

## АЛГОРИТМ PAGE RANK КАК СПОСОБ ПОВЫСИТЬ УЗНАВАЕМОСТЬ БРЕНДА

**Вицуп Сергей Игоревич**

магистрант, Самарский Государственный Экономический Университет, РФ, г. Самара

### PAGE RANK ALGORITHM AS A WAY TO INCREASE BRAND RECOGNIZATION

*Sergey Vitsup*

*Undergraduate, Samara State Economic University, Russia, Samara*

**Аннотация.** Данная статья будет посвящена алгоритму PageRank и анализу его работы. Мы выявим актуальность данного алгоритма и рассмотрим, как он может помочь компаниям повышать узнаваемость в сети Интернет, и, как следствие, увеличивать продажи.

**Abstract.** This article will be devoted to the PageRank algorithm and the analysis of its work. We will identify the relevance of this algorithm and consider how it can help companies increase awareness on the Internet, and, as a result, increase sales.

**Ключевые слова:** Интернет, сайт, алгоритм, PageRank, компания, предприятие, сеть, ранжирование, поисковая система.

**Keywords:** Internet, site, algorithm, PageRank, company, enterprise, network, ranging, search system.

В современные век информационных технологий всемирная сеть Интернет вмещает в себя более 5,3 млрд веб-страниц. С каждым годом этот показатель неустанно растет. Чтобы не запутаться во всемирной паутине, были разработаны специальные алгоритмы компьютерных программ, которые быстро предоставляли пользователям необходимую информацию по запросам – поисковые системы.

Поисковые системы загружают нам информацию по заданным ключевым словам, показывая только качественные веб-страницы и включая их индекс для качественного результата. Ранжирование веб-страниц, возникающих из исходящих запросов, помогает решать данную проблему, показывая первыми в списке наиболее посещаемые веб-страницы или веб-страницы, которые содержатся в каталоге.

В начале XXI века значимое положение во всемирной паутине сети Интернет занимает поисковая система Google, созданная Сергеем Брином и Ларри Пейджем. В сравнении с другими поисковыми системами того времени, Google взял первенство благодаря качественному результату поиска, используя алгоритм PageRank. Данный алгоритм был опубликован впервые в статье «The Anatomy of Search Engine». Итеративный алгоритм PageRank определяет важность веб-страницы исходя из оценки количества гиперссылок на

веб-страницу в предположении, что на «хорошие» и «важные» страницы ссылаются больше, чем на другие.

Работа поисковых систем значительно облегчает поиск информации. Всемирная паутина Интернета содержит в себе информацию, выступая своеобразным информационным хранилищем, где организация поиска необходимых материалов является одновременно, как и легким для пользователя, так и сложным процессом с определённой технической организацией [3].

Организация поиска информации для пользователя сложна тем, что Интернет содержит в себе множества ресурсов. Формирование ответа на интересующий запрос без специализированных технических методов и приемов простому пользователю – непосильная задача.

Сама организация работы поисковой системы проста: получив запрос на поиск информации от пользователя, система подбирает необходимую веб-страницу из индексной базы с релевантной информацией и предоставляет веб-страницы в виде списка.

Можно выделить два принципа работы поисковых алгоритмов: прямого действия и инвертированного индекса.

Алгоритмы работы прямого действия поисковых систем работают по принципу подбора документов, находящихся в индексе поисковой системы. Они требуют огромных технических и временных ресурсов, поэтому все действующие системы отказались от данного метода, либо используют ограничено [1].

Алгоритмы инвертированного индекса или обратного индекса рассматривает каждый документ со стороны индексации, то есть каждый термин состоит из ссылки на содержащий его документ. При таком методе возможно за короткие сроки получить необходимую веб-страницу, где остается только сортировка в списке.

Для сортировки релевантных веб-страниц при алгоритме инвертированного индекса используется ссылочное ранжирование – фактор для формирования рейтинга веб-странице при поисковой выдаче. Оно формируется из понятия: чем больше ссылок с тематическим текстом ссылаются на определенный сайт или страницу, тем выше его место в списке поисковой выдаче [4].

Пользователь, который выдал запрос в поисковую систему ищет информацию во всемирной паутине Интернет и посещает страницу ( $p$ ), что можно рассматривать как случайное блуждание в состоянии ( $p$ ). На каждом этапе пользователь переходит на другие веб-страницы, либо переходит по ссылкам на уже имеющиеся, при этом не возвращаясь на предыдущую и не переходит по одной ссылке дважды. Вероятность случайного перехода на страницу обозначим ( $d$ ), а вероятность перехода по ссылке внутри страницы обозначим ( $1-d$ ). Теперь представим формулу вероятности нахождения пользователя на заданной странице ( $p$ ):

$$R_{j+1}(p) = d + (1-d) \sum_{i=1}^k R_j(p_i) / C(p_i)$$

где  $R(p)$  — PageRank страницы,  $C(p)$  — число ссылок на странице,  $k$  — число ссылающихся на  $p$  страниц,  $d$  - коэффициент демпфирования (damping factor), который варьируется в пределах от 0,1 до 0,15.

Если масштабировать PageRank таким образом, что:

$$\sum_{i=1}^N R(p_i) = 1$$

где  $N$  — число всех страниц, для которых производится расчёт PageRank, то  $R(p)$  можно рассматривать как распределение вероятности по всем страницам.

Для более подробного вычисления алгоритма PageRank составляется матрица  $M$  размером  $(N*N)$ , где каждому элементу  $m_{ij}$  матрицы присваивается значение  $R_0(p)=1/N$  в том случае, если с  $i$ -й страницы имеется ссылка на  $j$ -ую, все оставшиеся элементы матрицы заполняются нулями. Таким образом, вычисление PageRank сводится к поиску собственного вектора матрицы  $M$ , что достигается умножением матрицы  $M$  на вектор  $R_j$  на каждом шагу итерации. Введение коэффициента демпфирования гарантирует, что процесс сходится.

Рассмотрев расчет алгоритма PageRank необходимо подумать о том, как увеличить индекс своей веб-страницы. Чем выше авторитет у заданной веб-страницы с размещенной ссылкой, тем больше увеличивается PageRank веб-страницы, на которую ведет данная ссылка. Данная схема работает и наоборот, чем больше ссылок на веб-странице, тем меньше будет ее показатель PageRank. Таким образом сайты, которые содержат просто наборы ссылок на различные веб-страницы имеют наименьший авторитет в учете показателя PageRank.

Индекс цитирования – это реферативная база данных различных публикаций, которая индексирует ссылки и затем показывает количественный показатель данных ссылок. При расчете алгоритма PageRank не учитывается текстовое содержимое веб-страницы или ссылки, акцентируется только их количество и качественный вес.

Не всегда высокий показатель PageRank для веб-страницы гарантирует первые места в списке выдачи поисковых систем по запросу пользователя. PageRank не является основным критерием ранжирования веб-страниц, а только дополнительным.

Не стоит забывать о том, что PageRank вычисляется не для целого сайта, а только отдельной веб-страницы, поэтому количество ссылок именно на странице очень важно. Необходимо помнить, что ссылки из каталогов веб-страниц необходимо указывать напрямую через CGI-скрипт, иначе они просто не будут учтены.

Поисковые роботы, которые погружают веб-страницы методом поиска в ширину, могут обнаружить только веб-страницы высокого качества на начальных этапах сканирования. В режиме работы поисковой системы Google при сканировании веб-страниц используются системные файлы Sitemap [2], которые предоставляют сами владельцы сайтов или веб-страниц. По мере посещения сайтов поисковый робот сам переходит по ссылкам веб-страниц на другие страницы. Больше внимание робот уделяет непосредственно новым ссылкам или неработающим вовсе. Робот автоматизировано сканирует определенные сайты, через какой период времени это надо делать и какое количество веб-страниц необходимо выбрать на каждом из них.

Большая часть компаний, которая занимается поисковыми системами не разглашают методы и принципы работ их поисковых роботов, рассматривая ее как коммерческую тайну, к которой нет доступа и источников.

На данный момент PageRank является единственным связующим показателем, который ранжирует все страницы сети интернет поисковыми роботами. Существуют и другие выявляющие алгоритмы, например, такой как HITS или алгоритм Клейнберга, который так же показывает качество веб-страницы. Данная метрика зачастую используется для ответа на широкую тему запросов пользователей в сети интернет. Она основывается на предположениях, что гиперссылки кодируют значительное количество скрытых авторитетных страниц. Алгоритм использует для вычисления две оценки: оценку авторитетного документа и оценку хаб-документа (посредника).

Алгоритм Клейнберга разработан для ранжирования результатов запроса в поисковой

системе и учитывает только небольшой набор страниц при вычислении авторитетных оценок.

Говоря о важности данного алгоритма для бизнеса и предприятий нельзя не отметить следующее. Сейчас большое количество компаний имеют свои сайты, на которых содержится вся информация как о самой компании, так и о ее продукте. Для того, чтобы о компании узнало большее количество человек можно пробовать раскрутить ее не только при помощи рекламы. Можно пробовать писать интересные статьи, заметки или публикации на сайте, которые будут содержать информацию о продукте и его полезности или, например, о процессе его производства. Все это приведет к тому, что потенциальный клиент сможет сам случайно попасть на сайт благодаря алгоритму PageRank и интересной статье. Ведь, по сути, написание интересного материала о товаре, который производит компания, не должно вызывать какую-то сложность. Скорее наоборот – это должно быть очень легко, так как высококлассные менеджеры должны уметь продавать товар, а значит и уметь красиво рассказать о нем, его свойствах и отличительных чертах.

### **Список литературы:**

1. Search Engine Watch: Up-to-date information on leading search engines. [Электронный ресурс] URL: <http://www.searchenginewatch.com/>.
2. The Google Search Engine: Commercial search engine founded by the originators of PageRank. [Электронный ресурс] URL: <http://www.google.com/>
3. K. Bharat and A. Broder. A technique for measuring the relative size and overlap of public web search engines. In Proceedings of the Seventh International World Wide Web Conference, 2017.
4. Ландэ, Д. В. Поиск знаний в Internet / Д. В. Ландэ. - М. : «Диалектика», 2016. - 272 с.
5. Грошев А. С. Информатика: Учебник для вузов / А. С. Грошев . Архангельск. Арханг. гос. техн. ун-т, 2021. - 470 с
6. Советов Б. Я. Информационные технологии: Учеб. Для вузов. - М.: Высш. шк., 2021. - 157 с.