

УДК 004.9

*О. Ю. Высокая, О. С. Машкова**

ПОДХОД К РЕАЛИЗАЦИИ ГЕОЗАВИСИМОГО ПОИСКА ПО ОРГАНИЗАЦИЯМ С ИСПОЛЬЗОВАНИЕМ РЕЙТИНГА БЛАГОНАДЕЖНОСТИ

Среди всех запросов, выполняемых к поисковым системам, от 15 до 30%, в зависимости от региона, составляют те, в ответ на которые пользователь ожидает получить местную, региональную информацию – например, об услугах в своем городе [1]. Такие запросы носят название геодеpendимых. В ответ на геодеpendимые запросы поисковая система показывает разные результаты поиска для разных регионов. Наиболее релевантные ответы находятся, как правило, на местных, региональных сайтах. Но это не значит, что по геодеpendимому запросу нельзя найти авторитетный общероссийский сайт или сайт, расположенный в другом регионе. Речь только о приоритете для локальных результатов при прочих равных.

Принадлежность сайта к тому или иному региону определяется по многим признакам – в том числе по указанным на нем контактам, региону, которому посвящена большая часть информации на ресурсе, и т.д. [1]. С одной стороны, указанная особенность формирования поисковой выдачи повсеместно используется при выполнении SEO-оптимизации своих веб-ресурсов теми компаниями, чьи офисы не расположены в заданном для поиска регионе, но которые готовы расширять зону оказания своих услуг на соседние области. С другой стороны, далеко не все местные организации имеют веб-ресурсы, которые будут либо вообще корректно проиндексированы поисковым роботом, либо после индексации будет корректно определен их регион. В итоге в результатах поисковой выдачи более высокие места часто занимают крупные компании из соседних регионов.

Для решения описанной проблемы предлагается формирование каталога организаций, предоставляющих свои услуги в конкретном регионе, с организацией локального поиска по ним. В рамках алгорит-

* Работа выполнена под руководством старших преподавателей ФГБОУ ВО «ТГТУ» М. Ю. Авдеевой и А. Н. Кагдина.

ма ранжирования результатов поиска по компаниям в качестве одного из параметров предлагается использовать рейтинг благонадежности организации из каталога.

Идея, связанная с принципом определения рейтинга благонадежности организации, заключается в следующем: о благонадежности организаций следует судить по количеству положительных отзывов, данных другими организациями, а также по благонадежности самих организаций. Причем определение тональности отзывов не сложно автоматизируется за счет применения алгоритмов машинного обучения.

Назначим каждой организации O показатель ее благонадежности $T(O)$. Предположим, что организация O_j дала l_j положительных отзывов о других организациях, с которыми она взаимодействовала в рамках, например, договорных отношений. Если один из этих отзывов относится к организации O_i , то O_j передаст $1/l_j$ своей благонадежности O_i . Рейтинг благонадежности организации O_i является суммой всех вкладов, сделанных отзывами на нее. То есть, если мы обозначим множество организаций, давших отзывы на организацию O_i через C_i , то:

$$T(O_i) = \sum_{O_j \in C_i} \frac{T(O_j)}{l_j}. \quad (1)$$

Очевидно, что для вычисления рейтинга благонадежности организации O_i сначала нужно знать благонадежность всех организаций, давших на нее отзывы.

Преобразуем проблему в математическую задачу. Создадим матрицу $\mathbf{H} = [H_{ij}]$, в которой запись в i -й строке и j -м столбце определяется следующим выражением:

$$H_{ij} = \begin{cases} 1/l_j, & \text{если } O_j \in C_i; \\ 0, & \text{в противном случае.} \end{cases}$$

Матрица \mathbf{H} имеет некоторые особые свойства. Во-первых, все записи неотрицательны. Кроме того, сумма записей в столбце равна единице, если организация, соответствующая этому столбцу, не давала

отзывов. Так как все записи неотрицательны и сумма записей в каждом столбце равна единице, матрица является стохастической.

Сформируем вектор $T = [T(O_i)]$, чьи компоненты есть рейтинги благонадежности организаций. Условие выше, определяющее рейтинг благонадежности, может быть выражено как:

$$T = \mathbf{HT}.$$

Другими словами, вектор T является собственным вектором матрицы \mathbf{H} (или стационарным).

Рассмотрим пример. Ниже показано представление небольшого набора организаций с положительными отзывами на другие компании, представленными стрелками (рис. 1).

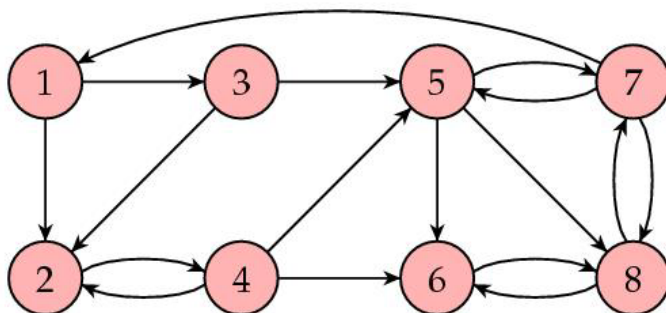


Рис. 1. Граф отзывов организаций

Соответствующая матрица:

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix}$$

со стационарным вектором

$$T = \begin{bmatrix} 0,0600 \\ 0,0675 \\ 0,0300 \\ 0,0675 \\ 0,0975 \\ 0,2025 \\ 0,1800 \\ 0,2950 \end{bmatrix}.$$

Этот вектор показывает, что организация 8 побеждает в рейтинге благонадежности.

Существует несколько способов поиска собственных векторов квадратной матрицы. Можно использовать метод, известный как степенной, для нахождения стационарного вектора T матрицы \mathbf{H} [2]. Процесс начинается с выбора вектора T_0 в качестве кандидата на T , а затем производится последовательность векторов T^k с помощью выражения:

$$T^{k+1} = \mathbf{H}T^k.$$

Общий принцип: последовательность T^k будет сходиться к стационарному вектору T [2]. Значения получаемого вектора T^k являются относительными показателями для сравнения благонадежности организаций с помощью таких утверждений, как «организация A вдвое благонадежнее, чем организация B ».

Теперь поскольку для каждой организации из каталога определен ее рейтинг благонадежности, предлагается осуществить формирование рейтинга благонадежности сотрудников, которые работали в некоторой организации из каталога, а теперь находятся в поисках работы. Введение персонального рейтинга благонадежности совместно с рейтингом благонадежности организаций позволит сформировать бизнес-среду вокруг предлагаемого в статье сервиса.

Назначим сотруднику E показатель его благонадежности $T(E)$. Предположим, что организация O_j дала l_j положительных отзывов о разных сотрудниках. Если один из этих отзывов относится к сотруднику E_i , то O_j передаст $1/l_j$ своей благонадежности E_i . Рейтинг благонадежности сотрудника E_i является суммой всех вкладов,

сделанных отзывами на него, т.е., если мы обозначим множество организаций, давших отзывы на сотрудника E_i через C_i , то:

$$T(E_i) = \sum_{O_j \in C_i} \frac{T(O_j)}{I_j}. \quad (2)$$

Очевидно, что вычисление рейтинга благонадежности сотрудников не представляет сложности, поскольку рейтинг благонадежности организаций был посчитан ранее.

Список литературы

1. Компания Яндекс – Технологии – Поиск с учетом региона [Электронный ресурс]. – Режим доступа : <https://yandex.ru/company/technologies/regions> (дата обращения 20.09.2019).

2. Клиот-Дашинский, М. И. Алгебра матриц и векторов : учебник для вузов / М. И. Клиот-Дашинский. – 2-е изд. – СПб. : Лань, 1998. – 160 с.

Кафедра «Электроэнергетика» ФГБОУ ВО «ТГТУ»