

Эта часть работы выложена в ознакомительных целях. Если вы хотите получить работу полностью, то приобретите ее воспользовавшись формой заказа на странице с готовой работой:

<https://studservis.ru/gotovye-raboty/kurovaya-rabota/218624>

Тип работы: Курсовая работа

Предмет: Программирование

СОДЕРЖАНИЕ

ВВЕДЕНИЕ 2

1. Аналитическая часть 3

1.1 Проведение анализа предметной области 3

1.2 Формулировка задачи 15

2. Практическая часть 16

2.1 Алгоритм парсинга сайта инструментами Data Mining 16

2.2 Обучение (получение векторов термов и списка категорий) 24

2.3 Результаты парсинга сайта 25

ЗАКЛЮЧЕНИЕ 28

СПИСОК ИСТОЧНИКОВ 29

Приложение А 32

ВВЕДЕНИЕ

Технология интеллектуального анализа данных (англ. data mining) помогает извлекать полезную информацию из различных баз данных. Хранилища данных хорошо работают с числовой информацией, но они не приносят успеха, когда речь идет о текстовой информации. XXI век предоставил нам неограниченные объемы информации в сети Интернет. Это хорошо, потому что больший объем информации обеспечивает большую осведомленность и лучшее знание. В настоящее время большая часть информации в бизнесе, промышленности, государственных и других учреждениях хранится в текстовой форме в базах данных. Интеллектуальный анализ текста представляет собой вариацию на область, называемую интеллектуальный анализ данных (data mining), которая служит для нахождения интересных образцов из огромных баз данных. Для обработки неструктурированных документов на основе различных предметных исследований были разработаны такие технологии поиска информации, как методы парсинга содержимого сайтов. При традиционном поиске, пользователь, как правило, ищет уже известные термины, которые были написаны кем-то другим.

С каждым днем количество веб-страниц растёт, увеличивается конкуренция между предприятиями в Интернете, сайтам становится сложнее занять верхние позиции в поисковой системе. Поэтому в настоящее время для повышения конкурентоспособности компании и продвижения веб-страницы используются методы поисковой оптимизации сайта - комплекса мер для поднятия позиции сайта в результатах выдачи поисковых машин по определенным запросам пользователей [1]. В данном исследовании рассмотрим основу стратегии парсинга сайтов.

Целью исследования является разработка и исследование алгоритма парсинга семантического ядра веб-сайтов на основе методов интеллектуального анализа данных с использованием среды программирования Python.

1. Аналитическая часть

1.1 Проведение анализа предметной области

Сеть Интернет является глобальной информационной сетью, все части которой связываются друг с другом в едином адресном пространстве, основанном на использовании протокола TCP/IP. Сеть Интернет построена из множества взаимосвязанных компьютерных сетей и предоставляет удаленный доступ к компьютерам, доскам объявлений, электронной почте, базе данных и дискуссионным группам [1].

Интернет – это перспектива коммуникаций. В настоящее время Сеть дает возможность получать видео из любого уголка мира, отправлять электронную почту с картинками, использовать Интернет-телефонию и организовывать международные конференции.

Веб-дизайном называют создание дизайна сайта и его отдельных частей, создание иллюстративных и рекламных материалов в Интернет. Часто услуги веб-дизайна предоставляются различными студиями веб-

дизайна.

Веб-сервер – это сервер, который обеспечивает снабжение информацией в службе глобального соединения. Веб-сервер хранит и предоставляет во внешнюю сеть информацию, которая имеется на веб-страницах. Веб-сервер является ответственным за обработку запроса клиента к веб-сайту и исполнение CGI-, JSP-, ASP-, PYTHON- и других приложений.

Веб-страница – это самостоятельная составляющая веб-сайта; документ, имеющий уникальный адрес (URL). Веб-страница может иметь динамическое или статическое построение. Как правило, веб-страницы организованы в виде гипертекста с включениями текста, графики, звука, видео или анимацию. В сети Интернет просмотр веб-страниц происходит посредством браузера.

В настоящий момент широко применяют концепцию веб-сайта, когда пользователю не нужно выходить из дома чтобы приобрести себе какой-либо товар или получить новую информацию [6].

На данный момент веб-сайт представляет собой новую, но уже устоявшуюся систему реализации товаров или услуг. По сути эта система является более эффективной по сравнению со стандартным «физическим» представлением информации. Применение веб-сайта в качестве информационной системы для электронной коммерции дает возможность:

а) улучшения качества обслуживания клиента, так как все действия по покупке товара производят в автоматическом режиме, а менеджеру следует только уточнять информацию и выдавать распоряжение о доставке товара;

б) снижения трудоемкости и напряженности труда персонала компании, минимизации ошибок в их действиях.

в) снижения издержек, так при этом отпадает необходимость в аренде здания для отдельного офиса.

Быстрое развитие информационных технологий, среди которых одно из главных мест занимает Интернет, появление и стремительный рост электронной коммерции стали основной причиной для развития новой ветви в нынешней концепции маркетинга взаимодействия - Интернет-маркетинга.

Однако авторы многих имеющихся сегодня в Интернет-публикаций трактуют понятие Интернет-маркетинга как продвижение каких-либо товаров или услуг в Интернете, или объясняют его еще более односторонне: как рекламу в сети Интернет, либо еще ограниченнее – только как баннерную рекламу. Подобные трактовки неверны именно в силу узости такого понимания рассматриваемого вопроса [7].

Как правило, Интернет-маркетинг включает целый набор дочерних отраслей. Они содержат не только баннерную рекламу и public relations, а также и методiku проведения маркетингового исследования в Интернете, а именно, изучение спроса и потребительской аудитории, изучение алгоритма формирования и обеспечения высокой эффективности рекламной кампании, способов рационального позиционирования торговой марки на рынках, и многое другое. Рассмотренное ранее узкое понимание данного предмета является, вероятно, своего рода «детской болезнью» Интернет-маркетинга. Это, в свою очередь, можно анализировать как беспристрастный положительный признак. Наличие подобных «детских болезней» свидетельствует о том, что Интернет-маркетинг у нас растет быстро и развивается активно.

Интернет-маркетинг — это теория и методология организации дела по маркетингу в среде Интернета [3].

Интернет имеет уникальные характеристики, которые серьезно отличаются от характеристик традиционных приемов маркетинга. Например, одним из главных качеств среды Интернет является ее гипермедийная природа. Это качество характеризуется высокой эффективностью в представлении и усвоении информации. В связи с этим значительно повышаются возможности маркетинга в усилении взаимосвязи предприятий и потребителей. Надо отметить, что многие специалисты в области PR и маркетинга оказались к этой революции не подготовлены. Они не смогли приспособиться, не воспринимают Интернет как уникальное киберпространство, где осознанно находятся миллионы людей. Поэтому многие компании были выброшены на обочину, стали одиночками и забытыми в этом киберпространстве [9].

Другая категория компаний смогла войти в Интернет, но не смогла добиться успехов, так как больше заботились о том, что они хотят получить от Интернета, не стремились при этом разобраться с требованиями новой целевой аудитории.

Анализируя определение маркетинга на электронном рынке, важно также ввести понятия, что является объектом и субъектом маркетинговой деятельности на данном рынке. Объектом маркетинговой деятельности на электронных рынках является информационно-аналитическое и экспертно-исследовательское функционирование фирм, с применением сетевых информационных технологий и систем:

а) выбор конкурентных позиций на данных рынках, на которых компания действует со своими товарами;

б) определение стратегии продвижения продукции и её распределения, выбор рекламной и ценовой

политик с учетом всех совокупностей факторов внешней и внутренней среды в условиях рисков и неопределенностей.

Субъектом маркетинговой деятельности на электронных рынках является, как и при использовании классического маркетинга, деятельность конкретных собственников или компаний по целенаправленному регулированию деятельности организаций. Регулирование производят по определенной технологии с применением систем методов анализа и обработки цифровых данных электронного рынка для реализации поставленной цели [12].

Процессы управления в маркетинге на электронных рынках отображают совокупность операций и процедур, которые выполняют работники маркетинговых служб компаний в нужной последовательности.

Он включает:

- а) сбор и анализирование данных о поведении рынка и конкурентов на нем с помощью применения возможностей сети Интернет, хранилища корпоративной базы данных;
- б) экспертизу стохастических и динамических процессов на данном сегменте рынка;
- в) построение модели психологических реакций поведения потребителей на электронном рынке, построение модели стратегии компании в условиях рисков и неопределенностей, формулировка новой стратегии развития текущего рынка, внедрения на рынок, стратегических альянсов и консолидации, стратегии диверсификации и др.

Перед тем, как переходить к принципам продвижения сайта, рассматриваем возможные уже решенные аналогичные задачи, для решения которых система электронной коммерции создается и продвигается.

- а) Модель Ad Supported (рекламная бизнес-модель). В рамках данной модели сайты создаются для образования с его использованием постоянной, четко сегментированной или, напротив, максимально широкой аудитории посетителей. Контакты с этой совокупностью пользователей продают рекламодателю или спонсору. Можно привести множество примеров. Например: www.citycaty.ru является крупнейшим рассылочным сервером, который получает доходы исключительно от использования рекламных сообщений в рассылках; www.autoca.ru является крупным автомобильным порталом Рунета,

СПИСОК ИСТОЧНИКОВ

1. Бондарчук Д. В., Тимофеева Г. А. Выделение семантического ядра на основе матрицы корреспонденций термов // Системы управления и информационные технологии. — 2015. — Т. 61, № 3.1. — С. 134—139.
2. Бьюзен Т. Интеллект-карты / пер. с англ. Ю. Константиновой. — М.: Манн, Иванов и Фербер, 2019. — 199 с..
3. Нейро-нечеткие методы в интеллектуальных системах обработки и анализа многомерной информации. — Томск: Изд-во Том. ун-та, 2014. — 442 с.
4. Data Mining [Текст] : учебное пособие [для студентов вузов, обучающихся по специальностям в области информационных технологий] / И. А. Чубукова. - 2-е изд., испр. - Москва : Интернет-университет Информационных Технологий : БИНОМ. Лаборатория знаний, 2013. - 382 с.
5. Loginom [Электронный ресурс]. — Режим доступа: URL: <https://basegroup.ru/deductor/download> (Дата обращения: 01.04.2021)
6. Serpstat [Электронный ресурс]. — Режим доступа: URL: <https://serpstat.com/ru/> (Дата обращения: 01.04.2021).
7. Wordstat yandex [Электронный ресурс]. — Режим доступа: URL: <https://wordstat.yandex.ru/> (Дата обращения: 01.04.2021).
8. XMind - Mind Mapping Software [Электронный ресурс]. — Режим доступа: URL: <https://www.xmind.net/> (Дата обращения: 01.04.2021).
9. Багиев Г. Л. и др. Маркетинг: Учебник для вузов. — М.: ОАО «Изд-во «Экономика»», 2018 г.
10. Багиев Г. Л., Успенский И. В., Ченцов В. И. Интерактивные модели маркетинговых решений на виртуальных рынках. — СПб: СПГУЭиФ, 2011 г.
11. Быков В. А. Электронный бизнес и безопасность. — М.: Радио и связь, 2015 г.
12. Бокарев Т. Энциклопедия Интернет-рекламы. — М.: Издательство «ПРОМО-РУ», 2016 г.
13. Гульяев А.К., Машин В.А. Уроки Web-мастерства. Технология. Дизайн. Инструменты.- М.:Корона-Принт, 2015г. - 448с
14. Галкин С. Е. Бизнес в Интернет. — М: «Центр», 2013 г.
15. Гуров Г. Г. Интернет для бизнеса. — М., 2013 г.
16. Евдокимов Николай , Лебединский Игорь Раскрутка веб-сайта. Практическое руководство; Вильямс - Москва, 2016. - 288 с.
17. Информационные системы в экономике. Под ред. проф. В. В. Дика. — М.: Финансы и статистика, 2012 г.

18. Клонингер, Курт Свежие стили Web-дизайна: как сделать из вашего сайта "конфетку"; М.: ДМК Пресс - Москва, 2017. - 232 с.
19. Лешев, Д. Создание интерактивного Web-сайта. Учебный курс; СПб: Питер - Москва, 2014. - 544 с.
20. Ломов, А.Ю. HTML, CSS, скрипты: практика создания сайтов. - СПб. : БХВ-Петербург, 2017. - 416с.
21. Матросов, А. Сергеев, М. Чаунин «HTML 4.0» Издательство «БХВ-ПИТЕР», Санкт - Петербург, 2014 - 340 с. конференции «Индустрия программирования 19». 2019 г. с.117-123.
22. Стрэндж К. (Strange K.). "Магический квадрат" СУБД для Хранилищ данных (Magic Quadrant for Data Warehouse DBMSs/). Gartner Research Note M-22-2154. 2014.
23. ADO Сборник рецептов. Для профессионалов. Билл Гамильтон.
24. М. Русинович, Д. Соломон. Внутреннее устройство Microsoft Windows, 6-е издание. 2013.
25. Дейт К. Дж. Введение в системы баз данных. - М.: Вильямс, 2019.
26. Джеффри Д. Ульман, Дженнифер Уидом. Введение в системы баз данных. - М.: Лори, 2011.
27. Бойко В.В., Савинков В.М. Проектирование баз данных информационных систем. - М.: Финансы и статистика, 2019.
28. А.Н. Наумов, А.М. Вендров, В.К. Иванов. Системы управления базами данных и знаний: Справ. Изд. / Под ред. А.Н. Наумова. - М.: Финансы и статистика, 1991. -352 с.: ил. 1.Волгина Н.А, Одеговага Ю.Г Экономика труда- М.: Экзамен, 2016.
29. Редько, В.Н.; Басараб, И.А. Базы данных и информационные системы; Знание, 2013.

Эта часть работы выложена в ознакомительных целях. Если вы хотите получить работу полностью, то приобретите ее воспользовавшись формой заказа на странице с готовой работой:

<https://studservis.ru/gotovye-raboty/kurovaya-rabota/218624>