

## **Анализ способов построения представлений web-документов для обработки семантическими поисковыми системами**

*О.Н. Половикова*  
*АлтГУ, г. Барнаул*

Современное направление развития Интернета заключается в переходе от документов, «читаемых компьютером» к документам, которые «понимаемы компьютером». Данная тенденция связана, прежде всего, с реализацией проекта – Semantic Web. Основная идея данного проекта заключается в переходе на новый уровень представления данных – уровень знаний и расширение возможностей их автоматизированной обработки. Мировое сообщество уже осознало необходимость принципиальных изменений Интернет-ресурсов и активно включилось в процесс реконструирования Web-пространства в пространство знаний.

Новые правила для представления документов в Web определяют и связывают их между собой таким образом, чтобы их данные можно было использовать для более эффективного исследования, автоматизации, интеграции и повторного использования в разных приложениях. Публикуемые данные могут быть общедоступными и обрабатываемыми автоматическими средствами так же, как и людьми [1]. Заинтересованные службы (модули, агенты) работающие в сети, могли получать из хранимого множества информационных ресурсов значимую для их функционирования информацию, т.е. извлекать знания.

Одной из основных задач концепции Semantic Web является решение проблем связанных с индексацией и поиском информации по словесному содержанию. Идея создания универсального средства семантического поиска информации на уровне нескольких информационных систем или для Web-пространства уже не является утопической и имеет реальное практическое воплощения, в качестве примера можно выделить семантическую поисковую систему SHOЕ.

В данном исследовании рассмотрены некоторые способы формализации информационных документов для их последующей адекватной обработки семантическими поисковыми системами. Под семантическими поисковыми системами будем понимать специализированные модули или совокупность связанных агентов (для распределенных систем), учитывающие смысловой контекст, в котором представлена обрабатываемая информация. Обычные поисковые системы: дескрипторные машины поиска или тематические каталоги, формируемые

специалистами в полуавтоматическом режиме, имеют принципиальные отличия от семантических поисковых систем.

Для машин поиска соответствие информационного ресурса пользовательскому запросу определяется результатом теоретико-множественной операцией (включение, пересечение) над множествами ключевых слов ресурса (его дескриптором) и слов запроса. Таким образом, обычные машины поиска не учитывают смысловое содержание рассматриваемых ресурсов. Сама форма составления пользовательского запроса (строка терминов с использованием специальных символов для обозначения булевых операций) позволяет выполнить только его лексический и/или синтаксический анализ, а значит невозможно сформулировать, а затем в процессе поиска учесть, реальные потребности пользователя.

Поиск с использованием какого-либо тематического каталога (общего или специализированного назначения) осуществляется посредством навигации по категориям разного уровня – последовательно выбираются тематические разделы и их подразделы, начиная с первого уровня, пока не будет предоставлен список гиперссылок на ресурсы соответствующей подкатегории. При таком подходе пользователю не нужно углубляться в специфику составления поискового запроса, тем более, что составление индекса для информационного ресурса выполняют эксперты (специалисты), следовательно, это описание, будет адекватно источнику. Но база данных ссылок (индексов) любого тематического каталога обычно имеет ограниченный объем, это связано с высокими трудозатратами процесса индексирования источников. И это даже не основной недостаток таких поисковых систем. Жесткая привязка к выбору конкретных категорий в процессе навигации по дереву тематических разделов, не позволяет пользователю оформить свои реальные потребности. Тем более что видение необходимых подкатегорий какой-либо категории каталога пользователями и экспертами может существенно отличаться. Поэтому пользователь может просто не найти нужную категорию, либо в выбранной категории будут отсутствовать удовлетворяющие потребностям пользователя ресурсы, по причине ограниченности объема БД ссылок каталога.

Необходимо заметить, что способы формализации информационных ресурсов выставляемых в глобальной сети, в объемном хранилище или в распределенной информационной системе взаимозависят от подходов к построению поисковых запросов и используемого критерия для определения соответствия ресурсов запросу. Невозможно построить семантический поиск ресурсов в системе без сочетания всех компонентов модели поиска: способа представления информационных

документов (или их заменителей), способа формирования запросов, критерия релевантности документов запросу.

С учетом современных требований к поиску web-ресурсов семантическая система должна интегрировать разнородные и распределенные знания в целостную картину предметной области. Для реализации такого поиска информационные ресурсы должны быть подготовлены для автоматизированной обработки их семантики – сформирован контент, язык запросов поисковой системы обязан оперировать с семантикой web-ресурсов, а поисковое приложение должно работать с распределенными знаниями (семантиками). Технологии Semantic Web обеспечивают взаимодействие всех составляющих модели семантического поиска.

В результате анализа информационных ресурсов сообщества Semantic Web были выделены следующие способы создания семантических представлений (формализации) информационных документов для их последующей обработки модулями семантической поисковой системой:

1. Аннотирование документа с использованием метаформатов (структура, стандарт для описания метаданных):

- метаданные встроены в сам ресурс;
- описания сохраняются и обновляются независимо от ресурсов.

2. Описание онтологии предметной области и семантическая разметка ресурса при помощи понятий семантической модели (онтологии).

На сегодняшний день разработано множество схем описания метаданных для информационного ресурса, также активно развиваются и языки для описания метаданных. Однако, базовыми стандартами для Semantic Web в данный момент признаются стандарты Dublin Core, FOAF, SIOC и DOAP. Чтобы встроить в документ его семантическое описание с использованием выбранной схемы построения метаданных не требуется внутреннего языка представления данных, для этих целей можно использовать язык XML специально разработанный формат описания ресурсов RDF. Метаданные, характеризующие смысловое содержание документа (контент) и предметной области (контекст), оформляются с использованием специализированных тегов – семантических тегов. Семантические теги стандартного HTML позволяют «внести знания» прямо в страницы [1].

Метаданные, оформленные использованием языка RDF, могут быть встроены непосредственно в ресурс (MsWord документ или HTML-страницу), а могут сохраняться и обновляться независимо от ресурсов. Многие из производителей программного обеспечения уже выпускают

ряд продуктов, которые автоматически формируют некоторый небольшой блок RDF-описания внутри документа. Второй подход является более универсальным, так как в этом случае метаданные могут быть созданы для любого ресурса [1]. В рамках проекта Semantic Web развитие получило направление по автоматическому созданию репозитория RDF-описаний ресурсов Интернет.

Для создания онтологий предметной области используют специально созданные для этих целей языки: Схема RDF (RDF Schema), OWL. Язык RDF Schema позволяет описывать структуру RDF-хранилища в терминах типов (классы, свойства) и отношений между ними, применяется для создания простых онтологий данных. Чтобы описать более сложные виды отношений следует использовать расширенный вариант RDF Schema – язык OWL, который позволяет описывать не только классы и свойства, но также задавать ограничения на их использование.

Модели предметной области, построенные на языке OWL, могут быть опубликованы в Web и одновременно использоваться модулями различных информационных систем, для того чтобы строить актуальные знания из ресурсов специфицированных данными онтологиями. Формирование новой онтологии предметной области может базироваться на имеющихся в сети онтологиях. Для концептуализаций содержания конечного ресурса можно использовать существующие в сети онтологии, которые по необходимости следует настроить под его специфику. Семантическая разметка документов на основе выбранной (или построенной) онтологии характеризует не только содержание этих ресурсов, а также семантику различных сервисов, предоставляющих эти документы конечным пользователям.

Следует выделить два перспективных направления в развитии проекта Semantic Web, связанных с использованием онтологий: создание визуальных сред для работы с онтологиями (построение, модификация, сопоставление и т.д.), разработка агентов автоматического построения семантических карт. Визуальные среды позволяют специалистам непосредственно «рисовать» онтологии, что помогает наглядно сформулировать и объяснить природу и структуру явлений. Семантические карты описывают концептуализацию содержания ресурса в виде OWL онтологий.

Как уже было отмечено выше, все составляющие модели поиска должны быть взаимозависимы. Поэтому при поступлении в систему пользовательского запроса для него также строится соответствующее представление, а метод его построения аналогичен методу построения представлений документов. Разметка документов с помощью мета-

форматов или онтологических терминов позволит производить автоматическую обработку их семантического содержания. Среди специальных язык запросов, которые умеют работать семантическим содержанием следует выделить SPARQL и RDF Query, которые базируются на обработке направленных графов (RDF-графов).

### **Библиографический список**

1. Андон Ф.И., Гришанова И.Ю., Резниченко В.А.. Semantic Web как новая модель информационного пространства Интернет [Режим доступа: <http://shcherbak.net/semantic-web-kak-novaya-model-informacionnogo-prostranstva-internet/>].
2. Басипов А.А., Демич О. В. Семантический поиск: проблемы и технологии // Вестник АГТУ. Сер.: Управление, вычислительная техника и информатика. 2012. №1. С 104-111.
3. Рабчевский Е.А. Булатова Г.И. Автоматическое построение онтологий для тематических поисковых систем [Режим доступа: <http://shcherbak.net/avtomaticheskoe-postroenie-ontologij-dlya-tematicheskix-poiskovyx-sistem/>].

## **Web-приложение в работе современной организации**

***Е.В. Прокопенко**  
КузГТУ, г. Кемерово*

С каждым годом web-приложения – вспомогательные программные средства, предназначенные для автоматизированного выполнения действий на web-серверах – приобретают все большую популярность из-за их универсальности, удобства использования и гибкости. За годы существования Интернета состав web-приложений, выполняемые ими функции, принципы и архитектура их построения претерпели значительные изменения – от простейших средств хранения HTML-страниц до решений, ориентированных на поддержку работы корпоративных информационных систем и их партнеров.

Web-системы имеют много преимуществ перед обычными системами, которые работают по технологии клиент-сервер. Достаточно разместить web-приложение на хостинге и можно работать с ним с любого компьютера, который имеет доступ к Интернету. С одной стороны это удобно, а с другой предъявляет дополнительные требования к надежности создаваемого программного обеспечения.

Главное преимущество web-приложений – это удобство в поддержке и администрировании: отсутствие необходимости установки приложения на каждое рабочее место, удобство при обновлении версий,