

СЕМАНТИЧЕСКАЯ ПАУТИНА ИЛИ БУДУЩЕЕ СЕТИ ИНТЕРНЕТ

А.О. Дементьев

ГОУ ВПО «Тамбовский государственный технический университет», г. Тамбов

Рецензент В.Е. Подольский

Ключевые слова и фразы: Интернет; информация; метаданные; программирование; семантическая паутина.

Аннотация: Рассмотрен вопрос построения семантической сети как способ решения проблем систематизации, поиска и доступа к информации в Интернете.

Самая громкая футурологическая утопия XX века – информационное общество. Родилась она в конце 1960-х – начале 1980-х годов. Ее теоретики Элвин Тоффлер, Даниел Белл, Маршал Маклюэн и другие обещали, что научное знание станет определяющим фактором развития современного общества, и промышленные корпорации будут вытеснены университетами – традиционными центрами создания и накопления знаний.

Однако, как показала практика, современный мир, несмотря на совпадение ряда предсказанных фоновых признаков, таких, например, как преобладание услуг и всеобщая компьютеризация, основывается не на знаниях, а на коммуникациях. Поэтому информационная теория плавно сменилась новой парадигмой «сетевого общества». Акцент здесь переносится с содержания на форму передачи и восприятия информации.

На заре возникновения Интернета в начале 1970-х годов, когда он всего лишь связывал компьютеры нескольких американских университетов, никто не предполагал, что эта паутина окутает всю Землю, станет массовой и привычной в использовании. Как известно, паутина основана на системе IP-адресов – уникальных числовых идентификаторов, которые присваиваются любому устройству, подключенному к Интернету, будь то компьютер, мобильный телефон или ПК, а также на системе доменных имен (**DNS**), которые пользователь, не затрудняя себя IP-номерах, вводит, когда хочет попасть на какой-либо сайт. Сегодня мы на пороге модернизации двух этих систем [1, 2].

Следующий шаг в будущее Интернета – разработка так называемой семантической паутины, или, как ее еще окрестили, Web 3.0. Ее идею в

Дементьев А.О. – аспирант кафедры «Криминалистика и информатизация правовой деятельности» ТамбГТУ, г. Тамбов.

2001 году опубликовал в журнале Scientific American изобретатель всемирной паутины Тим Бернерс-Ли.

Семантическая паутина – это часть глобальной концепции развития сети Интернет, целью которой является реализация возможности машинной обработки информации, доступной во Всемирной паутине. Основной акцент концепции делается на работе с метаданными (информация о данных или информация об информации), однозначно характеризующими свойства и содержание ресурсов Всемирной паутины, вместо используемого в настоящее время текстового анализа документов. В семантической паутине предполагается повсеместное использование, во-первых, универсальных идентификаторов ресурсов, или Uniform Resource Identifier (**URI**), а во-вторых – онтологий и языков описания метаданных [1].

Эта концепция была принята и продвигается Консорциумом W3. Для ее внедрения предполагается создание сети документов, содержащих метаданные о ресурсах Всемирной паутины и существующей параллельно с ними. Тогда как сами ресурсы предназначены для восприятия человеком, метаданные используются машинами (поисковыми роботами и другими интеллектуальными агентами) для проведения однозначных логических заключений о свойствах этих ресурсов.

Семантическая паутина – это надстройка над существующей Всемирной паутиной, которая призвана сделать размещенную в ней информацию более понятной для компьютеров. Машинная обработка возможна в семантической паутине благодаря двум ее важнейшим характеристикам.

1. Повсеместное использование универсальных идентификаторов ресурсов (**URI**). Традиционная схема использования таких идентификаторов в современном Интернете сводится к установке ссылок, ведущих на объект, им адресуемый. Очевидным свойством такой ссылки является возможность «загрузки» объекта, на который она указывает. Таким объектом может быть веб-страница, файл произвольного содержания, фрагмент веб-страницы, а также неявное указание на обращение к реально существующему физическому ресурсу по протоколу, отличному от HTTP (например, ссылки mailto:). Концепция семантической паутины расширяет это понятие, включая в него ресурсы, недоступные для скачивания. Адресуемыми с помощью URI ресурсами могут быть, например, отдельные люди, города и другие географические сущности, художественные артефакты и т. д. К идентификатору предъявляются несколько простых требований: он должен быть строкой определенного формата, уникальной, а также адресующей на реально существующий объект.

2. Повсеместное использование онтологий и языков описания метаданных. Современные методы автоматической обработки данных, доступных в Интернете, как правило, основаны на частотном и лексическом анализе текстового содержимого (хотя есть и исключения: Swoogle или Intellidimension Semantic Web Search Engine), которое прежде всего предназначено для восприятия человеком. В семантической паутине предлагается использовать форматы описания, доступные для машинной обработки (например, семейство форматов, часто упоминаемое в литературе как «Semantic Web Family»: RDF, RDF Schema или RDF-S, и OWL), в свою

очередь, использующие URI для адресации описываемых и описывающих объектов, а также онтологии и дескриптивные логики в качестве базовых математических формализмов.

Несмотря на все преимущества, предоставляемые семантической паутиной в случае ее внедрения, существуют сомнения в возможности ее полной реализации.

Разные комментаторы высказывают различные причины, которые могут быть препятствием к этому, начиная с человеческого фактора (люди склонны избегать работы по поддержке документов с метаданными, открытыми остаются проблемы истинности метаданных, и т. д.), и заканчивая косвенным указанием Аристотеля на отсутствие очевидного способа деления мира на концепты, что ставит под сомнение возможность существования онтологии верхнего уровня, критической для семантической паутины.

Необходимость описания метаданных так или иначе приводит к дублированию информации. Каждый документ должен быть создан в двух экземплярах: размеченным для чтения людьми, а также в машинно-ориентированном формате. Этот недостаток семантической паутины был главным толчком к созданию так называемых микроформатов.

Техническую часть семантической паутины составляет семейство стандартов на языки описания, включающее XML, XML Schema, RDF, RDF Schema, OWL, а также некоторые другие. Располагая их в порядке повышения уровня абстракции, реализуемого тем или иным языком, получаем:

- eXtensible Markup Language (**XML**) предоставляет синтаксис для определения структуры документа, подлежащего машинной обработке. Синтаксис XML не несет семантической нагрузки;

- XML Schema определяет ограничения на структуру XML-документа. Стандартный синтаксический анализатор языка XML в состоянии проверить произвольный XML-документ на соответствие его структуры так называемой *схеме документа*, описанной в XML Schema;

- Resource Description Framework (**RDF**) представляет собой простой способ описания экземплярных данных в формате *субъект-отношение-объект*, в котором в качестве любого элемента этой тройки используются только идентификаторы ресурсов. Существует стандартизованное отображение этих троек на XML-документы предопределенной структуры (т.е. консорциумом W3 определена *схема XML-документов*, содержащих RDF-описания), а также на другие форматы представления;

- RDF Schema описывает набор атрибутов (здесь их точнее назвать *отношениями*), таких, как `rdfs:Class`, для определения новых типов RDF-данных. Языком поддерживается также отношение наследования типов `rdfs:subClassOf`;

- Web Ontology Language (**OWL**) расширяет возможности по описанию новых типов (в частности, добавлением перечислений), а также позволяет описывать новые типы данных RDF Schema в терминах уже существующих (например, определять тип, являющийся пересечением или объединением двух существующих) [4].

Форматы описания метаданных в семантической паутине предполагают проведение логического вывода на этих метаданных, и разрабатывались с оглядкой на существующие математические формализмы в этой области. Математическое обоснование тех или иных конструкций языка описания необходимо для проведения заключений о свойствах программ, обрабатывающих данные в этом формате.

Особенно это относится к языку OWL. Базовым формализмом для него являются дескриптивные логики, а сам язык разбит на три вложенных подмножества (в порядке вложенности): OWL Lite, OWL DL и OWL Full.

OWL Lite предназначено для пользователей или приложений, которым необходима лишь классификационная иерархия сущностей и некоторые простые условия согласованности сущностей. OWL DL (Description Logic) рассчитано на пользователей, которым необходима максимальная степень выразительных возможностей языка без потери вычислительной полноты (без потери ни одного из семантических воплощений – содержательных толкований выводов, полученных формально-логическим путем) и разрешимости (вычисления будут закончены за конечное время). Уровень OWL DL ориентирован на существующие сегодня системы описания знаний и системы логического программирования. OWL Full рассчитано на пользователей, которым необходимы максимальные выразительные возможности языка и свобода выбора конечного формата RDT, но без каких-либо гарантий вычислительной полноты и разрешимости. OWL Full позволяет расширить смысл термина, взятого из какого-либо заданного словаря, и добавить его в онтологию.

Простая структура предикатов языка RDF, в свою очередь, позволяет использовать при его обработке опыт из теорий логических баз данных, логики предикатов, и т. д.

Одним из первых серьезных и популярных проектов, основанным на принципах семантической паутины, стал проект «Дублинское ядро», реализуемый инициативной организацией Dublin Core Metadata Initiative (DCMI). Это открытый проект, цель которого – разработать стандарты метаданных, которые были бы независимы от платформ и подходили бы для широкого спектра задач. Конкретнее, DCMI занимается разработкой словарей метаданных общего назначения, стандартизирующих описания ресурсов в формате RDF.

Для того чтобы семантическая сеть могла функционировать, компьютеры должны иметь доступ к структурированным хранилищам информации и множествам правил вывода, которые они могли бы использовать для проведения автоматических рассуждений. Исследователи в области искусственного интеллекта занимались изучением подобных систем задолго до возникновения Сети. Представление знаний, как часто называют эту технологию, в настоящее время в состоянии, сравнимом с тем, каким было понятие гипертекста до появления Сети: идея, несомненно, здравая, и уже существуют достаточно хорошие опытные образцы, но пока еще мир она не изменила. У нее уже есть зачатки важных приложений, но чтобы реализовать весь ее потенциал, необходимо связать ее в единую глобальную систему.

Привнести в Сеть логику (как то: способы применения правил вывода для проведения рассуждений, методы выбора тактик выполнения операций с данными и средства для ответов на вопросы) – вот та задача, которая стоит перед сообществом семантической Сети в настоящий момент. Комбинирование существующих математических и инженерных решений усложняет эту задачу. Эта логика должна быть, с одной стороны, достаточно сильной, чтобы позволять описывать сложные свойства объектов, а с другой – не настолько сильной, чтобы агента можно было поставить в тупик, дав ему парадоксальный запрос. К счастью, подавляющее большинство информации, которую мы хотим выразить, представляет собой нечто вроде «шестигранный болт является типом машинных болтов», что без труда вписывается в уже существующие языки, расширенные некоторыми дополнительными языковыми конструкциями.

Семантическая сеть – это не «просто» инструмент для решения чьих-то индивидуальных задач. Правильно организованная семантическая Сеть может, помимо прочего, способствовать эволюции человеческого знания в целом.

Человеческие усилия стеснены вечным противоречием между эффективностью малых независимо действующих групп людей и необходимостью согласования своих действий с широкой общественностью. Небольшой коллектив может быстро и эффективно создать нечто новое, однако зачастую это порождает своего рода субкультуру, концепции которой могут быть не поняты окружающими. С другой стороны, координирование действий большой группы чрезвычайно замедляет процессы и требует колоссального количества общения. Все в мире работает где-то посередине между двумя этими крайностями, с преобладанием тенденции начинать с малого – некой частной идеи – и постепенно продвигаться на пути к более широкому пониманию.

Существенным процессом становится объединение субкультур, когда требуется более широкий единый язык. Очень часто две независимо работающие группы разрабатывают очень сходные концепции, и попытка описать взаимосвязи между этими концепциями даст весьма существенную выгоду. Подобно финско-английскому языку или таблице преобразования мер и весов, эти взаимосвязи позволят общаться и сотрудничать, даже когда общность понятий не перешла еще в общность терминов.

Семантическая Сеть, именуя всякое понятие просто с помощью URI-идентификатора, даст возможность каждому выражать новые понятия, которые он изобретает, с минимальными усилиями. Ее универсальный логический язык позволит постепенно связать все эти понятия в универсальную Сеть. Эта структура сделает знания и достижения человечества доступными для анализа программными агентами и предложит нам новый класс средств, с помощью которых мы можем вместе жить, работать и учиться.

Список литературы

1. Dieter Fensel, Wolfgang Wahlster, Henry Lieberman, James Hendler, «Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential», MIT Press, ISBN 0-262-06232-1.
 2. John Davies, Dieter Fensel, Frank van Harmelen, «Towards the Semantic Web: Ontology-Driven Knowledge Management», John Wiley & Sons, ISBN 0-470-84867-7.
 3. Орлова, Д. Душа Интернета / Д. Орлова // Сынс. – 2008. – Август. – С. 24–27.
 4. Жишкевич, Н. Семантическая паутина [Электронный ресурс] / Н. Жишкевич. – Режим доступа : <http://shcherbak.net/2008/07>. – Загл. с экрана.
-

Semantic Web or the Future of the Internet

A.O. Dementyev

Tambov State Technical University, Tambov

Key words and phrases: Internet; semantic web; information; metadata; programming.

Abstract: The paper is devoted to the problem of a semantic network construction as the way of solving the problems of ordering, search and access to the information in the Internet.

© А.О. Дементьев, 2008