

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра математического и компьютерного моделирования

**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ПОИСКОВЫХ СИСТЕМАХ.
РАЗРАБОТКА ПРОГРАММ ПО РАСЧЕТУ PAGE RANK И ПОИСКА С
ПРИМЕНЕНИЕМ ВЕКТОРОВ С ПОМОЩЬЮ ЯЗЫКОВ PYTHON и C#**

Автореферат по бакалаврской работе

Студента 5 курса 561 Группы

направление 09.03.03 – Прикладная информатика

механико-математического факультета

Афанасьева Игоря Павловича

Научный руководитель
доцент, к.ф.-м.н., доцент

С. П. Шевырев

Зав. кафедрой
зав. каф., д.ф. – м. н.

Ю.А. Блинков

Саратов 2016

Введение. Выбранная тема бакалаврской работы является актуальной для Интернет - индустрии, так как с каждым днем все больше растет количество информации в сети Интернет. Однако далеко не всегда эта информация является качественной, актуальной и стоящей. В результате можно потратить большое количество времени для поиска необходимой информации. Единственным выходом для пользователя является поисковик, который хранит информацию об адресах и содержимом веб-страниц. Поисковики призваны помочь пользователю выбрать лучший документ среди сотен однотипных. Также на сегодняшний день существует масса продуктов известных компаний, которые отчасти являются персональными помощниками. Они основаны на разработках искусственного интеллекта в поиске. Данная тема также очень современна и актуальна, и потому рассматривается нами в этой бакалаврской работе.

С концептуальной точки зрения, вне всякого сомнения, существует универсальная форма интеллектуальности. И люди, и животные обладают небольшой долей этой способности и обнаруживают в себе конкретные проявления универсальной интеллектуальности. По-видимому, мы, люди, унаследовали большую часть этой универсальной способности. Но в связи с тем, что в основе нашей интеллектуальности лежит биологическая природа, мы лишены некоторых характеристик универсальной интеллектуальности (например, полноты и нейтральности).

Большинство исследователей в области компьютерных наук считают, что биологическая интеллектуальность доступна для воспроизведения и что интеллект не является исключительно атрибутом человека. Из этого предположения по существу следует вывод, что часть универсальной интеллектуальности можно перенести и на машину. Таким образом, искусственный интеллект – это искусственно созданный аналог интеллектуальности, вырабатываемой мозгом человека как биологического существа. Очевидно, что развитие технических средств приводит к получению иных результатов, чем биологическая эволюция, и в этом состоит объяснение того факта, что искусственный интеллект характеризуется другими свойствами

по сравнению с человеческим интеллектом (например, доскональностью во всем). Таким образом, искусственный интеллект – это еще одно проявление универсальной интеллектуальности.

Целью данной бакалаврской работы является рассмотрение теоретических основ понятия искусственного интеллекта в поисковых системах, его возникновения и методов построения. В практической части разрабатываются программы по двум составляющим искусственного интеллекта в поисковых системах.

Основное содержание работы. Техника сопутствует человеку в течение тысячелетий и представляет собой не что иное, как совокупный результат стремлений человечества к лучшей жизни. Сейчас, однако, находятся люди, которые считают, что развитие техники, напротив, ухудшает, а не улучшает жизнь. Проблемы, стоящие ныне перед людьми, различны по своим масштабам: от социальных сдвигов, вызванных технологическими изменениями, безработицы, загрязнения окружающей среды и угрозы ядерного уничтожения до отчуждения и неудовлетворенности работой и ее конкретными результатами. К этому можно добавить следующее. Не исключено, что именно сложность, порожденная техническим прогрессом, несет ответственность за не поддающиеся лечению недуги экономики и что значительную опасность начинают представлять технические системы, которые становятся настолько сложными, что вскоре людям будет недоставать знаний и понимания для управления ими.

Длительное время ошибочно считалось, что на выходе компьютера можно получить лишь то, что заложено в него на входе. Это представление, безусловно, находило себе подтверждение на протяжении последних 30 лет в большинстве работ, связанных с обработкой данных. Теперь, однако, неопровержимо доказано, что от компьютеров можно получить нечто совершенно новое, а именно знания. Эти знания в свою очередь могут принимать форму оригинальных идей, стратегий и решения реальных проблем.

Однако прогресс не стоит на месте, и каждый день появляется масса новых технологий и продуктов в различных сферах жизни человека. На сегодняшний день также существует множество успешно реализованных инноваций в сфере искусственного интеллекта. Рассмотрим понятие «искусственный интеллект» – это раздел информатики, изучающий возможность обеспечения разумных рассуждений и действий с помощью вычислительных систем и иных искусственных устройств. В настоящее время искусственный интеллект активно используется при разработке по следующим направлениям: представление знаний, игры и творчество, разработка

естественно-языковых интерфейсов и машинный перевод, распознавание образов, новые архитектуры компьютеров, интеллектуальные роботы, специальное программное обеспечение, обучение и самообучение. Также существует и активно используется миллионами людей искусственный интеллект в поисковых системах. Такие бренды как Apple, Microsoft, Google и Яндекс создали своих уникальных персональных помощников в поисковых системах с применением искусственного интеллекта: Siri, Cortana, GoogleNow и MatrixNet. Каждый пользователь может выбрать удобного для себя помощника, который будет упрощать взаимодействие с гаджетами.

Немного о каждом персональном помощнике и об обучающихся алгоритмах.

Siri - облачный персональный помощник и вопросно-ответная система, программный клиент которой входит в состав iOS, watchOS, macOS, и tvOS компании Apple. Данное приложение использует обработку естественной речи, чтобы отвечать на вопросы и давать рекомендации. Siri приспособливается к каждому пользователю индивидуально, изучая его предпочтения в течение долгого времени.

Первоначально Siri стало доступно в App Store как приложение для iOS от Siri Inc. Вскоре, 28 апреля 2010 года, Siri Inc. была приобретена Apple Inc. Но ещё до того, как Apple купила Siri, было объявлено, что их программное обеспечение будет доступно для телефонов BlackBerry и телефонов под управлением Android, затем эти планы были отменены из-за покупки

Кортана - виртуальная голосовая помощница с элементами искусственного интеллекта от Microsoft для Windows Phone 8.1, Microsoft Band, Windows 10, Android, Xbox One и iOS.

Впервые была продемонстрирована во время Конференции Build в Сан-Франциско 2 апреля 2014 года. Кортана была названа в честь героини серии компьютерных игр Halo - голос помощницы в версии для американского рынка принадлежит Джен Тейлор, которая также озвучивала Кортану в оригинальной

игре.

Google Now - это сервис персонализированного поиска Google, который понимает голосовые запросы. Основная его "фишка" заключается в абсолютной персонализации информации - приложение подстраивается под потребности и привычки пользователя. Мы рассмотрим основные возможности и команды Google Now.

Основное предназначение Google Now - выводить всю необходимую пользователю информацию до того, как он решится ввести поисковый запрос. Для этого система отслеживает текущее местонахождение человека, историю его запросов, интересы, маршруты, которыми он добирается до работы, сканирует его ящик электронной почты и так далее. Конечно, выглядит все это немного жутковато, и многие пользователи не хотят, чтобы Google знала о них настолько много. Тем не менее, тем, кто меньше боится за свои личные данные, Google Now может сильно облегчить жизнь.

Приложение Google Now позволяет быстро и удобно пользоваться Интернетом при помощи голосовых запросов. Это роднит его с ассистенткой Siri, предустановленной в "айфоны". Голосовой ввод Google очень точный и продуманный, а поисковая система надежно отыщет практически любую информацию.

MatrixNet - специальный алгоритм машинного обучения, благодаря использованию которого поисковые машины Яндекса стали учитывать гораздо большее количество критериев ранжирования, чем до введения этого алгоритма. Подчеркнём, MatrixNet является детищем российского поисковика Яндекс. Другие поисковые машины пользуются своими решениями.

Впервые о MatrixNet заговорили одновременно с введением поискового алгоритма «Снежинск» в 2009 году. Целью разработки MatrixNet было многократное увеличение количества факторов ранжирования и, как следствие, повышение качества поиска. В частности, применением особых

математических формул было достигнуто уменьшение количества ошибочных оценок важности тех или иных факторов. Говоря проще, поисковая машина, обученная с помощью MatrixNet, перестала выдавать желаемое за действительное, придавая несущественным поисковым критериям большой вес.

Другим нюансом MatrixNet является возможность настройки ранжирования сайтов под ту или иную категорию поисковых запросов без ухудшения качества поиска по остальным. Формула ранжирования, которую создаёт MatrixNet для обучения поисковой машины, порой содержит десятки тысяч значений различных коэффициентов. Возможность точной настройки процедуры ранжирования позволяет отсеивать, так сказать, нетематические страницы, содержащие тематические запросы. Всё это в конечном итоге ведёт к повышению точности поиска.

С введением алгоритма машинного обучения MatrixNet выдача Яндекса стала более релевантной несмотря на то, что количество просматриваемых страниц выросло в тысячи раз.

Главное назначение любой поисковой системы - давать ответы на вопросы пользователя. И не просто ответы - а такие, которые смогут удовлетворить интерес человека, и чтобы при этом он приложил поменьше усилий.

Когда человек делает запрос, то поисковая система не заходит непосредственно на сайты, а ищет по своей базе проиндексированных сайтов. В базе поисковик находит все страницы, которые полностью или частично совпадают с поисковым запросом или имеют какое-то отношение к нему и пытается расположить найденные страницы таким образом, чтобы релевантные страницы находились выше.

Составление всех результатов поиска в соответствии с запросами называется ранжированием. С ранжированием тесно связано качество поиска - получил ответ человек на свой вопрос, посетив предложенный сайт, или нет? А откуда поисковику знать, был ли получен ответ? Именно для этого нужен

искусственный интеллект - чтобы самостоятельно принимать решения о полезности страниц и еще и учиться при этом.

Как может поисковая система учиться? Прежде всего учитывая поведенческие факторы. Например, если человек задал поисковику вопрос, открыл несколько сайтов и на одном из них провел больше всего времени - можно считать, что именно там была найдена лучший или самый полный ответ и в будущем этот сайт можно поднимать выше в результатах поиска.

В данной бакалаврской работе раскрываются история создания искусственного интеллекта в поисковых системах, подходы и направления к построению искусственного интеллекта, современные продукты компаний в сфере искусственного интеллекта в поисковых системах. Также рассматриваются две составляющие искусственного интеллекта в поисковых системах: Page Rank и поиск с применением векторов. В практической части бакалаврской работы разрабатываются программы по выше изложенным составляющим искусственного интеллекта в поисковых системах.

В практической части бакалаврской работы написаны две программы по расчету Page Rank итерационным методом и методом графов. Для расчетов были использованы одинаковые входные данные, чтобы можно было в итоге сравнить, какой же из приведенных методов эффективнее. В итоге результаты получились одинаковыми, но для разного количества итераций. В итерационном методе осуществляется 100 итераций, а в методе графов 103, следовательно, расчет итерационным методом осуществляется быстрее и является более эффективным.

Также была написана программа по ранжированию документов с применением векторов. Работа программы показала, что применение векторов в поисковых системах очень эффективно и позволяет найти файл, более точно соответствующий запросу пользователя, основываясь лишь на фразе в запросе. В программе каждый отдельный документ задается вектором. Когда мы задаем запрос в поиске, программа сравнивает вектор файла и вектор нашего запроса, и чем больше вектора сходятся к одной параллели и чем меньше между ними

угол, тем больше они похожи. Структура бакалаврской работы выглядит следующим образом:

ВВЕДЕНИЕ

1 Теоретические основы искусственного интеллекта

1.1 История создания искусственного интеллекта

1.2 Подходы и направления к построению искусственного интеллекта

2 Искусственный интеллект в поисковых технологиях

2.1 Разработки компаний в сфере искусственного интеллекта в поисковых системах

2.2 Составляющие искусственного интеллекта в поисковых системах: PageRank и поиск с применением векторов

3 Разработка программ по составляющим искусственного интеллекта в поисковых системах с помощью языков Python и C#

3.1 Программа расчета PageRank

3.2 Программа поиска с применением векторов

ЗАКЛЮЧЕНИЕ

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

Заключение. В заключение данной бакалаврской работы, можно сделать вывод о том, что на сегодняшний день все большую роль играет искусственный интеллект. Человек все больше стремится упростить свою жизнь посредством автоматизации. Но для того, чтобы техника самостоятельно могла выполнять различные сложные задачи и в последствии полностью или частично заменить человека, ее необходимо «научить» мыслить подобно человеку.

В сфере информационных технологий искусственный интеллект изучен и широко применяется. Такие крупнейшие компании как Apple, Microsoft, Google и Яндекс создали своих уникальных персональных помощников в поисковых системах с применением искусственного интеллекта: Siri, Cortana, GoogleNow и MatrixNet. У каждого пользователя теперь есть возможность выбрать удобного для себя помощника, который будет упрощать взаимодействие с гаджетами.

В ходе данной бакалаврской работы были выполнены поставленные цели: были рассмотрены теоретические основы понятия искусственного интеллекта в поисковых системах, история его возникновения и методы его построения. В практической части были разработаны программы по двум составляющим искусственного интеллекта в поисковых системах: программа расчета PageRank и программа поиска с помощью векторов.