

МОДИФИЦИРОВАННЫЙ МЕТОД РАНЖИРОВАНИЯ НА ОСНОВЕ АЛГОРИТМА PAGE RANK

А. О. Сытежев, Л. В. Борисова

*Саратовский национальный исследовательский
государственный университет им. Н. Г. Чернышевского, Россия*
E-mail: Sytehzev.a@mail.ru, lvborisova27@gmail.com

В данной статье представлена модифицированная формула ранжирования, которая основана на логике алгоритма Page Rank. Продемонстрирован принцип ее работы.

MODIFIED RANKING METHOD BASED ON PAGE RANK ALGORITHM

A. O. Sytezhev, L. V. Borisova

This article presents a modified ranking formula based on the logic of the Page Rank algorithm. The principle of its operation is demonstrated.

В данной статье рассмотрен алгоритм ранжирования элементов, основанный на схеме Page Rank, получена модифицированная формула ранжирования, применимая к прикладным задачам. Потребность к сортировке элементов возникла очень давно. Она касалась различных сфер деятельности человека, затрагивая самые разнообразные профессии и отрасли. Наибольший скачок в развитии алгоритмов по сортировке элементов пришелся на момент развития информационных технологий и сети интернет. В 1996 году Сергей Брин и Ларри Пейдж начали работу над исследовательским проектом, который решал вопрос о необходимости сортировки интернет-страниц по степени важности этих страниц [1]. BackRub — поисковая система по Интернету, использующая новую идею о том, что веб-страница должна считаться тем «важнее», чем больше на неё ссылаются других страниц, и чем более «важными», в свою очередь, являются эти страницы. Данная система опиралась на алгоритм Page Rank, что давало наилучший эффект, по сравнению с конкурентами. Более подробно логика работы алгоритма представлена в [2].

Введем обозначения: d - коэффициент затухания, который может быть установлен в диапазоне от 0 до 1, PR - Page Rank страницы, $PR(T_i)$ - Page Rank i -й страницы, с которой осуществляется переход на рассматриваемую страницу, $C(T_i)$ - количество ссылок на другие страницы с i -й страницы.

Тогда алгоритм по нахождению веса страницы Page Rank определяется формулой [3]:

$$PR = (1 - d) + d \sum_{i=1}^n \frac{PR(T_i)}{C(T_i)} \quad (1)$$

Для подсчета веса страницы в этой формуле учитывается ряд факторов, таких как веса сторонних страниц, а также количество ссылок с данной страни-

цы на остальные, по которым определяется значимость конкретной страницы. Данный подход широко применялся в различных сферах, например, уже в 1998 году Google был одной из первых поисковых систем, внедривших данный алгоритм. В настоящее время система Page Rank несколько раз видоизменялась, создано много различных модификаций этого алгоритма [4]. Рассмотрим задачу конкурса научных статей. На конкурс приходит некоторое число n статей. Эти работы рассматриваются m судьями различной категории, которая оценивается рядом факторов, например, квалификация и опыт работы. Предположим, у конкурса тематика "Информационные технологии", в ходе которой работы, связанные с наиболее востребованными темами безопасности и защиты информации, могут получить грант, в силу чего данные темы находятся в приоритете. Применяя рассуждения алгоритма (1) Page Rank, необходимо определить значимость данных научных статей, в результате чего формируется порядок занесения статей в сборник.

Заметим, что в этом случае нет зависимости работ друг от друга, в силу чего коэффициент затухания d , применяющийся в Page Rank, можно убрать, т.е. $d=1$, но при этом совокупное количество мест, отданных данной статье, является решающим показателем в определении ее значимости, в силу чего сумма отданных голосов становится решающим фактором. В этом случае повлиять на значимость статьи также может категория судьи, распределяющего статьи, и актуальность темы, выбранной участником, таким образом, справедлива формула:

$$P_l = \left(\sum_{j=1}^n \sum_{i=1}^m x_{ji} \beta \alpha_i \right) + p_j \quad (2)$$

где x_{ji} - критерий ранжирования, β - коэффициент, а α – параметр, определяющий вес судьи. Параметр p_j представляет собой показатель важности сторонних критериев, влияющих на результат ранжирования.

Формула (2) обеспечивает сортировку данных по степени важности, тем самым делая возможным применение ее в различных сферах деятельности.

Исходные данные задачи сортировки статей можно представить в виде таблицы.

Исходные данные задачи сортировки статей

Статья (значимость) / места								
Судья(вес)	№1(p1)	№2(p2)	№3(p3)	№4(p4)	№5(p5)	№6(p6)	№7(p7)	№8(p8)
1(вес1)	X11	X12	X13	X14	X15	X16	X17	X18
2(вес2)	X21	X22	X23	X24	X25	X26	X27	X28
3(вес3)	X31	X32	X33	X34	X35	X36	X37	X38
4(вес4)	X41	X42	X43	X44	X45	X46	X47	X48
5(вес)	X51	X52	X53	X54	X55	X56	X57	X58

Из результатов судейства видно, что место статьи зависит от количества мест, отданных каждым судьей, при этом учитывается значимость судьи, отдающего голос, параметр α_i определяет вес судьи. Тогда данную зависимость отражает слагаемое $\sum_{j=1}^n \sum_{i=1}^m x_{ji} \beta \alpha_i$. Коэффициент β указывает на то, совпадает ли место, отданное судьей с местом $L(L_1 L_2 L_3 L_4 L_5 L_6 L_7 L_8)$ по которому мы ве-

дем отбор.

$$\begin{cases} \beta = 1, & x_{ji} = l \\ \beta = 0, & x_{ji} \neq l \end{cases}$$

За учет актуальности самой темы работы отвечает слагаемое p_j .

В итоге формула:

$$P_l = \left(\sum_{j=1}^n \sum_{i=1}^m x_{ji} \beta \alpha_i \right) + p_j$$

позволяет, задавая значение места L , по которому мы хотим распределить статьи, получить ранжированные списки.

Заметим, что формула (2) актуальна и для задач, где необходимо не только распределить все имеющиеся варианты, но и провести отбор нескольких вариантов. Например, при устройстве на работу каждый сотрудник заполняет анкету с навыками и умениями, которыми он обладает. Для работодателя ряд определенных умений играет более важную роль по сравнению со всеми остальными, тем самым, у каждого навыка появляется свой вес α_j . Этот показатель и будет решающим в определении востребованности сотрудника для компании. По совокупности умений и их весов будет определяться ранжированный список сотрудников, а по нему могут быть выделены самые востребованные. За коэффициент p_i будут отвечать сторонние факторы, например, показатели, не влияющие на умения сотрудника, но необходимые с другой точки зрения, например, пол, возраст, предыдущее место работы, рекомендации.

Таким образом, можно сделать вывод о том, что рассмотренный способ ранжирования актуален во многих сферах деятельности человека и применим ко многим прикладным задачам.

СПИСОК ЛИТЕРАТУРЫ

1. *Sergey Brin, Lawrence Page* The Anatomy of a Large-Scale Hypertextual Web Search Engine // Computer Science Department, Stanford University.
2. *Chris Ridings* PageRank Explained or Everything you've always wanted to know about PageRank. 2001.
3. *Тимонина А. В.* Модель задачи ранжирования и ее исследование // МФТИ (ГУ), 2009. Т. 5. С. 139-156.
4. Google PageRank жив: почему он всё ещё важен. [Электронный ресурс]. URL: <https://ahrefs.com/blog/ru/google-pagerank/> (дата обращения: 01.10.2021).