

Солдатова Г. Т., Белинский П. Д., Шадрин М. Е.

ПРОБЛЕМА ПОВЫШЕНИЯ КАЧЕСТВА ПОИСКОВЫХ ЗАПРОСОВ

Гульнара Тагировна Солдатова

кандидат педагогических наук, доцент

gulnara.soldatova@rsyvu.ru

ФГАОУ ВО «Российский государственный профессионально-педагогический

университет», Россия, Екатеринбург

Павел Дмитриевич Белинский

студент Кафедры информационных систем и технологий,

antichristymer@gmail.com

ФГАОУ ВО «Российский государственный профессионально-педагогический

университет» Россия, Екатеринбург

Максим Евгеньевич Шадрин

студент Кафедры информационных систем и технологий,

maximustir777@gmail.com

ФГАОУ ВО «Российский государственный профессионально-педагогический

университет» Россия, Екатеринбург

THE PROBLEM OF QUALITY IMPROVEMENT SEARCH QUERIES

Gulnara Tagirovna Soldatova

Candidate of Pedagogical Sciences, Associate Professor

Russian State Vocational Pedagogical University, Russia, Yekaterinburg,

Pavel Dmitrievich Belinsky

student of the Department of Information Systems and Technologies,

Russian State Vocational Pedagogical University, Russia, Yekaterinburg,

Maxim Evgenievich Shadrin

student of the Department of Information Systems and Technologies,

Russian State Vocational Pedagogical University, Russia, Yekaterinburg,

Аннотация: В статье представлены результаты исследования качества поисковых запросов студентов. Анализ полученных данных показал необходимость оптимизации поисковых запросов. В рамках работы были обработаны и реструктурированы алгоритмы построения запросов в различных поисковых системах, что привело к уменьшению времени составления корректного поискового запроса и увеличению качества получаемой пользователем информации. При составлении указанных алгоритмов применялись орграфы и алгоритм Дейкстры.

Ключевые слова: Поисковый запрос, орграф, опрос, исследование, алгоритм Дейкстры, пользователь.

Abstract: The article presents the results of a study of the quality of students' search queries. Analysis of the data obtained showed the need to optimize search queries. As part of the work, the algorithms for constructing queries in various search engines were processed and restructured, which led to a reduction in the time for compiling a correct search query and an increase in the quality of information received by the user. Digraphs and Dijkstra's algorithm were used in the compilation of these algorithms.

Keywords: Search query, orgraph, poll, research, Dijkstra's algorithm, user.

В настоящее время развивающихся систем программирования и создания различных приложений, пользователи используют различные поисковые сервисы. Разработчики стараются периодически обновлять, оптимизировать свои программные продукты. В частности, совершенствование работы поисковых систем связывают с увеличением числа оказываемых услуг, расширением секторов поисковых запросов, а также разработкой приложений, создающих определенный уровень комфорта для пользователя.

Цель данной работы — проанализировать запросы пользователей и создать плагин, встраиваемый непосредственно в поисковую систему и облегчающий поиск научной и профессиональной информации.

В рамках исследования были произведены наблюдения и проанализировано среднее время, которое среднестатистический пользователь проводит за ПК, конструируя и отправляя различные формы поисковых запросов. В процессе обработки результатов были обнаружены основные «паразитные» временные потери. Разработанный нами плагин позволяет решить проблемы низкого КПД запросов пользователей и сократить временные издержки на поиск информации.

В работе определены следующие задачи:

- провести опрос студентов с целью определения удовлетворенности качеством получаемой информации в поисковых системах;
- проанализировать полученные в ходе опроса данные;
- изучить возможность оптимизации поисковых запросов с точки зрения теории графов;
- разработать плагин, позволяющий оптимизировать поиск информации.

Определимся с основными понятиями, используемыми в данном исследовании.

Под информационным поиском будем понимать поиск неструктурированной информации в различных источниках согласно информационным потребностям пользователя [1; 3].

Данный процесс имеет следующую операционную последовательность: сбор, обработка, предоставление информации.

Запрос — способ выражения пользователем своих потребностей в информации с помощью поисковых систем.

Пользователь может задавать поисковые запросы различных форматов, используя как привычные текстовые запросы, так и изображения в виде специализированных файлов, музыку и любые другие форматы мультимедийной информации.

Выделяют следующие основные виды поиска: полнотекстовый поиск (поиск по всему текстовому содержанию источника), поиск по метаданным

(поиск по свойствам документа, содержащимся в системе — название документа, дата его создания, размерность, авторы и т.д.), поиск изображений.

Полнотекстовый поиск осуществляют, например, такие поисковые системы, как Яндекс, Google и т.п. Для ускорения нахождения информации полнотекстовый поиск использует изначально созданные индексы. Одной из самых распространённых технологий для индексов реализуемых полнотекстовым поиском можно выделить такой, как инвертированный индекс.

Аналогом процесса нахождения по метаданным является диалог поиска в файловой системе, такой как, MS Windows.

В качестве примеров систем, осуществляющих поиск по изображениям, можно привести Polar Rose, Piollator и т.п.

В процессе данной работы были использованы следующие понятия теории графов и алгоритмы, применяемые для решения задачи оптимизации пользовательских запросов.

Граф — схема, состоящая из точек (вершины графа) и связывающих эти точки попарно отрезков и дуг (рёбра графа).

Графом может быть описана схема поискового запроса, в которой вершины графа — этапы работ, а рёбра показывают последовательность исполнения этих работ. Используя данный граф, можно рассчитать время непосредственного осуществления проекта, а также найти самый короткий путь осуществления этих работ.

Для реализации задач, связанных с информационным поиском, в исследовании используются графы, все рёбра которых ориентированы. Такие графы называются ориентированными графами или орграфами.

При создании схематичного изображения системы запросов пользователей нами использовался взвешенный граф, т.е. граф, в котором каждому ребру сопоставляется число — вес ребра.

В задаче оптимизации поисковых запросов требуется построить маршрут от одной вершины к другой, учитывая различные ограничения.

Уточним понятия «маршрут», «цепь», «путь» и «цикл».

Маршрут — последовательность вершин и рёбер графа, следуя по которым, можно попасть из одной вершины в другую.

Цепь — маршрут без повторяющихся рёбер.

Путь — цепь, все вершины в которой различны (за исключением, быть может, начальной и конечной вершин).

Цикл — цепь, в которой совпадают начальная и конечная вершины, а все остальные вершины различны.

В данном исследовании решалась задача нахождения кратчайшего пути в графе, отражающим реализацию поиска нужной информации. Под кратчайшим путем понимается путь, суммарный вес которого минимален.

В рамках анализа и проработки проблемы был выбран алгоритм Дейкстры [2,4].

Алгоритм Дейкстры — мощный инструмент в задачах поиска кратчайшего пути от одной вершины графа до всех остальных.

Для корректной отработки заданного алгоритма необходимо, чтобы исследуемый граф не имел рёбер, отрицательного веса и петель.

Доказательство необходимости оптимизации и реорганизации поисковых систем осуществлялось посредством социального опроса. В работе изучены результаты тестирования 96 человек (студентов 1–4 курса очной формы обучения различных направлений Российского государственного профессионально-педагогического университета).

Результаты опроса приведены в виде диаграмм (рис. 1–6).

Большинство опрошенных оценило свои навыки использования сети интернет и составления в ней поисковых запросов на уровень выше среднего (рис. 1). Данный показатель студенты оценивали по 10-балльной шкале, в которой 1 балл соответствовал самому низкому уровню.

Как вы можете оценить свои навыки в составлении поисковых запросов?

96 ответов

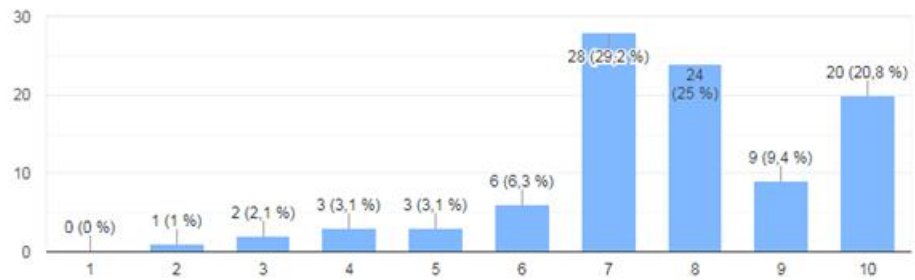


Рисунок 1 — Анализ респондентами своих навыков

Относительно тематики запросов, студенты составляют как научные запросы, так и запросы развлекательного характера (рис. 2).

Какие запросы вы делаете чаще всего?

95 ответов

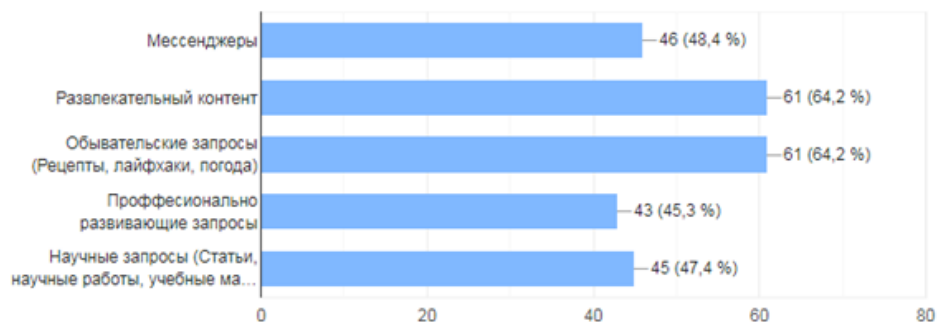


Рисунок 2 — Анализ предпочтительных категорий запросов среди респондентов

Что касается использования специальных символов, то в подавляющем большинстве случаев, студенты отказываются от их использования либо пользуются ими крайне редко (рис. 3)

Пользуетесь ли вы специальными символами и поисковыми операторами при работе с запросом информации в браузере?

96 ответов

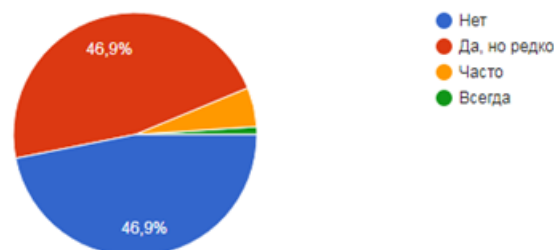


Рисунок 3 — Анализ среди опрошенных частоты использования специальных символов

Далее, на рисунках 4, 5, 6 студентами проанализированы собственные навыки в составлении поисковых запросов — стабильность и полнота получаемой информации, а также оценка скорости получаемой информации. В большинстве случаев, опрашиваемые все же находят нужную им информацию, но часто возникает потребность в перефразировании запросов. Стоит отметить, что присутствует и вероятность получения бесполезной информации, что может замедлить поиск необходимой. Данные показатели студенты оценивали по 5-балльной шкале, в которой 1 балл соответствовал наименьшему значению.

Как часто у вас возникала потребность перефразирования запроса для получения более полной информации?

96 ответов

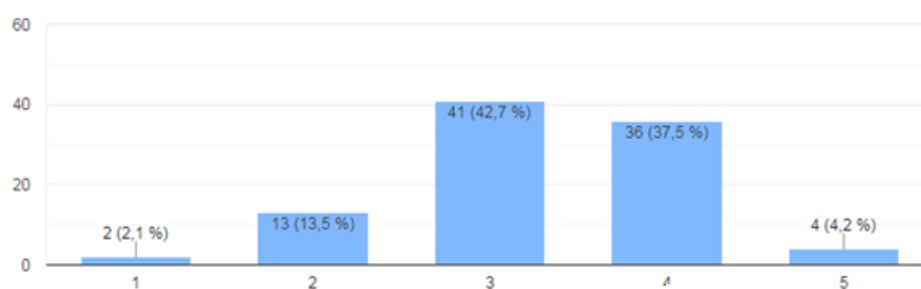


Рисунок 4 — Анализ навыка составления корректных запросов

Как много ненужной информации вы находили при поиске нужной?

96 ответов

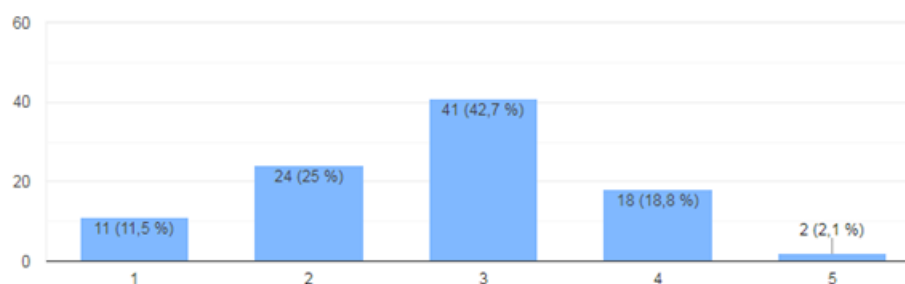


Рисунок 5 — Анализ корректности алгоритмов подбора информации интерфейсом браузерной системы запросов

В среднем, насколько быстро происходит поиск нужной вам информации?
96 ответов

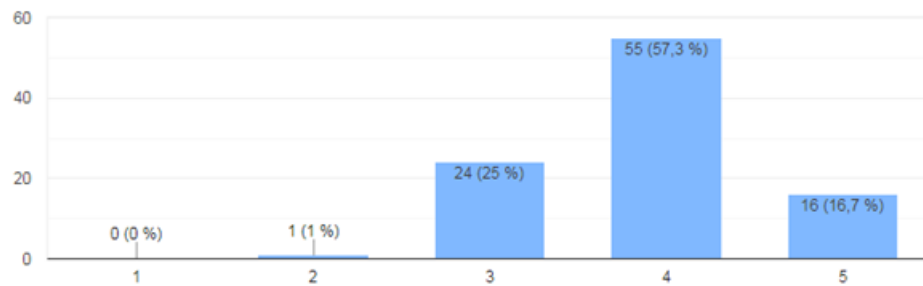


Рисунок 6 — Анализ скорости поиска информации

Исходя из проведенного анализа результатов опроса, можно сделать вывод: пользователи тратят много времени на поиск полноценной корректной информации, не замечая паразитных потерь, не пытаются решить данную проблему или увеличить свой навык пользования поисковыми системами. Однако данная проблема актуальна и требует внимания.

В ходе изучения указанной проблемы был разработан плагин, внедряемый напрямую в систему поиска и необходимый для облегчения поиска информации.

Анализируя поиск с помощью графов, демонстрирующих путь исполнения пользовательского запроса, без плагина и с его применением, была выявлена значительная разница в скорости получения пользователем необходимой информации. Анализ осуществлялся с помощью алгоритма Дейкстры на полученных графах.

Таким образом, в ходе исследования была выявлена проблема необходимости оптимизации поисковых запросов, а также рассмотрена возможность решения данной проблемы с помощью применения разработанного плагина.

Список литературы

1. *Глоссарий* по информационному обществу / под общ. ред. Ю. Е. Хохлова. Москва: Ин-т развития информ. общества, 2009. 160 с. Текст: непосредственный.

2. *Коньшева, Л. К.* Дискретная математика / Л. К. Коньшева. Екатеринбург: Изд-во Рос. гос. проф.-пед. ун-та, 2010. 205 с. Текст: непосредственный.

3. *Минькович, Т. В.* Информационные технологии: понятийно-терминологический аспект / Т. В. Минькович. Текст: электронный // Образовательные технологии и общество. 2012. Т. 15, № 2. С. 371–389. URL: http://ifets.ieee.org/russian/depository/v15_i2/pdf/1.pdf (дата обращения: 05.11.2022).

4. *Харари, Ф.* Теория графов / Ф. Харари; под ред. Г. П. Гаврилова; пер. с англ. В. П. Козырева. 5-е изд., доп. Москва: УРСС: Ленанд, 2018. 297 с. Текст: непосредственный.