

**МИНОБРНАУКИ РОССИИ**

**федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«Московский государственный технологический университет «СТАНКИН»**

**(ФГБОУ ВО МГТУ «СТАНКИН»)**

|  |  |
| --- | --- |
| **Институт**  информационных технологий  и систем управления | **Кафедра**  информационных систем |

**Отчет по лабораторной работе №4**

по дисциплине «**Мировые информационные ресурсы и сети**»

на тему: «А**нализ работы поисковых систем**»

|  |  |
| --- | --- |
| Студент  группы ИДБ-14-13 | \_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Матвеев А.С.**  подпись |
| Руководитель  преподаватель | \_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Солдатов А.В.**  подпись |

**Москва**

**2018**

**Цель лабораторной работы:**

Составить 10 различных поисковых запросов, используя операторы языка поисковых запросов. По каждому из составленных запросов провести поиск в трёх различных поисковых системах.

Результаты представить в виде таблицы (1-ая колонка - "Описание поискового запроса", 2-ая - 4-ая колонки - "Синтаксис запроса в ... (название поисковой системы) и количество найденных документов").

Для любого поискового запроса для каждой поисковой системы определить следующие показатели эффективности поисковых систем: точность выдачи и информационный шум.

В качестве заключения сделать выводы о релевантности и качестве поиска в целом и сравнить выбранные поисковые системы.

**Ход выполнения:**

Используя операторы языка поисковых запросов составлены 10 поисковых запросов в поисковых системах “Google”, “Яндекс”, “Yahoo!”.

Запросы и результаты по ним представлены в таблице 1.

Таблица 1

**Поисковые запросы**

|  |  |  |  |
| --- | --- | --- | --- |
| Описание поискового запроса | Синтаксис запроса в Google и количество найденных документов | Синтаксис запроса в Яндекс и количество найденных документов | Синтаксис запроса в Yahoo! и количество найденных документов |
| Страницы, содержащие слова «котики» или «енотики», или оба слова | котики OR енотики  ~135 тыс. | котики | енотики  ~106 млн. | котики OR енотики  ~1,04 млн. |
| Страницы, содержащие цитату «Недопустимы доказательства» | "Недопустимые доказательства"  ~23,6 тыс. | "Недопустимые доказательства"  ~5 млн. | "Недопустимые доказательства"  ~50,9 тыс. |
| Страницы, содержащие фразу «с переподвыподвертом» в том виде, в котором она представлена | "с переподвыподвертом"  ~8,69 тыс. | !с переподвыподвертом  ~53 млн. | "с переподвыподвертом"  ~20,3 тыс. |
| Страницы обязательно содержащие слова «площадь», «Лондон» и возможно содержащие слово «Кинг» | Кинг +площадь +Лондон  ~472 тыс. | Кинг +площадь +Лондон  ~15млн. | Кинг +площадь +Лондон  ~30,7 тыс. |
| Поиск страниц с цитатой «К нам на… рассол», содержащей пропущенное слово | К нам на \* рассол  ~10 | "К нам на \* рассол"  =258 | К нам на \* рассол  ~93,5 тыс. |
| Поиск страниц, содержащих слово «безумие» и не содержащих слово «радужное» | безумие –радужное  ~6,01 млн | безумие –радужное  ~130 млн. | безумие –радужное  ~517 тыс. |
| Поиск страниц в домене рф | финальная стадия site:.рф  ~23,4 тыс. | финальная стадия domain:рф  ~19 млн. | финальная стадия site:.рф  ~101 тыс. |
| Поиск страниц, обязательно содержащих «штрафы», не содержащих «Недвижимость» и опционально содержащих «налоги» в домене рф | -Недвижимость + штрафы налоги site:.рф  =108000 | -Недвижимость +штрафы налоги domain:рф  ~191000 | -Недвижимость +штрафы налоги domain:ru  ~131000 |
| Страницы, обязательно содержащие «Грудинин» и одно из «счета» или «налоги», или оба и «счета» и «налоги», не содержащие «Швейцария» | +Грудинин +(счета OR налоги) –Швейцария  ~2 | +Грудинин +(счета | налоги) –Швейцария  ~29 млн. | +Грудинин +(счета OR налоги) –Швейцария  =5 |
| Поиск страниц содержащих фразу «Улица … фонарей» с пропущенным словом и обязательно содержащих точную фразу «сезон 1» | Улица \* фонарей +"сезон 1"  ~3,47 млн. | "Улица \* фонарей + сезон 1"  ~3 тыс. | Улица \* фонарей +"сезон 1"  ~31,8 тыс. |

Для расчетов значения точности выдачи и информационного шума поисковой системы был выбран следующий запрос:

**-Недвижимость +штрафы налоги site:.рф**

Поисковая система **“Google”**

* множество релевантных запросов: ***а*** = 107982,
* множество нерелевантных запросов: ***b*** = 18,
* множество релевантных, но не выводимых системой: ***c*** = 207,
* точность выдачи: ***TB*** = (107982/ (107982+207)) \*100%=99,8%,
* информационный шум: ***ИШ*** = (18/ (107982+18)) \*100%=0,02%.

Поисковая система **“Yandex”**

* множество релевантных запросов: ***a*** = 190932,
* множество нерелевантных запросов: ***b***=68,
* множество релевантных, но не выводимых системой: ***c*** = 311,
* точность выдачи: ***TB*** = (190932/ (190932+311)) \*100%=99,83%,
* информационный шум: ***ИШ*** = (68/ (190932+68)) \*100%=0,03%.

Поисковая система **“Yahoo!”**

* множество релевантных запросов: ***а*** = 130968,
* множество нерелевантных запросов: ***b*** = 32,
* множество релевантных, но не выводимых системой: ***c*** = 520,
* точность выдачи: ***ТВ*** = (130968/ (130968+520)) \*100%=99,6,
* информационный шум: ***ИШ*** = (32/ (130968+32)) \*100%=0.024%

**Вывод:** поисковые системы “Google”, “Yandex”, “Yahoo!” выводят релевантные запросы примерно с вероятность в 98-99% с учетом использования специальных операторов. Поисковая система Yandex выдает наибольшее количество запросов по сравнению с Google и Yahoo, но многие страницы повторяются.