

АНАЛИЗ ЭФФЕКТИВНОСТИ ВЕБ-САЙТА НА ОСНОВЕ МЕТОДОВ WEB MINING

Пехова Александра Сергеевна

студент института экономики и управления АПК, Российский государственный аграрный университет – МСХА имени К. А. Тимирязева, РФ, г. Москва

WEBSITE PERFORMANCE ANALYSIS BASED ON WEB MINING METHODS

Alexandra Pekhova

Student of the Institute of Economics and Management of the Agro-Industrial Complex, K.A. Timiryazev Moscow State Agricultural Academy, Russia, Moscow

Аннотация. В эпоху развития цифровых технологий интернет становится богатейшим источником информации, все чаще можно услышать такое понятие, как цифровая трансформация общества и в этой связи актуальность webmining становится неоспоримой в стремительно набирающем темпы электронном бизнесе. Этим и обусловлено исследование данной темы.

Abstract. In the era of digital technology development, the Internet is becoming the richest source of information, it is increasingly possible to hear such a concept as the digital transformation of society, and in this regard, the relevance of webmining is becoming indisputable in the rapidly gaining pace of electronic business. This is the reason for the study of this topic.

Ключевые слова: web mining, искусственный интеллект, машинное обучение, интеллектуальный анализ данных.

Keywords: web mining, artificial intelligence, machine learning, data mining.

Использование инструментов и методов webmining позволяет быстро и с максимальной эффективностью выполнить анализ существующего веб-ресурса по различным критериям и предоставить аналитический материал для оценки эффективности ресурса и рекомендации по устранению выявленных дефектов.

С каждым годом всемирная паутина разрастается с невероятной скоростью. Объем данных, хранящихся в интернете, растет в геометрической прогрессии и по состоянию на 2022 году объем данных, которые пользователи создают и копируют, достиг 44 зеттабайта или 44 триллионов гигабайт [3]. Необходимость автоматического анализа информации из интернета вызвана высокой доступностью огромного количества постоянно пополняющейся информации, а также растущей популярностью веб-услуг среди всех категорий пользователей. Такая популярность информационных технологий отражена в действующих стандартах РФ ГОСТ Р 58608-2019 [1] и ГОСТ Р ИСО/МЭК 38506-2022 [2].

Развитие Веба в глобальную информационную инфраструктуру позволило обычным пользователям быть не только потребителями информации, но ее создателями и распространителями. В этой связи для эффективного решения задач поиска, структурирования и анализа в основном хаотично организованной информации в сети предназначено новое направление в методологии анализа данных – webmining [4].

Webmining развивается на пересечении таких дисциплин как обнаружение знаний в базах данных, эффективный поиск информации, искусственный интеллект, машинное обучение и обработка естественных языков [5].

В настоящее время пользователи интернета все чаще сталкиваются с такими основными проблемами, как небольшой процент действительно нужной информации среди множества ссылок, которые предоставляют поисковые системы и низкая повторяемость вызовов, связанная с невозможностью индексировать все Web-ресурсы. В результате возникают трудности поиска неиндексированной информации, которая нужна пользователю.

Для решения этих проблем используются различные технологии, напрямую или косвенно разрешающие их. К таким технологиям относятся, например, базы данных, информационный поиск, обработчики естественных языков и другие. К этому же числу технологий относится и веб-майнинг.

Веб-майнинг — это процесс методов интеллектуального анализа данных для автоматического обнаружения и извлечения информации из веб-документов и служб. Основная цель веб-майнинга — обнаружение полезной информации из всемирной паутины и моделей ее использования. Использование методов веб-майнинга помогает улучшить возможности поисковой системы в Интернете, классифицируя веб-документы и идентифицируя веб-страницы.

Веб-майнинг также используется для прогнозирования поведения пользователей, оптимизации целевой страницы сайта, для веб-поиска (Google, Yahoo) и вертикального поиска (FatLens, Vesome).

Интеллектуальный анализ веб-контента — это приложение для извлечения полезной информации из содержимого веб-документов. Веб-контент состоит из нескольких типов данных — текста, изображения, аудио, видео и т. д. Данные контента — это группа фактов, которые предназначены для веб-страницы. Он может предоставить эффективные и интересные шаблоны о потребностях пользователей. Текстовые документы связаны с анализом текста, машинным обучением и обработкой естественного языка. Этот майнинг также известен как интеллектуальный анализ текста. Этот тип анализа выполняет сканирование и анализ текста, изображений и групп веб-страниц в соответствии с содержанием ввода.

Интеллектуальный анализ веб-структуры — это приложение для обнаружения информации о структуре из Интернета. Структура веб-графа состоит из веб-страниц как узлов и гиперссылок как ребер, соединяющих связанные страницы. Интеллектуальный анализ структуры в основном показывает структурированное резюме конкретного веб-сайта. Он определяет отношения между веб-страницами, связанными информацией или прямой связью. Для определения взаимосвязи между двумя коммерческими веб-сайтами, анализ веб-структуры может быть очень полезен.

Интеллектуальный анализ использования Интернета — это приложение для выявления или обнаружения интересных шаблонов использования из больших наборов данных. И эти шаблоны позволяют вам понять поведение пользователей или что-то в этом роде. При анализе использования Интернета пользователи получают доступ к данным в Интернете и собирают данные в виде журналов. Таким образом, интеллектуальный анализ использования Интернета также называется интеллектуальным анализом журналов [6].

Поиск ресурсов предполагает поиск различных Web-источников (преимущественно текстовых) по ключевым словам. Данный этап разделяют на два класса: поиск документов и поиск сервисов, которые реализуются поисковыми роботами, такими как, например,

WebCrawler и AltaVista.

После того как ресурсы найдены, из них должна быть извлечена информация, подвергаемая анализу. Часто этот этап называют препроцессинг, т. к. он заключается в подготовке найденных ресурсов непосредственно к анализу. Такая подготовка заключается в преобразовании текстов, путем удаления стопслов, стеммингов, извлечением фраз и словосочетаний и т. п. Другими словами, результатом данного этапа должна быть информация, пригодная для анализа.

На этапе обобщения к обработанной информации уже применяются методы DataMining, который непосредственно и осуществляет обобщение и последующую интерпретацию полученных результатов.

На сегодняшний день существует достаточно большое количество разнообразных инструментов с помощью, которых можно провести всесторонний анализ любого интересующего нас ресурса, доступ к которому возможен по его URL-адресу.

Webmining является новым перспективным направлением анализа интернет-ресурсов для оптимизации структуры веб-сайтов, получения знаний о посетителях сайта, описания социальных сетей и сообществ, а также для автоматического поиска и структуризации информации из интернета.

Методами Webmining можно оптимизировать страницы сайтов, выявлять неучтенные категории целевой аудитории и многое другое.

Детальное изучение полученных статистических данных и рекомендации по оптимизации ресурса помогут владельцу ресурса сделать свой сайт более привлекательным и удобным для посетителей и клиентов, а значит потенциально позволят увеличить свою клиентскую базу и прибыль, что является конечной целью существования любого коммерческого предприятия.

Список литературы:

1. ГОСТ Р 58608-2019 «Информационные технологии. Стратегическое управление ИТ. Структура и модель» (утвержден и введен приказом Федерального агентства по техническому регулированию и метрологии от 25 июля 2022 года N 679-ст. Дата введения 2021-01-01). — URL : https://www.rst.gov.ru/portal/gost/home/standarts/catalognational?portal:componentId=3503536e-2ac1-4753-8ed1-09a92fee02de&portal:isSecure=false&portal:portletMode=view&navigationalstate=JBPNS_rO0ABXdKAAZhY3Rpb24AAAABABBjb25jcmV0ZURvY3VtZW50AARmcm9tAAAAQAEMTAwMAAGZG9jX2lkAAAAQAfNDMwNzEAB19fRU9GX18* (дата обращения 2022-09-10),
2. ГОСТ Р ИСО/МЭК 38506-2022. Национальный стандарт Российской Федерации "Информационные технологии (ИТ). Управление ИТ. Применение ИСО/МЭК 38500 для управления инвестициями в ИТ" (утвержден и введен приказом Федерального агентства по техническому регулированию и метрологии от 25 июля 2022 года N 679-ст. Дата введения 2022-30-11). — URL : <https://www.consultant.ru/cons/cgi/online.cgi?req=doc;base=OTN;n=34173#eQbv6QTW0cSjEDqx> (дата обращения 2022-09-10).
3. Prowebscraper, 2022.- URL: <https://prowebscraper.com/blog/web-mining-tools/>(дата обращения 2022-07-10),
4. businessdataanalytics, 2007-2022.- URL: <http://businessdataanalytics.ru/WebMining.htm> (дата обращения 2022-07-10),
5. interface.ru, 2022.- URL: <https://www.interface.ru/home.asp?artId=25898> (дата обращения 2022-07-10),

6. geeksforgeeks, 2022.- URL: <https://www.geeksforgeeks.org/web-mining/> (дата обращения 2022-07-10),

7. studfile.net, 2022.- URL: <https://studfile.net/preview/6139731/page:93/> (дата обращения 2022-07-10),