

Оценка качества информационного поиска в слабо структурированных источниках на основе метаданных и базы знаний.

Автор: А.А. Егошина - канд. техн. наук, доцент, А.С. Вороной - ассистент
Донецкий Национальный Технический Университет
Кафедра систем искусственного интеллекта

Источник: Научные труды ДонНТУ выпуск 16 (204) Информатика, кибернетика 2012 и вычислительная техника

Аннотация:

Предлагается технология повышения эффективности поиска в слабо структурированных базах данных с Web-интерфейсом на основе метаданных, представляющих структуру баз данных и базы знаний, описывающей семантику хранимых данных. Проведена оценка качества информационного поиска в слабо структурированных источниках на основе метаданных и базы знаний для различных категорий пользователей с использованием критериев полноты и точности. Результаты экспериментов показали, что при реальных значениях числа пользовательских запросов к базам данных предлагаемая технология обеспечивает комфортное для пользователей время обработки запросов, сокращение нагрузки на серверы системы и приемлемые для всех категорий пользователей полноту и точность. С ростом числа запросов преимущества предложенной технологии возрастают.

Общая постановка проблемы

Основной проблемой при работе с распределенными и слабо структурированными информационными ресурсами является сложность точной формулировки запроса - подбора ключевых слов, которые предстоит искать в документах или базах данных.

В последнее время активное развитие получило направление в информационных технологиях, использующее стандарт метаданных, который позволяет пользователям совершать поиск в большом количестве таблиц баз данных и уверенно определять местонахождение интересующей информации. Метаданные определяют ортогональный основному уровню описания информации (который формируется такими понятиями, как классы, типы данных и др.) уровень описания свойств [1]. Использование метаданных, в особенности семантических, позволяет эффективно решать такие задачи работы со знаниями как поиск, категоризация и рекомендация знаний.

Организация базы знаний для семантического поиска на основе онтологий в web-ориентированных реляционных базах данных

Для повышения эффективности информационного поиска в web-ориентированных реляционных базах данных в работе [2] предлагается подход к извлечению информации из слабо структурированных источников на основе метаданных и базы знаний.

Использование базы знаний, как некоторого унифицированного интерфейса для решения задач над множественными неструктурированными источниками информации, освобождает пользователя от необходимости находить релевантные источники, задавать запросы к каждому из них по отдельности и вручную сопоставлять информацию из них.

В связи со значительным объемом ресурсов (таблиц базы данных) и их слабой структурированности в работе [3] предлагается хранить метаданные отдельно от ресурса в хранилище метаданных на отдельном сервере. Для организации базы знаний используется онтологический подход, который позволяет отразить семантику ресурса.

Первоначально база знаний содержала только онтологии, на основе которых для пользователей формировались ответы в виде html-фрагментов. Однако, в процессе эксплуатации системы обнаружилось, что такой подход требует значительных вычислительных ресурсов, т.е увеличивается нагрузка на сервер.

Поэтому было принято решение о расширении базы знаний шаблонами html- фрагментов, что незначительно увеличивает затраты памяти, но значительно уменьшает нагрузку сервера и снижает временные затраты на обработку запроса пользователя.

В работе [3] разработана структура БЗ, состоящей из двух компонентов: хранилища онтологий (SearchIndex) и хранилища html- шаблонов., как это показано на рисунке 1.

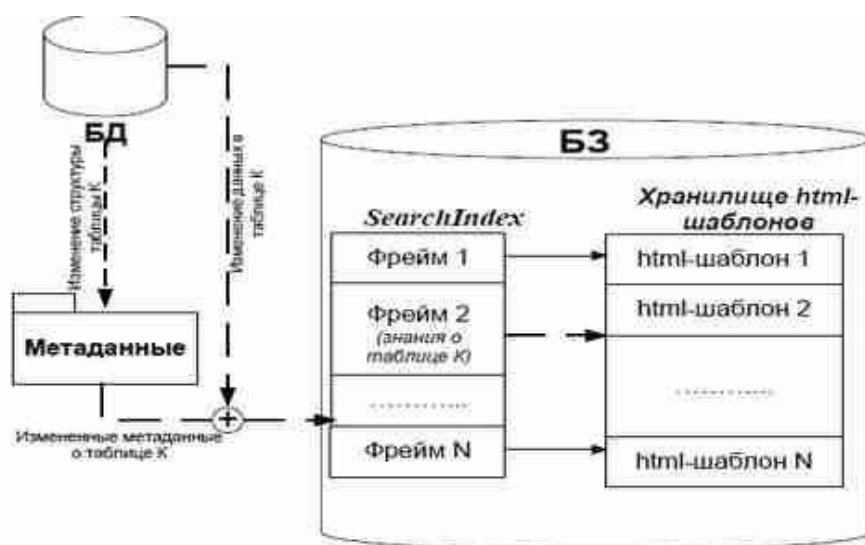


Рисунок 1 - Структура базы знаний

Заключение

В настоящей работе проведена оценка качества информационного поиска в слабо структурированных источниках на основе метаданных и базы знаний. Данная структура и схема хранилища метаданных сокращает нагрузку на серверы системы, что позволяет снизить требования к аппаратному обеспечению.

Предложенная структура базы знаний, содержащая не только онтологии предметной области, но и html-шаблоны, значительно снижает временные затраты на обработку запроса пользователя.

Использование представленной технологии информационного поиска с использованием базы знаний и метаданных, улучшает критерии полноты и точности поиска для различных категорий пользователей.

Литература

1. [Электронный ресурс]. - Режим доступа: w3.org;
2. Егошина А.А. Повышение эффективности извлечения информации из слабо структурированных источников на основе метаданных и базы знаний / А.А. Егошина, А.С. Вороной // Збірник наукових праць ДонНТУ. Серія «Інформатика, кібернетика і обчислювальна техніка». – 2011. - № 13(185). - С. 44-47.;
3. [Электронный ресурс]. - Режим доступа: prometheusresearch.com