

УДК 004.8

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ПОИСКОВЫХ СИСТЕМАХ: ОБЗОР СОВРЕМЕННОГО СОСТОЯНИЯ ТЕХНОЛОГИЙ

Лукша Б.Н., студент гр. 981063, Лаптёнок Н.В., студент гр. 981063

*Белорусский государственный университет информатики и радиоэлектроники,
Институт информационных технологий,
г. Минск, Республика Беларусь*

Савенко А.Г. – магистр техн. наук

Аннотация. В работе представлен обзор современного состояния технологий искусственного интеллекта, которые используются в поисковых системах для ранжирования результатов поиска, обработки естественного языка в поисковых запросах, голосового поиска и поиска по изображению. Представлены результаты исследований, доказывающих эффективность и перспективность использования искусственного интеллекта в поисковых системах.

Ключевые слова. Поисковые алгоритмы, поисковый запрос, релевантность, машинное обучение, поисковые системы, искусственный интеллект, голосовой поиск, поиск по изображению, ранжирование результатов запроса.

Введение. Современные поисковые системы в значительной степени полагаются на искусственный интеллект для работы, и знание того, как работает искусственный интеллект в поиске, может помочь лучше ранжировать веб-сайты. Искусственный интеллект в поиске отвечает практически за все: от того, какие результаты поиска вы видите, до того, какие связанные темы будут предложены для дополнительно изучения. Сегодня современная поисковая система не может функционировать без сложного искусственного интеллекта.

Искусственный интеллект – это общий термин, описывающий ряд интеллектуальных технологий, способных учиться и совершенствоваться самостоятельно, творчески подходить к решению задач [1]. Можно выделить следующие современные технологии, использующие искусственный интеллект: машинное обучение; глубокое обучение; нейронные сети; голосовой поиск; компьютерное зрение; распознавание изображений; генерация естественного языка (NLG); обработка естественного языка (NLP) и др. Все эти технологии считаются искусственным интеллектом, хотя и предназначены для достижения разных результатов и находятся на разных уровнях зрелости. Но, тем не менее, все эти технологии объединяет то, что машины с искусственным интеллектом способны обучаться и совершенствоваться без постоянного вмешательства человека. Как только эти машины будут обучены людьми, они смогут работать автономно и продолжать обучаться на своих результатах, со временем становясь все лучше и лучше. Это обусловлено тем в том числе тем, что технологии искусственного интеллекта находят закономерности в данных, а затем используют эти закономерности для дальнейшего прогнозирования. Применительно к поисковым системам такие прогнозы могут включать в себя:

- какой продукт пользователь захочет купить в следующий раз;
- какой контент пользователь хочет получить далее;
- что пользователь пытается написать, и какой язык, скорее всего, будет следующим;
- что пользователь пытается сказать, и каков наиболее вероятный правильный ответ.

Когда системы с искусственным интеллектом делают прогнозы, они учатся на этих прогнозах, чтобы со временем становиться еще более «умнее». Пример такого применения искусственного интеллекта – функция Smart Compose в Gmail. Уже несколько лет назад Gmail мог предугадывать простые фразы и слова, которые пользователь собирался ввести следующими. Сегодня искусственный интеллект Smart Compose так хорошо изучил миллиарды электронных писем, что теперь может закончить за вас целые предложения. Более того, тот же тип технологии искусственного интеллекта сейчас настолько развит, что уже может писать целые статьи самостоятельно [2].

Такая способность к обучению делает искусственный интеллект более мощным, чем традиционное программное обеспечение, существовавшее до него. Именно поэтому искусственный интеллект является основополагающим для любой поисковой системы, используемой сегодня.

Сегодняшний поиск слишком сложен для людей или традиционных машин. По результатам исследования установлено, что только Google обрабатывает колоссальные 63 000 поисковых запросов каждую секунду, или более двух триллионов поисковых запросов в год [3].

Даже самая большая команда людей не смогла бы эффективно обработать такой объем поиска. И традиционное программное обеспечение, которое не может находить закономерности в данных и делать прогнозы, также не справляется с этой задачей.

Без искусственного интеллекта предоставлять точные результаты поиска в режиме реального времени с такой скоростью и масштабом, как в сети Интернет крайне трудная задача. Именно

поэтому сегодня искусственный интеллект может быть применим практически в каждой части поисковой системы, при решении таких задач, как:

- индексация страниц, опубликованных в Интернете, и понимание их содержания;
- интерпретация поисковых запросов путем понимания человеческого языка;
- сопоставление запросов с наиболее точными и качественными результатами;
- оценка и переоценка качества контента для постоянного улучшения результатов поиска.

Фактически, результат поиска в таких системах является прямым результатом решений, принятых искусственным интеллектом. Для лучшего понимания аспектов применения искусственного интеллекта в поисковых системах рассмотрим решаемые для этого задачи.

Поисковые рейтинги. Поисковые системы чаще всего используют искусственный интеллект для ранжирования веб-страниц, видео и другого контента в результатах поиска.

Google (и другие поисковые системы также) полагаются на сложный искусственный интеллект, чтобы определить, как контент будет ранжироваться. Алгоритмы, используемые этими системами искусственного интеллекта, имеют множество правил, определяющих приоритетность различных факторов, от типов ключевых слов в вашем контенте до пользовательского опыта вашего сайта.

Вместе эти факторы помогают искусственному интеллекту поисковой системы определять, какие страницы показывать в ответ на ваш поисковый запрос.

В то время как Google дает указания о том, что важно для его алгоритмов, никто за пределами компании не имеет полного представления о том, как эти алгоритмы принимают решения [4].

На самом деле большая часть современного поискового маркетинга сосредоточена на том, чтобы узнать как можно больше о том, как работает этот искусственный интеллект, чтобы компании могли ранжировать контент в результатах поиска.

Понимание поисковых запросов. Современные поисковые системы должны понимать естественный человеческий язык, чтобы находить пользователям нужную им информацию. Другими словами, чтобы предоставить вам правильный контент, поисковые системы должны понимать, что вы спрашиваете в первую очередь. Для этого используются модули по обработке естественного языка (NLP) – области искусственного интеллекта, посвященной обучению компьютеров понимать нашу письменную речь. И Google, и Microsoft используют NLP, чтобы понимать своих пользователей. Исследование Вашингтонского университета рассмотрело Яндекс, четвертую по величине поисковую систему в мире, и отметило, что у нее также есть некоторые передовые приложения NLP и машинного обучения [5].

Разработчики Яндекса обнаружили, что могут взять все предыдущие поисковые запросы своих пользователей и использовать их для оптимизации будущих поисков. Создав эти персонализированные результаты поиска, они увеличили рейтинг кликов примерно на 10% [6].

Search Engine Journal также указывает, что поиск Google использует некоторые из тех же практик, оптимизируя результаты поисковой системы на основе недавних предыдущих запросов и позволяя пользователям выполнять поиск только по фотографии (для этого уже используется другая область искусственного интеллекта). Сегодня Google продвинулся в понимании поисковых запросов на несколько гигантских скачков вперед. Используя предварительно обученную языковую модель BERT, поисковая система компании теперь понимает полные предложения. Это позволяет предположить, что BERT может понимать контекст поиска, а не только ключевые слова в нем [6].

Пример одной из схем обработки запроса на естественном языке представлен на рисунке 1.

Контроль качества. Ранее некоторым специалистами по SEO удавалось обманывать поисковые системы с помощью сомнительных приемов, которые известны как «черные методы SEO». К ним относятся агрессивное наполнение ключевыми словами, маскировка, невидимый текст и так далее. Независимо от метода, у них была одна цель: обыграть каждую страницу результатов поисковой системы (SERP). Безусловно, это наносило ущерб поисковым системам, потому что страницы, которые были в верхней части их результатов, необязательно были с контентом самого высокого качества.

В настоящее время компании обновили свои алгоритмы и используют искусственный интеллект для отделения высококачественного контента от низкокачественного спама. Каждое из этих «изменений алгоритма» потенциально оказывает огромное влияние на то, как компании ранжируют свой контент и страницы в поиске [6].

Голосовой поиск и поиск по изображению. Голосовой поиск и поиск изображений – это новые возможности поиска, которые стали возможными также благодаря искусственному интеллекту. Технологии искусственного интеллекта, такие как NLP, стали настолько продвинутыми, что они действительно могут понимать человеческие голоса в режиме реального времени, например, при использовании Google Assistant. Эти системы на базе искусственного интеллекта могут понимать человеческую речь, а затем переводить их в результаты поиска [7].

Технологии искусственного интеллекта для распознавания изображений, могут определять, что изображено на изображении, а затем формулировать поисковый запрос и предоставлять релевантные результаты поиска по этому изображению.

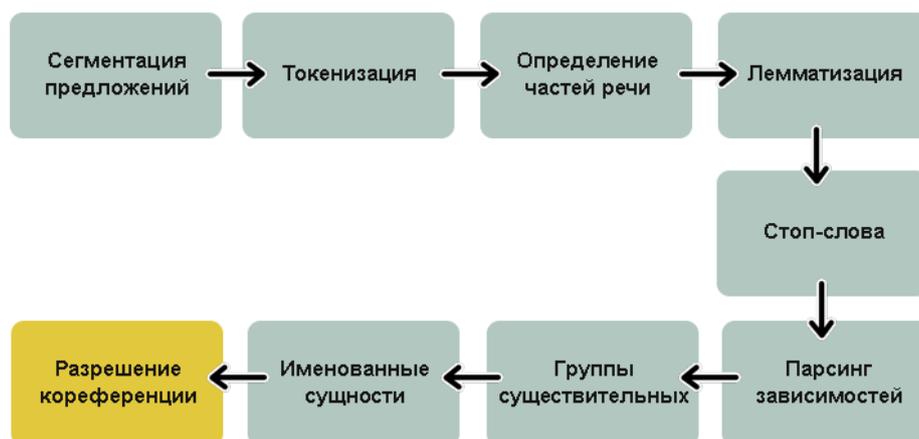


Рисунок 1 – Пример схемы обработки запроса на естественном языке

Заключение. Искусственный интеллект меняет оптимизацию поисковых систем, поскольку он позволяет одновременно анализировать и интерпретировать собираемые поисковыми системами данные для предоставления пользователям более качественных результатов.

Если компании хотят извлечь выгоду из этой «революции искусственного интеллекта», им необходимо работать над улучшением того, как они предоставляют контент своим пользователям. С точки зрения маркетинга, чем больше пользователи смогут удовлетворить свои потребности, тем больше поискового трафика будет приходить на сайты поисковых систем.

Список использованных источников:

- 1 Whitby, B. *Artificial Intelligence: A Beginner's Guide* / B. Whitby. – London : Oneworld Publications, 2008. – 192 p.
- 2 Создание умного сочинение в стиле Gmail с помощью языковой модели char ngram [Электронный ресурс]. Режим доступа: <https://towardsdatascience.com/gmail-style-smart-compose-using-char-n-gram-language-models-a73c09550447>. Дата доступа: 05.04.2022.
- 3 25 Google Search Statistics to Bookmark ASAP [Электронный ресурс]. Режим доступа: <https://blog.hubspot.com/marketing/google-search-statistics>. Дата доступа 05.04.2022.
- 4 Zhang L. *Sentiment Analysis and Opinion Mining* / L. Zhang, B. Liu. – Boston : Springer, 2017. – 905 p. DOI: 10.1007/978-1-4899-7687-1_907.
- 5 Search Personalization Using Machine Learning [Электронный ресурс]. Режим доступа: http://faculty.washington.edu/hemay/search_personalization.pdf. Дата доступа 05.04.2022.
- 6 Marketing Artificial Intelligence Institute [Электронный ресурс]. Режим доступа: <https://www.marketingaiinstitute.com/blog/how-search-engines-use-artificial-intelligence>. Дата доступа 05.04.2022.
- 7 Google Mobile Blog. Voice Search arrives in 13 new languages [Электронный ресурс]. Режим доступа: <https://blog.google/products/search/voice-search-arrives-in-13-new-languages>. Дата доступа 05.04.2022.

UDC 004.8

ARTIFICIAL INTELLIGENCE IN SEARCH ENGINES: A REVIEW OF THE CURRENT STATE OF TECHNOLOGY

Luksha B.N., Laptenok N.V.

*Institute of Information Technologies of the Belarusian State University of Informatics and Radioelectronics,
Minsk, Republic of Belarus*

Savenko A.G. – Senior Lecturer, Master of Engineering Sciences

Annotation. The paper presents an overview of the current state of artificial intelligence technologies that are used in search engines for ranking search results, natural language processing in search queries, voice search and image search. The results of studies proving the effectiveness and prospects of using artificial intelligence in search engines are presented.

Keywords. Search algorithms, search query, relevance, machine learning, search engines, artificial intelligence, voice search, search by image, ranking of query results.