

УДК 004.738.52

Проблема и технология семантического поиска

Елемесова Салтанат Ержигитовна

магистрант

Международный университет информационных технологий
yelemessova.saltanat@gmail.com

Андатпа. Бұл мақалада семантикалық іздеу жүйесінің проблемалары және жұмыс істеу алгоритмі талқыланады. Сондай-ақ, семантикалық іздеу жүйесінің алгоритмінің мүмкіндіктері сипатталады. Семантикалық іздеу жүйесін жүзеге асыру проблемалары анықталды.

Кілт сөздер: *семантикалық іздеу, интернет, іздеу жүйесі, онтология.*

Аннотация. В статье обсуждается проблемы и возможности алгоритма работы семантического поиска. Также в статье описаны отличительные особенности алгоритма семантического поиска. Выявлены проблемы реализации семантических поисковых систем.

Ключевые слова: *семантический поиск, интернет, поисковая система, онтология.*

Abstract. Challenges and opportunities of working semantic search are considered in the paper. The article also describes the features of semantic search algorithm. The existing realization problems of semantic search engines are revealed.

Key Words: *semantic search, Internet, search engine, ontology.*

Введение

Большие объемы прикладных задач позволяет рассматривать Интернет в качестве источника самых разнообразных знаний. Но такие факторы как повторяемость и постоянный рост количество ресурсов в Интернете усложняют задачу. Поиск необходимой и релевантной информации зачастую требует значительных временных затрат. Постоянное изменение и увеличение объемов ресурсов требует систематизации и структурирования. Для эффективного поиска необходимо выполнение нескольких требований. Важнейшими из них являются скорость обработки поискового запроса и возможность грамотного ранжирования информационных ресурсов. Целью настоящего исследования является рассмотрение существующих поисковых систем, рассмотрение технологий семантического поиска [1] и определение специфических проблем, касающихся

поиска документов в семантической паутине с использованием запросов на естественном языке.

Общий алгоритм построения системы семантического поиска

В данном разделе описывается алгоритм работы семантического поиска. Но прежде всего, нужно отметить, что семантический поиск-это процесс поиска документов по их содержанию, т.е. релевантность документа определяется семантически, а не синтаксически. Объединив некоторое количество технологий, семантический поиск разрешает сотворить мета-движок для фильтрации SWD-документов, отдаваемых поисковым приспособлением Swoogle[2]. Swoogle-это поисковая машина по онтологиям, которая обладает собственным языком составления сложных запросов [3]. Извлекая из них метаданные, Swoogle подсчитывает зависимости между документами. Несмотря на то, что она в последнее время предназначается в качестве SWD-индексирующей системы, используемая поисковая технология основана на поиске лексического соответствия терминов запроса и проиндексированных названий онтологических классов и свойств. Целью использования этой технологии является подтверждение того, что применяя предложенный метод семантического поиска, точность поиска для простого запроса может быть существенно улучшена. На таблице ниже сравниваются современные поисковые системы с семантической поисковой системой.

Таблица 1: Сравнение поисковых систем

	Google Index YANOO!	Семантическая поисковая система
Задача поиска	Найти ресурсы, содержащие слова запроса	Найти информацию - ответ на запрос
Кто определяет наличие ответа	Пользователь анализирует весь найденный поисковиками информационный "мусор" в поисках смыслового ответа	Поисковая система автоматически выбирает смысловые ответы на вопросы пользователя
Запрос пользователя	Набор ключевых слов. Игнорируются вопросительные слова, предлоги, союзы и др.	Нормальное предложение. Учитываются все лингвистически важные слова
Навигация по найденным ссылкам	Нужна. Пользователь должен искать информацию на указанных ресурсах	Не требуется. Ответы в выдаче поисковой системы

Алгоритм семантического поиска позволяет автоматически преобразовывать запросы на естественном языке в онтологии запроса, употребляемые при онтологическом поиске/ранжировании SWD-документов. На рисунке 1 изображена

схема семантического поиска. Зависимые онтологии присоединяют в метод устранения неоднозначности посредством словаря WordNet. WordNet-это электронный тезаурус или семантическая сеть, который использует механизм автоматического отображения на смысловые значения для разрешения неоднозначности терминов запроса.

AUTOMS – это автоматический инструмент, который использует алгоритм для создания онтологий запроса. Соответствие между SWD-документом и переформулированным запросом определяется с помощью этого инструмента AUTOMS. AUTOMS комбинирует лексические, семантические и структурные поисковые методы. Лексический поисковый метод, используя методы для измерения синтаксической схожести, вычисляет соответствие имен онтологических концептов. Структурный поисковый метод определяет соответствие понятия онтологии с учетом терминов в его окрестности. Окрестность концепта включает те понятия, которые зависят от него. И в конце, метод семантического отображения имеет дело с сопоставлением между смысловыми значениями определений понятий. Вычисление семантического отображения может зависеть от внешней информации, найденной в словарях, тезаурусах или зависимых онтологиях.

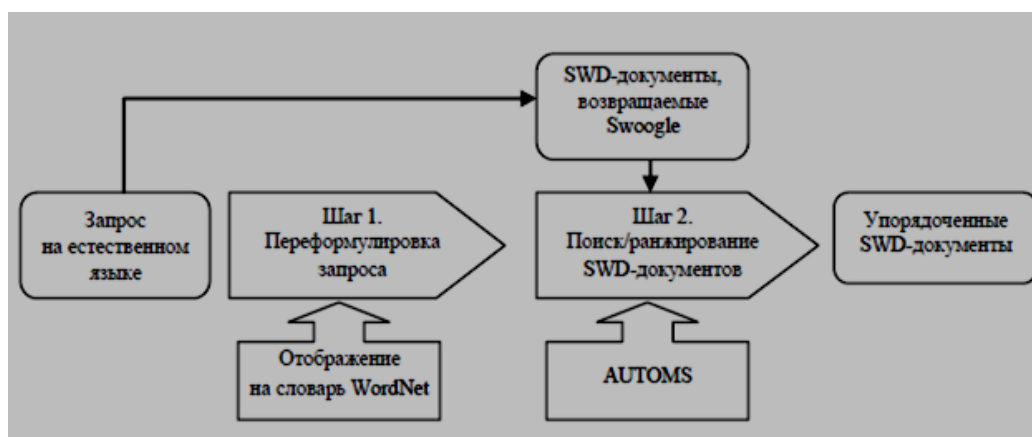


Рисунок 1: Схема алгоритма системы семантического поиска

Основываясь на том, насколько хорошо полученные документы соответствуют онтологии запроса, алгоритм позволяет проводить ранжирование полученных SWD документов. Устанавливается это числом соответствий между онтологией запроса и SWD-документом. Позиция SWD-документа в финальной выдаче будет высокой, если соответствие между SWD-документом и онтологией запроса соответственно будет большой. Фактически набор SWD-документов, который участвует в ранжирующем алгоритме, является списком документов,

полученных путем передачи запроса в свободной форме в Swoogle. Таким образом, ранжирование может быть представлено как фильтрующая обработка SWD-документов, возвращаемых ключевым поиском (рисунок 1).

Несмотря на то, что семантическая поисковая система способствует поиску информации в сети, есть несколько серьезных недостатков. Рассмотрим, какие недостатки имеет семантическая поисковая система на данный момент:

1) Огромное количество неструктурированных документов. Для использования семантическими поисковыми системами эти документы должны быть семантически размечены. А это тоже непростая задача, так как это требует развития проблемно-ориентированных онтологий.

2) Алгоритм нуждается в автоматизации процесса преобразования запросов свободной формы в формальный вид. Методы, позволяющие автоматизировать этот процесс, в настоящее время являются объектом исследования. Построение отображения онтологий предметных областей на формальные запросы также активно исследуется.

Заключение

В статье был рассмотрен алгоритм семантического поиска и его проблемы в настоящее время. Предоставлены проблемы, которые возникают при реализации семантической поисковой системы. Семантический поиск нуждается в автоматизации процесса преобразования запросов свободной формы в формальный вид. Это считается большой трудностью на данный момент. Представленный алгоритм пока не способен полноценно обеспечивать поддержку запросов свободной формы.

Использованная литература

- [1] Guha R. Semantic search / R. Guha, R. McCool, E. Miller // Proceedings of the 12th international conference on World Wide Web.–N.Y.ACM Press, 2003.– P. 700–709.
- [2] Кутовенко А. Профессиональный поиск в интернете. – СПб.: Питер, 2011. – 201 с.
- [3] Bernstein A., Kaufmann E., Fuchs N. Talking to the semantic web – a controlled english query interface for ontologies // AIS SIGSEMIS Bulletin. – 2005. – N 2. – P. 42–47.