

С.И. Коровкин

Беларусь, Гомель, ГГУ имени Ф. Скорины

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И МЕТОДЫ АНАЛИЗА WEB-СТРАНИЦ В СЕТИ ИНТЕРНЕТ

В настоящее время все больше людей предпочитают получать и делиться информацией в сети Интернет: оставляют всевозможные отзывы о товарах, различных продуктах, местах, где побывали, и т.д. Чтобы повысить качество предлагаемых товаров и услуг, компаниям важно знать, как реагируют на их деятельность клиенты. Это очень большое количество информации, которое нужно обработать, структурировать и анализировать. В результате оценки данной информации компании могут делать выводы о дальнейших путях развития и способах привлечения новых клиентов. Таким образом, создание приложения для сбора и анализа различных моделей данных, таких как отзывы о товарах, записи из социальных сетей и др., является актуальной разработкой.

Web Mining – это использование методов интеллектуального анализа данных для автоматического обнаружения веб-документов и услуг, извлечения информации из веб-ресурсов и выявления общих закономерностей в Интернете.

Целью информационного извлечения является превращение коллекции документов в легко усвояемую и проанализированную информацию. Благодаря динамике и разнообразию веб-содержимого, создание ручного режима систем информационного извлечения не представляется возможным. В связи с этим большинство систем по извлечению данных сосредотачивают внимание на конкретные веб-сайты. Другие используют обучающие машины или методы интеллектуального анализа данных и способны извлекать веб-документы в автоматическом или полуавтоматическом режиме. С этой точки зрения Web Mining является частью процесса извлечения информации из Интернета.

Сбор данных – непростая задача, так как используемые алгоритмы могут быть очень сложными, а данные не всегда доступны в одном месте. При разработке приложения следует учесть следующее:

1. Методология получения данных и взаимодействие с пользователем.
2. Проблемы с производительностью.
3. Проблемы с разнородными типами данных.
4. Хранение данных.
5. Анализ данных.
6. Визуализация данных.