

УДК 004

JEL коды: D89

08.00.13

В.С. Магомадов

Чеченский государственный университет, г. Грозный, ул. А. Шерипова 32, ассистент кафедры «Бизнес-информатика» факультета информационных технологий

V.S. Magomadov

Chechen State University, Grozny, 32 Sheripova Street, teaching assistant in the Department of Business Informatics, Faculty of Information Technologies

Семантическая Паутина Semantic Web

Аннотация: предметом данной статьи является Семантическая паутина, которая также известна как Web 3.0. Статья нацелена на то, чтобы выявить одни из основных проблем, возникающих в процессе развития Семантической паутины. Рассматривается то, что уже сделано для того, чтобы решить эти проблемы, которые замедляют процесс реализации полного потенциала Семантической паутины. В дополнение к этому, в статье рассматриваются некоторые проблемы, имеющиеся у Всемирной паутины в ее нынешнем виде. Статья объясняет то, почему на сегодняшний день существует острая необходимость развития этой технологии. Предоставляется ответ на вопрос, является ли семантическая паутина полезной для людей, и какие именно проблемы, касающиеся нашей текущей Всемирной паутины, могут быть ею решены. Напоследок, приводятся примеры некоторых организаций, которые уже применяют принципы Семантической паутины в своей деятельности.

Ключевые слова: Семантическая паутина, RDF, OWL, Knowledge Graph, Open Graph, Всемирная паутина.

Abstract: The basis of this paper is to investigate into the Semantic Web. The paper covers some of the major issues and challenges which are related to the development of the Semantic Web. It also gives some insight into what has been achieved so far to overcome those issues and challenges. In addition, the paper discusses some issues with the current web, that is, it explains why it is required to develop the Semantic Web in the near future, how it can be useful and what type of web related issues can be solved with it. Finally, it gives some examples of who are using Semantic Web concepts in their industry at the moment.

Keywords: Sematic Web, RDF, OWL, Knowledge Graph, Open Graph, Web.

С самих ранних дней Всемирной паутины, исследователи осознали необходимость в понимании семантики информации, содержащейся в паутине, чтобы интеллектуальные системы были в состоянии более эффективным образом обрабатывать развивающуюся сеть документов. Одни из самых первых предложений включали в себя маркировку различных видов ссылок, чтобы отличать, например, страницы, описывающие людей от тех, которые описывают проекты, события и т.д. К концу 1990-х годов эти усилия привели к углубленному исследованию области компьютерных наук, которая стала известна как Семантическая

паутина. За последнее десятилетие, стремление исследователей к тому, чтобы у программных агентов была возможность общаться друг с другом в сети умным образом, привело к достижениям в разных областях: определение языков и стандартов для описания и запроса семантики ресурсов в сети; разработка приемлемых и эффективных способов взаимодействия с этими представлениями и их запроса; и определение онтологий, которые описывают Веб-данные, чтобы обеспечить более эффективную интероперабельность [1].

В недалеком прошлом, только программисты могли позволить себе разрабатывать Веб-страницы. Однако, на сегодняшний день, Всемирная паутина сильно отличается от того, что было в прошлом. Это означает, что теперь, при желании, кто угодно может разрабатывать Веб-страницы. Это стало возможным благодаря средствам разработки, которые предоставляются Вебом. Эти средства делают процесс разработки Веб-страниц намного проще, чем раньше. Некоторые компании даже предоставляют доступ к данным, закодированным на их Веб-страницах. Проблема с этими страницами состоит в том, что они предназначены только для чтения человеком, т.е. машины не могут читать эти страницы. В данном случае, компьютеры мы используем только для того, чтобы отображать информацию. Компьютеры не обладают навыками чтения и принятия решений. В отличие от человека, они также не в состоянии видеть связи. В результате всего этого, для выполнения самых важных задач, на сегодняшний день, невозможно обойтись без вмешательства человека.

Целью Семантической паутины является развитие навыков чтения и использования Веба компьютерами. Достигнуто это может быть посредством добавления метаданных в Веб-страницы, в результате чего машины будут в состоянии понимать контент Всемирной Паутины. Конечно же, тут не говорится о том, что компьютеры будут обладать самосознанием. Речь идет об инструментах, с помощью которых компьютеры будут более эффективными в поиске, обмене и интерпретации информации. Принципы Семантической паутины уже применяются некоторыми вебсайтами, но большая часть инструментов на данный момент находятся в процессе разработки.

Проблемы Всемирной паутины в ее настоящем виде

Семантическая паутина обещает быть значительно улучшенным вариантом существующей Всемирной паутины, имеющей несколько важных проблемы, которые нельзя игнорировать. Считается, что эти проблемы станут еще хуже при дальнейшем росте паутины. Ниже будет рассказано, с какими именно проблемами сталкивается Всемирная паутина на сегодняшний день:

1. Поиск информации. Проблема заключается в том, что поисковые системы, осуществляющие поиск, основываясь на ключевых словах, извлекают большой объем неуместной информации. Это означает, что те или иные ключевые слова используются в неправильном контексте. Не является секретом, что такие поисковики как Google или Yahoo выводят на экран тысячи страниц, не имеющих значение для пользователей, когда они вводят несколько ключевых слов для поиска. Пользователям, затем, приходится вручную перебирать эту информацию, чтобы извлечь именно ту информацию, которую они искали. Этот процесс порой занимает очень много времени. Эта проблема может быть устранена при помощи поисковиков, которые могут находить Веб-страницы с синтаксически разными, но семантически одинаковыми словами, используя онтологии.

2. Извлечение информации. На данный момент, для извлечения необходимой информации, невозможно обойтись без человека, который в состоянии осуществлять поиск и чтение информации. По причине того, что у автоматических агентов отсутствует здравый смысл, они не могут извлекать информацию из текстовых представлений.

3. Поддержка. Процесс поддержки слабо структурированных ресурсов является довольно-таки трудоемким и занимает много времени, тем более, когда их становится всё больше и больше. Чтобы эти ресурсы были последовательными, обновленными и правильными, существует необходимость в механическом представлении семантики и ограничений [2].

Достижения

Семантическая паутина является областью, привлекающей внимание большого количества исследователей. В связи с этим, в этой области есть определенный прогресс. Одним из знаменательных достижений является нормализация RDF (Resource Description Network, Среда описания ресурса) и OWL (Web Ontology Language, Язык Веб-онтологий). Академические исследования в этой сфере привели к развитию методологий для онтологической инженерии, отладки и модуляризации. В результате этого, появилось понимание того, что из себя представляют распространенные языки онтологий. В свою очередь, это понимание способствовало развитию более масштабируемых решений для логического вывода. Кроме того, были разработаны более эффективные инструменты моделирования. В свете всех этих достижений, не могли остаться в стороне и такие крупные компании, как Oracle и IBM. Они находятся в процессе разработки крупномасштабных хранилищ данных, которые будут поддерживать стандарты Семантической паутины. Вдобавок, существуют компании, специализирующиеся в этой сфере, например, Aduna, Altova, TopQuadrant и Ontoprise. Эти компании предоставляют средства, облегчающие применение семантических технологий в различных компаниях [3].

Проблемы

Семантическая паутина сталкивается с рядом проблем, которые должны быть решены для дальнейшего развития. Эти проблемы можно разделить на 3 категории: технические, научные и проблемы, связанные с бизнесом.

Главной технической проблемой Семантической паутины является его репрезентативная сложность. Здесь речь идет о языках OWL и RDF, являющихся основными языками Семантической паутины, в основе которых лежит XML. Данные языки являются довольно-таки сложными для изучения учеными и математиками, не говоря уже о людях, не имеющих какого-либо технического опыта. Разработчики главным образом были нацелены на завершенность и гибкость, что привело к созданию документов, являющихся сложными в плане анализа.

Другая не менее важная проблема Семантической паутины заключается в том, что она применяет классический метод «снизу-вверх», потому что Веб имеет огромный объем данных, которые подлежат трансформации. Когда речь идет о масштабах Всемирной паутины, аннотирование информации является задачей крайне трудоемкой. Тут возникает вопрос о разработке алгоритма, способного делать выводы на основе текстовой информации не хуже людей. Некоторые эксперты склоняются к такому мнению, что создание подобного алгоритма является невозможным.

Вышеупомянутые технические проблемы являются крайне сложными, но существуют еще и не менее сложные научные проблемы, связанные с Семантической паутиной. Согласно математической системе Курта Гёделя, не существует логической системы, являющейся и завершенной, и последовательной. Из этого следует, что существуют вещи, недоказуемые логикой, а также есть проблемы, не имеющие решения.

Проблема, связанная с бизнесом, также является одной из самых трудных для решения. Когда речь идет о бизнесе, возникают следующие вопросы:

- Как именно рекламировать Семантическую паутину?
- Какого рода бизнес можно создать на основе Семантической паутины? и т.д.

Чтобы тот или иной бизнес начал использовать стандарты Семантической паутины, необходимо провести эффективную маркетинговую кампанию, но это сделать крайне сложно при нынешнем состоянии Семантической паутины. Есть острая необходимость в формулировке пользы для конечного потребителя, чтобы возник резонанс среди людей. Это особенно важно при использовании вышеупомянутого метода «снизу-вверх». Если у пользователей появится понимание пользы Семантической паутины для их деятельности, то, возможно, это станет для них стимулом для аннотирования и трансформации их данных. Однако, при нынешней академической направленности Семантической паутины, сложно показать, в чем именно заключается польза для людей [4].

Использование Семантической паутины

Некоторые принципы Семантической паутины уже применяются, но предстоит разработать еще больше технологий, чтобы Семантическая паутина стала для нас реальностью. Несмотря на то, что эти технологии находятся в процессе разработки, никому неизвестно, когда мы начнем пользоваться полнофункциональной Семантической паутиной.

В числе инструментов, применяющих стандарты семантического Веба, есть Knowledge Graph (Граф знаний) – поисковый инструмент, разработанный компанией Google.

Цель данного инструмента заключается в том, чтобы помочь пользователю получать наиболее релевантные результаты поиска. По мнению Google, при помощи Knowledge Graph, пользователи теперь в состоянии получать самые уместные для них результаты поиска.

Knowledge Graph представляет из себя базу знаний, применяемой Google для того, чтобы улучшить результаты, выдаваемые ее поисковой системой посредством информации, извлекаемой из разных источников. Эта информация отображается в поле справа на компьютерах, или сверху на мобильных устройствах. Информация, включаемая в Knowledge Graph, значительно увеличилась с момента запуска. Первоначальный объем информации был утроен на протяжении первых семи месяцев, позволив дать ответ на около одну треть из 100 миллиардов ежемесячных запросов.

Семантическая паутина является сферой, требующей дальнейшего углубленного изучения, чтобы она могла быть внедрена во Всемирную паутину. Некоторые технологии в этом направлении уже разработаны, но они являются лишь маленькой частью того, что предстоит выполнить для реализации полного потенциала Семантической паутины. На данный момент существующие проблемы являются серьезным барьером, для преодоления которого необходимо инвестировать много средств и времени. Есть необходимость в переосмыслении нынешнего подхода к этой области.

Статья демонстрирует некоторые примеры организаций, придерживающихся стандартов Семантической паутины в определенной степени, но данный момент, не существует вебсайта или поисковика, который был бы полностью построен на основе принципов Семантической паутины. Однако, существующие примеры являются показателем того, что есть определенный прогресс.

По мере того, как ранние исследования перешли на эти более крупные, более прикладные системы, сегодняшнее исследование Семантической паутины меняется: оно основывается на более ранних основах, но оно сгенерировало более разнообразный набор направлений. Поскольку упомянутый ранее Граф знаний все чаще использует семантические представления, это привело к развитию функциональности приложений нового поколения (мобильные службы здравоохранения, картографические и торговые ассистенты и многие другие). Поскольку эти приложения становятся все более важными для рекламы и электронной коммерции, представления, которые они использовали, стали менее формальными и точными, чем предполагали многие ранние исследователи Семантической паутины [5].

Поскольку разработчики стремятся обеспечить структуру и организацию помимо простого связывания данных, они не часто используют формальную семантику, стандартизованную на семантических Веб-языках. Современные семантические подходы используют широко распределенный, гетерогенный сбор данных с помощью облегченной интеграции данных на основе потребностей. Эти подходы используют преимущества сосуществования множества различных, иногда противоречивых онтологий различного уровня детализации, не учитывая всеохватывающих или формально правильных онтологий. Кроме того, мы начинаем видеть более широкое использование текстовых данных, доступных в сети на сотнях различных языков, для обучения агентов с искусственным интеллектом, которые поймут, что пользователи пытаются сказать в данном контексте и какая информация наиболее уместна для целей пользователей в данный момент времени. Эти проекты все чаще используют семантическую разметку, доступную в Интернете [6].

Литература

1. М.С. Гарин, Е.В. Романенко. Интеллектуальный семантически ориентированный подход к автоматизации работы туристического агентства // Инженерный вестник Дона, 2012, № 2 URL: <http://ivdon.ru/ru/magazine/archive/n2y2012/811>
2. Р.Р. Файзрахманов. Новый подход к обеспечению веб-доступности для незрячих пользователей на основе улучшения навигационных характеристик веб-страниц // Инженерный вестник Дона, 2012, № 2 URL: <http://ivdon.ru/ru/magazine/archive/n4p2y2012/1442>
3. Антониу, Г., Грос, П., Хармелен, Ф., Хоекстра, Р. (2016) *Семантический Веб*. Москва: ДМК Пресс.
4. Амерланд, Д. (2014) *Семантический поиск Google*. Нью-Йорк: Pearson Education.
5. Маннинг, К., Рагхаван, П., Шютце, Х. (2014) *Введение в информационный поиск*. Москва: Вильямс.
6. Салтыков, С. (2015) *Семантический Веб*, <http://seotuition.ru/Поисковоепродвижение/ОбучениеSEO/СемантическийВеб.aspx>

References

1. M.S. Garin, E.V. Romanenko. *Intellectual semantically oriented approach to the automation of the work of the travel agency* // *Engineering Bulletin of the Don*, 2012, No. 2 URL: <http://ivdon.ru/en/magazine/archive/n2y2012/811>
2. R.R. Fayzrakhmanov. *A New Approach to Web Accessibility for Blind Users Based on Improving the Navigation Characteristics of Web Pages* // *Engineering Bulletin of the Don*, 2012, No. 2 URL: <http://ivdon.ru/en/magazine/archive/n4p2y2012/1442>
3. Antonio, G., Gros, P., Harmelin, F., Hoekstra, R. (2016) *Semantic Web*. Moscow: DMK Press.

4. Amerland, D. (2014) *The semantic search of Google*. New York: Pearson Education.
5. Manning, K., Raghavan, P., Schütze, H. (2014) *Introduction to Information Search*. Moscow: Williams.
6. Saltykov, S. (2015) *Semantic Web*, [http://seotuition.ru/Search promotion / Training SEO / Semantic Web.aspx](http://seotuition.ru/Search%20promotion/Training%20SEO/Semantic%20Web.aspx)