

## **Семантическая паутина и поисковая оптимизация: особенности взаимодействия**

*Верхова Ольга Геннадьевна*

*Брянский государственный университет имени академика Ивана Георгиевича Петровского*

*Магистрант*

*Иванова Наталья Александровна*

*Брянский государственный университет имени академика Ивана Георгиевича Петровского*

*Кандидат технических наук, доцент кафедры информатики и прикладной математики*

### **Аннотация**

Данная статья посвящена анализу взаимодействия поисковых систем, семантической паутины и поисковой оптимизации. Проведенное исследование раскрывает возможности семантической паутины по поисковой оптимизации веб-ресурсов.

**Ключевые слова:** Семантическая паутина, Семантический Веб, поисковая оптимизация, сниппеты.

### **The semantic web and search engine optimization: features of interaction.**

*Verkhova Olga Gennadyevna*

*Bryansk State University named after Ivan G. Petrovsky*

*Master*

*Ivanova Natalia Aleksandrovna*

*Bryansk State University named after Ivan G. Petrovsky*

*Ph.D., assistant professor of computer science and applied mathematics*

### **Abstract**

This article analyzes the interaction between search engines, semantic web and search engine optimization. The study reveals the possibility of the semantic web to search engine optimization web resources.

**Keywords:** SemanticWeb, SEO, snippet

Глобальная сеть Интернет постоянно развивается. На сегодняшний день ориентация ее развития направлена на создание высококачественных сервисов и их содержимого, и данная концепция получила название Web 3.0.

Семантическая паутина (или семантический веб) рассматривается как одна из фундаментальных составляющих Web 3.0, и представляет собой

совокупность связанных данных. Основная идея семантической паутины заключается в добавлении семантической разметки, которая не только описывает данные, но и обеспечивает возможность создания связей между разрозненной информацией, хранящейся на разных веб-ресурсах.

Таким образом, семантический веб служит для улучшения качества результатов поисковых запросов, а также позволяет отвечать на пользовательские вопросы без переходов на веб-ресурсы. Кроме того, использование семантической паутины обеспечивает более формальное решение проблемы канонических URL и, в долгосрочной перспективе, может решить проблему повторяющегося контента [1].

Также семантические веб-технологии позволяют владельцам веб-ресурсов более точно решить проблему неоднозначностей для поисковых систем. Например, столкнувшись с двумя ресурсами по запросу «лук», доступными только в формате HTML, поисковая система должна обработать данные, чтобы по возможности выяснить, о чем именно шла речь. Семантическая паутина позволяет не только точно указать поисковой системе, что речь идет про «лук – растение», а не «лук – оружие», но и связать между собой понятия «лук-растение» и «чеснок-растение», таким образом связывая и веб-ресурсы.

Алгоритмы поисковых систем считают релевантными ресурсы с более детальным описанием веб-страниц, так что использование семантически связанных упорядоченных данных повышает видимость веб-ресурса в поисковом запросе.

В 2008 году Google начал собирать рейтинг сниппетов – кратких описаний веб-страниц сайтов, релевантных запросам – для Yelp, Citysearch, CNET, TripAdvisor и Download.com. На его основе Google провел детальный анализ веб-ресурсов с целью структурировать данные в удобную для обработки форму. Такого рода структурирование данных, безусловно, уже существовало с начала использования алгоритма Hilltop, приобретенного Google в 2003 году. Hilltop описывает механизм для извлечения ключевых фраз на основе HTML, и определяет их отношения со ссылками на странице (по принципу «ключевая фраза -> определение-> URL»).

Очевидно, что поисковым системам «легче» обрабатывать структурированную информацию. Постоянно растет число поддерживаемых поисковыми системами форматов и осуществляется поддержка по их использованию веб-разработчиками. Так в 2009 году Google объявил о поддержке RDF и микроформатов и продолжает расширять этот список в том числе и онтологиями для электронной коммерции GoodRelations [2]. Яндекс также поддерживает ряд микроформатов и семантическую разметку [3].

На самом деле, несмотря на относительно медленное внедрение технологий Web 3.0, структурированные данные уже везде, и даже если не используется RDF; то поисковым системам передаются структурированные данные в формате XML. Например, информация, полученная из блогов, может отображаться в Google RSS. Структурированные сведения, собранные

в Интернет-магазинах, также могут отображаться в Google Покупки и Яндекс.Маркет.

Поисковые системы определили два элемента для создания сниппетов:

1. Структурированные данные (в виде разметки товаров микроформатами hProduct, GoodRelations или Schema.org);
2. Загрузка информации о ресурсах в Google Покупки или Яндекс.Маркет [4].

Так что использование структурированных данных почти наверняка поможет улучшить видимость сайта в поисковых системах.

Кроме того, использование семантической разметки позволяет упорядочить внутреннюю архитектуру сайта.

Веб-ресурсы, созданные с использованием RDF на основе SPARQL-запросов и OWL-онтологий, получают хорошо структурированными и удобными для пользователей. Использование онтологий решает проблему каннибализации, так что поисковые системы отдадут предпочтение именно семантическим веб-ресурсам [5].

При выдаче поискового запроса перед поисковой системой стоят две задачи – фильтрация результатов и их упорядочивание. Так как внедрение в поисковые системы структурированных данных в конечном счете приводит к повышению видимости в поисковых системах, может возникнуть желание создавать спам, основанный на технологиях семантического веб.

Но с другой стороны, для получения данных фильтрация спама становится все менее важной. То есть, данные, которые пользователь запросит или просмотрит, будут содержать относительно не так много спама. Кроме того, с учетом персонализации результатов поисковых запросов, чем больше пользователь использует определенный тип информации на конкретных сайтах, тем больше вероятность того, что поисковые системы покажут пользователю данные именно с этих сайтов [6].

Помимо фильтрации структурированных данных от спама, поисковые системы упорядочивают эти данные. И как бы ни высокопоисковые системы оценивали структурированные данные, они будут ранжировать результаты согласно своим алгоритмам.

Поисковые системы не ранжируют данные, основываясь исключительно на семантической значимости ресурса, в отличие от запросов SPARQL из DBpedia. Релевантность запроса определяется относительно пользователей. Таким образом, структурированные данные никогда не будут преобладать над полезными данными, но релевантный ресурс вполне может быть семантическим. Актуальность и достоверность информации с точки зрения семантической паутины определяется источником данных и цифровыми подписями, в то время как со стороны поисковой оптимизации – цитированием и пользовательскими оценками ресурса. В действительности, в отличие от семантической паутины, поисковые системы напрямую определяют важность веб-ресурсов их актуальностью.

Традиционно, поисковые системы использовались для поиска информации на веб-ресурсах по запросам. Однако все чаще различные типы данных предоставляются пользователям из разных источников, что позволяет

находить новую информацию не только поисковыми запросами, например, из RSS, ленты новостей Facebook или сообщений в Twitter. Такого рода данные все чаще отображаются на страницах результата поиска.

В поисковых системах существуют так называемые косвенные связи, и на их основе посредством машинной обработки определяются возможные соединения связанных данных. На основе профилей пользователей в социальных сетях поисковые запросы анализируются, и информация, полученная в их результате, соотносится непосредственно с пользовательскими предпочтениями.

Web3.0 влечет за собой переход от ориентирования поисковой оптимизации на рекламу к использованию связанных данных и семантической паутины (рис. 1). Традиционная цель поисковой оптимизации заключается в привлечении пользователей на сайт. Но если процесс обработки связанных данных для поисковых систем упростится, то снизится и число пользователей, просматривающих веб-ресурсы.

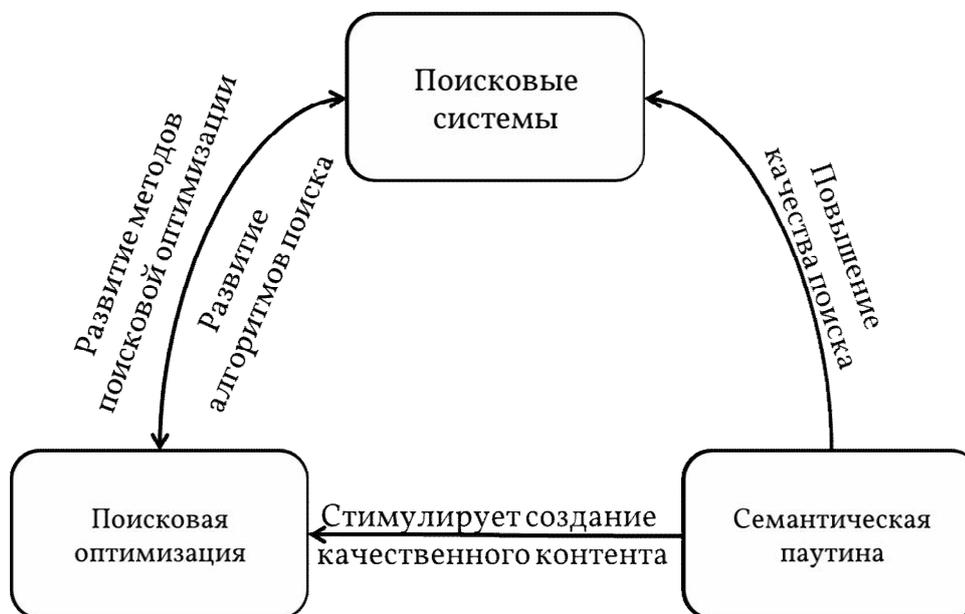


Рисунок 1 – Взаимодействие между поисковыми системами, поисковой оптимизацией и семантической паутиной

Предоставление поисковыми системами ответов на вопросы непосредственно в результатах поиска повлекло за собой неприятные для поисковой оптимизации последствия. В частности, пользователи перестали непосредственно переходить на веб-ресурсы для получения ответов на свои вопросы.

Но подобные изменения – результат эволюции сети от традиционного поиска к информационному развертыванию. Таким образом, создание семантически связанных ресурсов с полезной пользователям информацией – закономерный процесс развития Интернета, способствующий развитию и поисковой оптимизации в том числе.

**Библиографический список**

1. Understanding the Semantic web: Pros and Cons [Электронный ресурс] – URL: <http://www.bypeople.com/semantic-web-pros-and-cons/> (Дата обращения 28.05.2016).
2. GoodRelations [Электронный ресурс] – URL: <https://www.w3.org/wiki/GoodRelations> (Дата обращения 26.05.2016).
3. Вопросы о семантической разметке [Электронный ресурс] – URL: <https://yandex.ru/support/webmaster/schema-org/semantic-faq.xml> (Дата обращения 24.05.2016).
4. hProducts [Электронный ресурс] – URL: <http://microformats.org/wiki/hproduct> (Дата обращения 28.05.2016).
5. SEO, the Semantic Web and Information Discovery [Электронный ресурс] – URL: <http://www.seoskeptic.com/seo-semantic-web/> (Дата обращения 25.05.2016).
6. Семантический веб и микроформаты [Электронный ресурс] – URL: <http://www.intuit.ru/studies/courses/611/467/lecture/28810%253Fpage%253D3> (Дата обращения 27.05.2016).