

Р. В. Огородник
(БГУИР, Минск)
**МЕТОДЫ РАНЖИРОВАНИЯ
ПОИСКОВОЙ ВЫДАЧИ В СИСТЕМАХ
ПОЛНОТЕКСТОВОГО ПОИСКА**

В качестве проблем современных поисковых систем выделяют в отдельную категорию проблему полноты поиска и релевантности страницы поисковой выдачи. Поисковая выдача (страница результатов поиска) – страница, генерируемая системой в ответ на поисковый запрос пользователя. К наиболее известным методам ранжирования поисковых результатов относится семейство методов, основанное на присваивании весов полям, по которым производится поиск в зависимости от их важности. В настоящей работе используется этот метод, как наиболее удачный с некоторыми дополнениями, позволяющими увеличить релевантность поискового запроса к результатам поиска, определить не только полноту вхождения всех слов поискового запроса в статью, но и их близость.

В ходе проведенной работы для улучшения качества результатов поисковых систем и обеспечения наибольшей полноты выдачи информации из больших хранилищ данных, как, например, новостной портал, были выработаны следующие стратегии и дополнения к существующему основному методу:

- 1) задание доминирующих весов для ключевых полей, по которым ведется поиск;
- 2) настройка сети поисковых слов: если пользователь обращается с поисковым запросом, то во время его сессии хранится вся история его запросов, которые связываются в сеть;
- 3) помощь пользователей поиска: если по поисковому запросу найдена статья и дизайном предусмотрено наличие кнопки «Это то, что я искал», то статье по данному поисковому запросу назначается дополнительный вес;
- 4) нейронная сеть, классифицирующая поисковые запросы по категориям и соотносящая их с категориями статей, по которым ведется поиск, действующая в роли вспомогательной в случае омонимичных запросов или же запросов, в которых присутствуют многозначные слова.

В случае нулевого количества найденных элементов (или в случае, если количество найденных элементов меньше порогового значения) будет выполняться предложение других поисковых запросов.

Литература

1. Аксенов, А. Sphinx overview [Электронный ресурс]. – Электронные данные. – Режим доступа: <http://sphinxsearch.com>
2. Hatcher, E. Lucene in Action / E Hatcher, O. Gospodnetić. – London, Marmek, 2004. – 456 p.
3. Snowball [Электронный ресурс]. – 2011. – Режим доступа <http://snowball.tartarus.org/algorithms/russian/stemmer.html>
Дата доступа: 02.05.11