

Перспективы применения технологии семантического WEB при разработке интернет — порталов торговых площадок

А. Н. Скарина, email: anya.skarina@mail.ru

Государственный университет морского и речного флота имени адмирала С.О. Макарова

***Аннотация.** В данной статье описываются перспективы реализации концепции семантического Web при разработке интернет – порталов торговых площадок с учетом анализа современных поисковых систем и состояния современной сети Интернет, а также проблематика внедрения семантики на этапах создания интернет – порталов.*

***Ключевые слова:** Семантический Web, поисковые системы, структурированность данных, представление данных.*

Введение

Стремительное развитие веб – технологий в силу своей популярности и дешевизны порождает быстрый рост объемов информационного наполнения контента. Формат представления информации в сети в большинстве случаев ориентирован на людей и лишь в некоторых случаях допускает автоматическую обработку программными агентами, в результате чего возникает проблема выделения смысла информационных сообщений в автоматизированном режиме – поиск актуальной информации, необходимой пользователю, доставляет большие трудности, так как объемы web-пространства не позволяют оперативно обновлять базы данных информационно-поисковых систем. Поэтому дальнейшее развитие Интернет многие ученые связывают с концепцией семантического web, основной идеей которого является понимание информации компьютерами, выделение ими наиболее подходящих по тем или иным критериям данных и (уже после этого) предоставление информации пользователям [1]. Это означает, что автоматическая обработка информации в рамках семантического web предполагает в себе взаимодействие серверов друг с другом на основе анализа смысловых связей между объектами и понятиями, хранящимися в сети Интернет, которые должны отбирать лишь ту информацию, которая будет реально полезна пользователю.

1. Анализ состояния современной сети Интернет и принципы работы современных поисковых систем

По мнению разработчика концепции Тима Бернерса-Ли, семантический web реализует функциональное расширение существующей сети Интернет, в котором информация будет иметь конкретное смысловое значение, дающее возможность людям и компьютерам работать с достаточно высокой степенью взаимопонимания и согласованности [2].

Функциональное расширение предусматривает группировку разнообразной информации в единую систему, в которой каждому смысловому объекту данных будет соответствовать специальный синтаксический блок (тэг), который будет образовывать единую иерархическую структуру, на основе которой должен функционировать семантический web.

Для реализации семантического web разработчикам необходимо учитывать принципы работы современных поисковых систем и состояние современной сети Интернет относительно семантики:

- Во-первых, теги в HTML не несут семантической нагрузки;
- Во-вторых, процент полезной информации меньше процента разметки;
- В-третьих, разметка, в том числе, из-за большой сложности и вложенности (например, проблема табличной верстки) содержит много ошибок [3].

Так как информация в сети Интернет предназначена только для просмотра человеком, можно выделить следующие принципы работы поисковых систем:

- Во-первых, весовые коэффициенты на основе расположения слов;
- Во-вторых, важность слова в зависимости от тега;
- В-третьих, релевантность;
- В-четвертых, анализ "веса" ссылок в зависимости от количества ссылок, указывающих на данную страницу [3].

Таким образом, можно выявить ряд проблем, таких как непонимание, и следствием, невозможность анализа смысла информации машинами, сложность и неудобство поиска информации пользователем, так как результат часто является неудовлетворительным, и последнее, SEO – оптимизаторы, которые "играют" на несовершенстве алгоритмов поисковых систем, умышленно нарушая правильность разметки для кратковременного эффекта высоких позиций контента в поисковых запросах.

Разработчик концепции Тим Бернерс-Ли предполагает, что семантический web позволит решить вышеперечисленные проблемы.

При применении семантического web важной особенностью является то, как информация будет представлена в сети. В данном случае, функциональной системой должны учитываться определенные требования представления информации со следующими свойствами:

- Возможность обладать компактным синтаксисом;
- Возможность работать с большими базами данных;
- Наличие эффективного и мощного механизма вывода;
- Наличие четкой определенной семантики;
- Обладание достаточной описательной возможностью [3].

Все это позволяет расширить возможность поиска информации в сети Интернет, ведь он становится не просто средством обработки/отбора документов и оценки связи слов, например, стоящих друг от друга на определенном расстоянии, которым при формировании результатов запроса расставляются веса, влияющие на их порядок при выдаче.

2. Реализация концепции семантического web

В процессе реализации концепции семантического web получили широкое развитие синтаксические методы представления информации языковыми средствами семантических технологий. В настоящее время данные методы представлены слоями:

- Слой XML, который представляет структуру данных. XML обеспечивает синтаксис для структурированных документов, но не налагает на них никаких семантических ограничений на содержание этих документов;
- Слой RDF, который представляет смысловое описание данных;
- Слой Ontology, который служит для представления общепринятых соглашений о смысловом описании структур данных;
- Слой Logic, который позволяет выполнять интеллектуальный вывод;
- Слой Proof, который исполняет функцию поддержки взаимодействия между программными агентами на нужном доверительном уровне, обеспечивая понимание того, каким образом была извлечена необходимая информация.

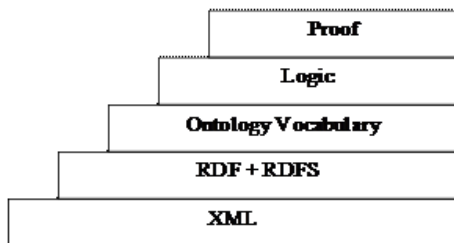


Рис. 1. Слои архитектуры семантического web

Семантический web дает пользователям возможность получить качественный результат поиска информации, а владельцам сайтов, соответственно, дает возможность получить больше целевого трафика, благодаря тому, что пользователь действительно находит полезную и нужную для него информацию.

Структурирование данных повышает ценность информации, благодаря постоянной структуре она может быть использована во многих отношениях.

Совокупность из вышеописанных технологий позволяет в структурированном виде представить данные и повысить уровень соответствия запросам пользователей.

При создании интернет-порталов торговых площадок начать работу следует с анализа поискового спроса, который позволит получить ответы на важные вопросы, а именно, как распределить товары по категориям, какие товары популярны, какие свойства товаров интересны пользователям и т.д.

Контент - менеджер заполняет интернет – портал определенным содержанием, и для того, чтобы грамотно разделить каталог на информационные блоки, необходимо использовать при верстке страницы микроформаты, которые позволят представить данные интернет – портала в четко сформированном виде для поисковых машин, что повысит вероятность правильной индексации контента.

При разработке интернет-порталов торговых площадок на основе технологии семантического WEB также не стоит забывать о создании семантического ядра, которое подразумевает под собой описание контента интернет – портала с помощью ключевых слов, которые определяются тематикой сайта – для этого необходимо всего лишь определить необходимую словоформу, изъять из семантического ядра ключевые слова, которые не дают трафика. Поэтому проработка структуры каталога состоит из трех этапов:

- Сбор и кластеризация семантического ядра;
- Определение групп запросов;
- Построение максимально полной, удобной и понятной структуры каталога.

Через семантику мы узнаем о потребностях целевой аудитории и, в дальнейшем, прорабатываем дополнительные разделы при необходимости.

3. Проблематика внедрения семантического web

Несмотря на видимые преимущества использования концепции семантического web, данная технология имеет много недостатков:

- В настоящее время существует огромное количество автоматизированных интернет-порталов, которые не имеют никакой семантики, однако у других программ, таких как агенты, нет никакого способа разыскать в сети подобную программу, выполняющую ту или иную функцию. Этот процесс, называемый обнаружением сервисов, станет возможным лишь после появления единого языка, позволяющего описывать сервисы, чтобы агенты могли «понимать», что позволяет делать данный сервис и каким образом им пользоваться. Так, например, поисковые системы умеют сравнивать как функционалы сайтов, допустим, наличие кнопки «купить» или «избранное», так и цены на представленные товары.

- Существует проблематика многоязычного контента, так как присутствует необходимость поддерживать доступ к информации независимо от того, на каком языке она была представлена первоначально. В основном, данная проблема возникает из-за того, что, во-первых, бренды в каталогах интернет-портала употребляются только на русском или английском языке, во-вторых, есть только один вариант названия услуги, объекта, товара;

- Необходимость описания метаданных приводит к дублированию информации. Создание и публикация содержимого будет занимать в два раза больше времени, так как необходимо будет представлять информацию не только для человека, но и для машин.

Заключение

Семантический web – это многогранная, динамичная и постоянно развивающаяся концепция, которая имеет возможность представлять информацию с сети Интернет в структурированные знания, с которыми смогут работать специальные интеллектуальные агенты и различные сервисы, взаимодействующие между собой и оперирующие понятиями соответствующих предметных областей. Цепочки логических рассуждений, осуществляемые агентами, позволят получать

информацию, представленную в Сети, в разрозненном виде, то есть из разных информационных источников.

Использование семантического ядра является основой дальнейшей оптимизации для интернет-порталов торговых площадок, так как это является преимуществом для логики поисковых запросов и правил работы поисковых систем. Это значит, что сайт будет удобен для потенциальных клиентов, потому что им будет проще найти именно нужный магазин. Семантика значительно упростит задачу дальнейшего продвижения интернет-портала торговой площадки.

Список литературы

1. W3C [Электронный ресурс]: Semantic Web. – Режим доступа: <https://www.w3.org/standards/semanticweb/>
2. Scientific American [Электронный ресурс]: The semantic Web. – Режим доступа: <https://www.scientificamerican.com/article/the-semantic-web/>
3. Электронный каталог НОУ ИНТУИТ [Электронный ресурс]: Семантический веб и микроформаты. – Режим доступа: <https://intuit.ru/studies/courses/611/467/lecture/28810>