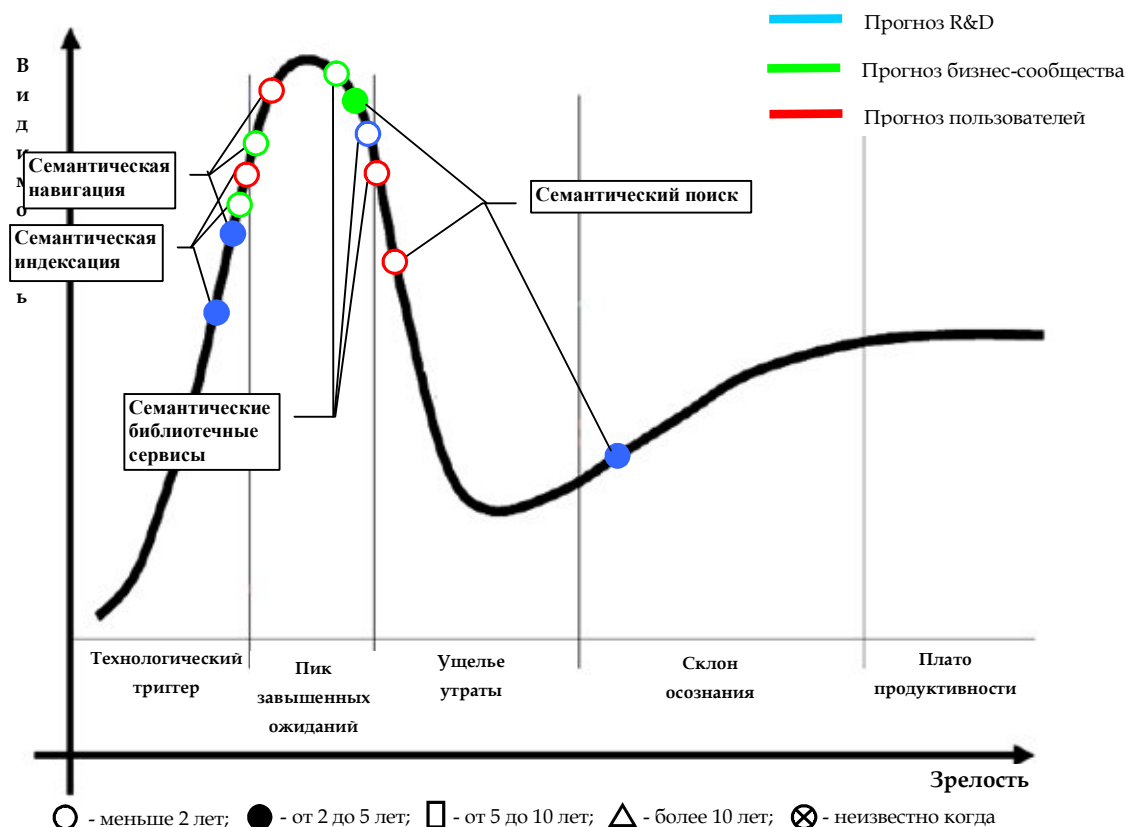


Рисунок 7 – Прогнозная кривая для технологического тренда «Семантические Электронные Библиотеки»

Барьеры, риски внедрения технологии и факторы неопределенности ее развития, по нашему мнению, связаны с недостаточностью законодательной базы для использования оцифрованных информационных ресурсов, уже представленных в электронных библиотечных коллекциях и циркулирующих в социальных сетях и в Интернет в целом, с научно-техническими проблемами семантизации контента и его семантической индексации, а также с проблемами создания действительно новых семантических сервисов.

2.4.7. Выводы по тренду

Как и в предыдущем случае, выше рассмотрены некоторые из основных аспектов технологического тренда «Семантические электронные библиотеки», дано описание тренда, кратко представлены проекты, продукты и сервисы, а также новые потребительские свойства и ожидаемые эффекты тренда и дана оценка сроков внедрения приложений. Идентифицированы технологии-конкуренты и технологии, от которых зависит развитие тренда, фиксирован разрушающий потенциал, барьеры и риски внедрения тренда.

ЗАКЛЮЧЕНИЕ

В работе представлены экспертные оценки по

трем из основных трендов в области использования семантических технологий в настоящее время – трендам открытых связанных данных, семантического социального Вэба и семантическим электронным библиотекам. При этом рассмотрены некоторые из основных аспектов этих новых технологических трендов, приведены их описания, кратко представлены проекты, продукты и сервисы трендов, а также их новые потребительские свойства и ожидаемые эффекты и дана оценка сроков внедрения приложений. Идентифицированы технологии-конкуренты и технологии, от которых зависит развитие тренда, фиксирован разрушающий потенциал, барьеры и риски внедрения тренда.

Таким образом, в настоящей работе сформирована методологическая основа для исследований и разработок в области автоматизации процессов выявления новых технологических трендов, которая, в свою очередь, должна, по нашему мнению, базироваться на семантических технологиях.

Работа выполнялась при поддержке Программы фундаментальных исследований НИУ ВШЭ в 2011 г.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- [Лапшин, 2010] Лапшин В.А. Онтологии в компьютерных системах. М.: Научный мир, 2010.
- [Хорошевский, 2008] Хорошевский, В.Ф. Пространства знаний в сети Интернет и Semantic Web (Часть 1) / В. Ф. Хорошевский // Искусственный интеллект и принятие решений. - 2008. - № 1.
- [Хорошевский, 2009] Хорошевский, В.Ф., Пространства знаний в сети Интернет и Semantic Web (Часть 2), / В. Ф. Хорошевский // Искусственный Интеллект и Принятие решений, 2009. - № 4.
- [Хорошевский, 2012] Хорошевский, В.Ф., Пространства знаний в сети Интернет и Semantic Web (Часть 3), / В. Ф. Хорошевский // Искусственный Интеллект и Принятие решений, 2012. - № 1 (в печати).
- [Benjamins et al., 2002] Benjamins, V. R., Contreras, J., Corcho, O., Gomez-Perez, A. Six Challenges for the Semantic Web, URL: http://www.cs.man.ac.uk/~ocorcho/documents/KRR2002WS_BenjaminsEtAl.pdf
- [Berners-Lee, 2003] Berners-Lee, T. The Semantic Web and Research Challenges, URL: <http://www.w3.org/2003/Talks/01-sweb-tbl/slide1-0.html>
- [Berners-Lee, 2006] Berners-Lee, T. Linked Data: 2006-07-27, URL: <http://www.w3.org/DesignIssues/LinkedData.html>
- [Bizer et al., 2009] Bizer, C., Heath, T., Berners-Lee, T. Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, (2009) Vol. 5(3), Pages 1-22.
- [CERIF, 2010] CERIF Data Models. URL: <http://www.eurocris.org/Index.php?page=CERIFreleases&t=1>
- [CommonTag, 2012] CommonTag home page. URL: <http://com-montag.org/Home>
- [COST, 2009] A Semantic Digital Library of Urban Models and Resources. Project Home Page. URL: <http://www.sbf.admin.ch/htm/dokumentation/publikationen/international/cost/cd2011/cost/C09.0062.html>
- [Europeana, 2012] Europeana Portal. URL: <http://www.europeana.eu/portal/>
- [GARTNER, 2012] Gartner home page, URL: <http://www.gartn-er.com/technology/research.jsp>
- [IDC, 2012] . IDC home page, URL: <http://www.idc.com>
- [JeromeDL, 2009] JeromeDL (eLibrary with Semantics). URL: <http://www.jeromedl.org/>
- [LATC, 2012] Linked Open Data Around the Clock project home page, URL: <http://latc-project.eu/>
- [LOD, 2012] Linking Open Data project home page. W3C SWEO. URL: <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
- [LOD2, 2012] Linking Open Data 2 project home page, URL: <http://lod2.eu/Welcome.html>
- [Mills, 2006] Mills, D. Semantic Wave 2006. Executive Guide to Billion Dollar Markets. A Project10X Special Report. January 2006.
- [MOAT, 2012] MOAT (Meaning Of A Tag) home page. URL: <http://moat-project.org>
- [Muruca, 2012] Muruca Home Page. URL: <http://www.muruca.org/>
- [Ocellli, 2006] Ocellli, S. Technological Convergence vs Knowledge Integration, Presented in: a Les Journées Annuelles Transdisciplinaires de réflexion au Moulin d'Andé, Colloques AFSCET, May 13-14, 2006.
- [OWL, 2009] OWL, W3C Documentation, 2009, URL: <http://www.w3.org/TR/owl-guide/>
- [PlanetData, 2012] PlanetData project home page, URL: <http://www.planet-data.eu/>
- [ReadWriteWeb, 2010] ReadWriteWeb Top 10 Semantic Web Products. URL: http://www.readwriteweb.com/archives/top_10_sema-ntic_web_products_of_2010.php
- [SCOT, 2012] SCOT (Social Semantic Cloud of Tags) home page. URL: <http://www.scot-project.org/>
- [SemLib, 2012] SemLib Project Home Page. URL: <http://www.s-emlibproject.eu/>
- [Shotton, 2010] Shotton, D. Introduction the Semantic Publishing and Referencing (SPAR) Ontologies. October 14, 2010. URL: <http://opencitations.wordpress.com/2010/10/14/introduci-ng-the-semantic-publishingand-referencing-spar-ontologies/>
- [SIOC, 2012] SIOC initiative home page. URL: <http://sioc-project.org/>

[SKOS, 2009] SKOS (Simple Knowledge Organization System). W3C Recommendation 18 August 2009. URL: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>

[SPAR, 2012] Semantic Publishing and Referencing Ontologies. URL: <http://sempublishing.svn.sourceforge.net/viewvc/sempublishing/SPAR/index.html>

[SPARQL, 2008] SPARQL Query Language for RDF, W3C Recommendation 15 January 2008. URL: <http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>

[Tofler, 2006] Tofler, A. Third Wave, Bantam Books, 2006, ISBN 0-553-24698-4

[Zemanta, 2012] Zemanta home page. URL: <http://www.zemant-a.com/about/>

SEMANTIC TECHNOLOGIES: EXPECTATIONS AND TRENDS

Khoroshevsky V.F.

*Federal State Budgetary Institution of Science,
Dorodnicyn Computing Centre of RAS,
Moscow, Russia*

khora@ccas.ru

State of art in the field of a new domain «Semantic technologies» is discussed in the paper. Main attention is given to the analysis of new technological trends in the field and an expert estimation of their importance and appearance of real applied products on the basis the usage of methods and tools of semantic processing of the information.