

Эффективность расширения запросов поисковых систем с помощью специальных операторов языка запросов

© Игорь Кураленок, Иван Блеканов, Данила Дюгуров, Леонид Еремин, Игорь Михайлюк, Василий Писарь, Константин Федоров

СПбГУ, ПМ-ПУ
ik@intellij.com

Аннотация

В данной статье рассматривается влияние использования специальных операторов языков запросов поисковых систем на результаты поиска. Были выбраны пятьдесят запросов к поисковым системам Google, Yandex, Aport и Rambler, каждый из запросов был расширен с помощью специальных операторов языков запросов, и было проведено сравнение результатов, возвращенных вышеупомянутыми поисковыми системами.

1. Введение

Не секрет, что каждый автор имеет свой стиль изложения информации, и даже классические методы поиска, такие, как tf-idf, булевский поиск и другие, не обеспечивают достаточно хорошего результата поиска. Для устранения данной проблемы лидеры мирового рынка поисковых машин (www.yandex.ru, www.google.com, www.rambler.ru и aport.ru) применили так называемый «поиск по фразе» (phrase querying) и расширение языка запросов дополнительными специальными операторами. Однако, несмотря на эти улучшения, большинство пользователей не использует их при составлении запросов, обходясь поиском по ключевым словам: по данным анализа логов различных поисковых систем, запросы со специальными операторами составляют от пяти до двадцати пяти процентов. Лишь в единичных случаях этот процент больше – например, при использовании узкоспециализированных поисковых систем ([2], [8]).

При поиске информации в Интернете с помощью различных поисковых систем пользователь часто сталкивается с тем, что его не удовлетворяет ответ поисковой системы. В таком случае пользователю приходится переформулировать свой запрос, пытаясь достичь ответа, более соответствующего информационной потребности пользователя. Перед пользователем встает вопрос: стоит ли пытаться изменять свой запрос специальными операторами? В этой ситуации перед пользователем стоит выбор: попытаться переформулировать свой запрос,

используя специальные операторы, определенные в спецификации языка данной поисковой системы, или полностью изменить свой запрос – то есть придумать новый, что, конечно, сложнее ([3], [6]).

Если оценить влияние расширенных запросов на ответы поисковой системы, то можно понять, имеет ли смысл изменять запрос, используя специальные операторы.

2. Эксперимент

В эксперименте оценивались первые десять ответов каждой системы (как наиболее интересные с точки зрения пользователя).

Для эксперимента были выбраны четыре популярные русскоязычные поисковые системы: Google, Yandex, Rambler и Aport. Для сравнения этих систем было составлено пересечение словарей специальных операторов этих систем. Также для эксперимента из запросов РОМИП-2004 были выбраны 50 запросов. Каждый запрос был «расширен» с помощью специальных операторов [5]: каждый исходный запрос был дополнен специальными операторами; при этом на расширенный запрос налагались следующие ограничения: 1) расширенный запрос не должен менять порядок слов исходного запроса; 2) расширенный запрос не должен менять словоформы исходного запроса. Например, запрос «мерседес из германии» (без кавычек) дал четыре запроса: «мерседес из германии», мерседес "из германии", мерседес AND "из германии", мерседес + "из германии". Для каждого исходного запроса было получено несколько расширенных запросов – от одного до семи.

Для каждого исходного запроса и для соответствующих ему расширенных запросов был получен ответ системы (для каждой поисковой системы эксперимент был проведен отдельно). Далее находилось число совпадений между ответами на базовый запрос и ответами на расширенный запрос (то есть, количество совпадающих ссылок). Если для базового запроса было несколько версий расширенного запроса, то находилось число совпадений для каждого расширенного запроса, которое затем усреднялось.

Для каждого базового запроса было получено усредненное число совпадений расширенных запросов с базовым. Для каждой поисковой системы эти числа так же были усреднены.

В конце концов, для каждой поисковой системы было получено среднее число совпадений ответов на базовый и расширенный запросы, которое, умноженное на 10(так как всего рассматривалось 10 ответов системы), дает процент совпадающих ответов.

3. Итоги эксперимента

Эксперимент дал следующие результаты:

процент совпадений результатов для исходного запроса и для расширенных версий в среднем составил: для aport.ru – 75.7%; для rambler.ru – 61.7%; для google.ru – 61.2%; для yandex.ru – 47.4%

4. Итоги эксперимента

В отличие от других исследований, в данном эксперименте не ставилась цель сравнить между собой влияние различных специальных операторов на ответ поисковой системы (как это делалось, например, в [4], [7]). Наоборот, целью исследования было выяснить, насколько целесообразно для пользователя пытаться расширить свой запрос, используя специальные операторы языков запросов.

Также целью данного эксперимента был анализ вышеупомянутой проблемы на основе достаточно большого набора данных – 50 базовых запросов и около 250 расширенных.

Наибольший процент совпадений у поисковой системы Апорт – 75.7% – следовательно, при ее использовании добавление специальных операторов изменит лишь два-три ответа системы из десяти. Лучшее всего – с точки зрения статистики – на изменение запроса реагирует Yandex, давая в среднем больше половины новых ответов. В итоге, наиболее выгодный для пользователя результат даст модернизация исходного запроса в Yandex'e, наименее – в Апорте. Полученные результаты косвенно подтверждаются результатами, приведенными в [1]: хороших результатов можно добиться, и применяя специальные операторы языка запросов, и обходясь без них.

В дальнейшем планируется расширить эксперимент, учитывая не только «голый» ответ системы, но и релевантность каждого возвращенного системой документа по отношению к запросу, а также изменение положения документа в первой «десятке» для расширенного и исходного запросов. Также хотелось бы провести обратный эксперимент, взяв реальные запросы, использующие специальные операторы, и сравнить ответы поисковых систем на эти запросы и на эти же запросы без специальных операторов.

Литература

- [1] Добров Б.В., Добров Г.В., Лукашевич Н.В., Резников Я., Штернов С.В. – Исследование методов трансформации запросов в первом туре Кубка Яндекса.
http://company.yandex.ru/grant/2005/02_Loukachevitch_103039.pdf
- [2] Broder A. A Taxonomy of Web Search. SIGIR Forum 36(2), 2002
- [3] C. Buckley, G. Salton, J. Allan, and A. Singhal. Automatic query expansion using SMART: TREC 3. In NIST Special Publication 500-225: The Third Text REtrieval Conference (TREC-3), pages 69{80. Department of Commerce, National Institute of Standards and Technology, 1995.
- [4] Eastman C.M., Jansen B.J., Coverage, Relevance, and Ranking: The Impact of Query Operators on Web Search Engine Results – ACM Transactions on Information Systems, V.21, No.4, 2003, p.383-411
- [5] Help Pages -- Google Help Center
<http://www.google.com/support/bin/static.py?page=searchguides.html&ctx=advanced>
Апорт – Справочник по языку поисковых запросов. <http://www.afort.ru/help/?p=3>.
Yandex – Язык поисковых запросов. <http://help.yandex.ru/search/?id=481939>.
Rambler – Язык поисковых запросов. <http://www.rambler.ru/doc/help.shtml>
- [6] Christoph Hölscher, Gerhard Strube. Web Search Behavior of Internet Experts and Newbies. Proceedings of WWW9. 2000.
<http://www9.org/w9cdrom/81/81.html>
- [7] Jansen, Bernard J. "The effect of query complexity on Web searching results". Information Research, 6(1).
<http://InformationR.net/ir/6-1/paper87.html>
- [8] Rose D.E., Levinson D. Understanding User Goals in Web Search // Proceedings of the 13th World Wide Web Conference. – New York, 2003, p. 13 – 19

Efficiency of search engines' queries extension using query language operators

Igor Kuralenok, Ivan Blekanov, Daniil Dyugurov,
Leonid Eremin, Igor Mihailyuk, Vasiliy Pizar,
Konstantin Fedorov

This paper examines the influence of using query language operators on search results. There were chosen 50 queries to Yandex, Rambler, Aport and Google, each was extended with query language operators and acquired results were compared.