

УДК 001.894.2

## ТЕХНОЛОГИЯ WEB MINING

А. Н. Борисова  
Научный руководитель – Т. Г. Долгова

Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева  
Российская Федерация, 660037, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31  
E-mail: a.anariel@mail.ru

*Рассматриваются эффективные решения задач поиска, структурирования и анализа в основном хаотично организованной информации в сети с помощью нового направления в методологии анализа данных – Web Mining.*

*Ключевые слова: извлечение и анализ информации, технологии Web Mining, Call Mining.*

## WEB MINING TECHNOLOGY

A. N. Borisova  
Scientific supervisor – T. G. Dolgova

Reshetnev Siberian State Aerospace University  
31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation  
E-mail: a.anariel@mail.ru

*The paper deals with the problem of effective, structuring and analysis largely chaotically organized information on the network with the new directions in the methodology of data analysis – Web Mining.*

*Keywords: extraction and analysis of information, technology Web Mining, Call Mining.*

Web Mining – это современная технология, которая была создана для исследования и дальнейшего извлечения информации из сервисов и web-документов. Данная технология использует методы Data Mining.

Data Mining – технология, которая изучает процессы нахождения новых и потенциально полезных для человека знаний в базах данных [1]. Технология Data Mining использует методы сразу нескольких научных направлений, основные из которых – статистика, искусственный интеллект и системы баз данных.

Необходимость автоматической обработки и анализа информации, полученной из сети Интернет, вызвана высокой доступностью для пользователя большого количества постоянно пополняющейся и обновляющейся информации [2].

Можно выделить следующие проблемы работы с информацией, полученной из глобальной сети:

- поиск значимой информации;
- создание новых знаний вне информации, доступной на Web;
- персонализация информации;
- изучение потребителя или индивидуального пользователя.

Для решения данных проблем используют различные технологии, которые могут напрямую или косвенно разрешать их. К таковым относятся: базы данных, информационный поиск, обработки естественных языков и др.

Выделяют 4 основных этапа применения данной технологии:

1. Поиск ресурсов.
2. Извлечение информации.
3. Обобщение.
4. Анализ.

Web Mining рассматривается в одном ряду с такими интернет-технологиями, как получение информации (Information Retrieval – IR) и извлечение информации (Information Extraction – IE) [3]. Однако, имея с ними много общего.

В области Web Mining выделяют следующие направления анализа:

- извлечение Web-контента (Web Content Mining);
- извлечение Web-структур (Web Structure Mining);
- исследование использования Web-ресурсов (Web Usage Mining).

Технология Web Mining тесно связана с другими направлениями Data Mining. Для анализа текстовой информации используются методы Text Mining. Для анализа изображений, видео- и аудиоинформации используется Multimedia Mining и Call Mining.

На сегодняшний день наибольшее распространение технология получила при решении задач в сфере бизнеса [4].

В бизнес-аналитике Web Mining решает следующие задачи:

- кластеризация и классификация посетителей сайта;
- поиск популярных наборов, ассоциативных правил;
- нахождение зависимостей при пользовании услугами сайта.

Одним из самых популярных инструментов Web Mining для корпоративных информационных систем стала технология Call Mining («добыча» и анализ звонков). Цель использования Call Mining – упрощение поиска в аудио-архивах, содержащих записи переговоров между операторами и клиентами.

Среди фирм – разработчиков новой технологии – компании CallMiner, Nexidia, ScanSoft и Witness Systems [5]. Было разработано два кардинально отличающихся между собой подхода – на базе фонетического анализа и на основе преобразования речи в текст.

Примером ПО для фонетической аналитики служит продукция американской компании Nexidia. Типичная инсталляция продукции от разработчика Nexidia для компании обходится в сумму от 100 до 300 тыс. долларов.

Самым ярким примером ПО на базе подхода преобразования речи в текст является система CallMiner. Стоимость внедрения этой системы в среднем составляет около 450 тыс. долларов.

Аналитики отмечают, что за последние годы интерес к системам на основе Call Mining значительно возрос [5].

В заключение можно сказать, что использование Web Mining-технологий повышает оперативность работы компаний и снижает стоимость обработки информации.

### Библиографические ссылки

1. Data Mining [Электронный ресурс] // Википедия. URL: [https://ru.wikipedia.org/wiki/Data\\_mining](https://ru.wikipedia.org/wiki/Data_mining) (дата обращения: 21.12.2014).
2. Технологии анализа данных [Электронный ресурс] // Base Group. URL: [http://www.basegroup.ru/library/web\\_mining/basic\\_conceptions](http://www.basegroup.ru/library/web_mining/basic_conceptions) – BaseGroup Labs (дата обращения: 21.12.2014).
3. Web Mining: интеллектуальный анализ данных в сети Internet [Электронный ресурс] // Управление знаниями. URL: <https://sites.google.com/site/upravlenieznaniami/tehnologii-upravleniaznaniami/text-mining-web-mining/web-mining> (дата обращения: 22.12.2014).
4. Подходы к извлечению данных из веб-ресурсов [Электронный ресурс] // Хабрахабр. URL: <http://habrahabr.ru/post/99918> (дата обращения: 23.12.2014).
5. Сферы применения Data Mining [Электронный ресурс] // Data Mining. URL: <http://bug.kpi.ua/stud/work/RGR/DATAMINING/spheresofapplication.html> (дата обращения: 23.12.2014).

© Борисова А. Н., 2015