

ПРИМЕНЕНИЕ МЕТОДОВ DATA MINING ДЛЯ АНАЛИЗА ПОВЕДЕНИЯ ПОСЕТИТЕЛЕЙ WEB САЙТА

USING DATA MINING TECHNIQUES FOR ANALYSING NAVIGATIONAL BEHAVIOUR OF WEB SITE VISITORS

Бедорев А.А., alex@krc.karelia.ru, Институт прикладных математических исследований КарНЦ РАН, г. Петрозаводск, Россия

Bedorev A.A. alex@krc.karelia.ru, Institute of Applied Mathematical Research KarRC RAS, Petrozavodsk, Russia

Журавлев Д.Н., zhuravle@krc.karelia.ru, Институт прикладных математических исследований КарНЦ РАН, г. Петрозаводск

Zhuravlev D.N., zhuravle@krc.karelia.ru, Institute of Applied Mathematical Research KarRC RAS, Petrozavodsk, Russia

An approach for analyzing navigational behaviour of Web site visitors is proposed. The targets of such analysis are optimizing Web site logical structure and analyzing thematic interests of Web site visitors. Applying regular episodes search algorithm to a database that contains logical, thematic and hyperlink structure of Web site allows making deeper analysis of commercial Web site efficiency and produces recommendations for optimizing its structure.

В настоящее время, для анализа посещений Web сервера широко применяются рейтинговые показатели и критерии, построенные на основе подсчета количества обращений к Web страницам сервера. Недостатками применения рейтинговых показателей являются то, что они не во всех случаях адекватно отражают уровень популярности информационного ресурса и не могут помочь в оптимизации структуры Web сайта, т.к. не учитывают структуры связей между Web страницами сайта. Более качественную информацию для повышения эффективности функционирования Web сайта дает применение методов Data Mining, в частности алгоритма поиска регулярных эпизодов. Целями анализа поведения посетителей Web сайта являются: оптимизация структуры сайта; определение тематической направленности интересов посетителей Web сайта; выявление попыток сканирования содержимого сайта, а также определение параметров сканирования.

Для решения вопросов, связанных с анализом поведения посетителя Web сайта предлагается реализовать подход, основанный на применении методов Data Mining (поиска регулярных эпизодов) для анализа данных, хранящихся в таблицах реляционной СУБД. При этом в одну таблицу помещается информация, отражающая структуру Web сайта на различных уровнях (тематическом, структурном и в терминах файловой системы), а во второй таблице хранится информация о посещениях Web сайта. При сопоставлении информации из указанных таблиц посредством стандартных SQL запросов в сочетании с применением методов Data Mining (поиском регулярных эпизодов), становится возможным более детальный и глубокий анализ поведения посетителей Web сайта.

Для заполнения информацией первой таблицы предлагается использовать специально разработанную программу, выполняющую последовательное сканирование всех Web страниц сайта. Информация для второй таблицы извлекается из лог-файла Web сервера, хранящего историю посещений сайта.

После заполнения вышеуказанных таблиц, к базе данных применяется алгоритм поиска регулярных эпизодов. Найденные регулярные эпизоды в данном случае являются наиболее часто встречающимися цепочками посещений страниц Web сайта. На основе полученных данных составляются рекомендации по оптимизации структуры Web сайта, следующего вида:

- перемещение (изменение порядка следования) гиперссылок на Web странице в соответствии с частотой их использования посетителями;
- установка дополнительных гиперссылок между первым и последним элементами (Web страницами) регулярно встречающихся цепочек посещений;
- установка дополнительных гиперссылок на последние элементы регулярно встречающихся цепочек посещений в соответствующем структурном (тематическом) разделе.