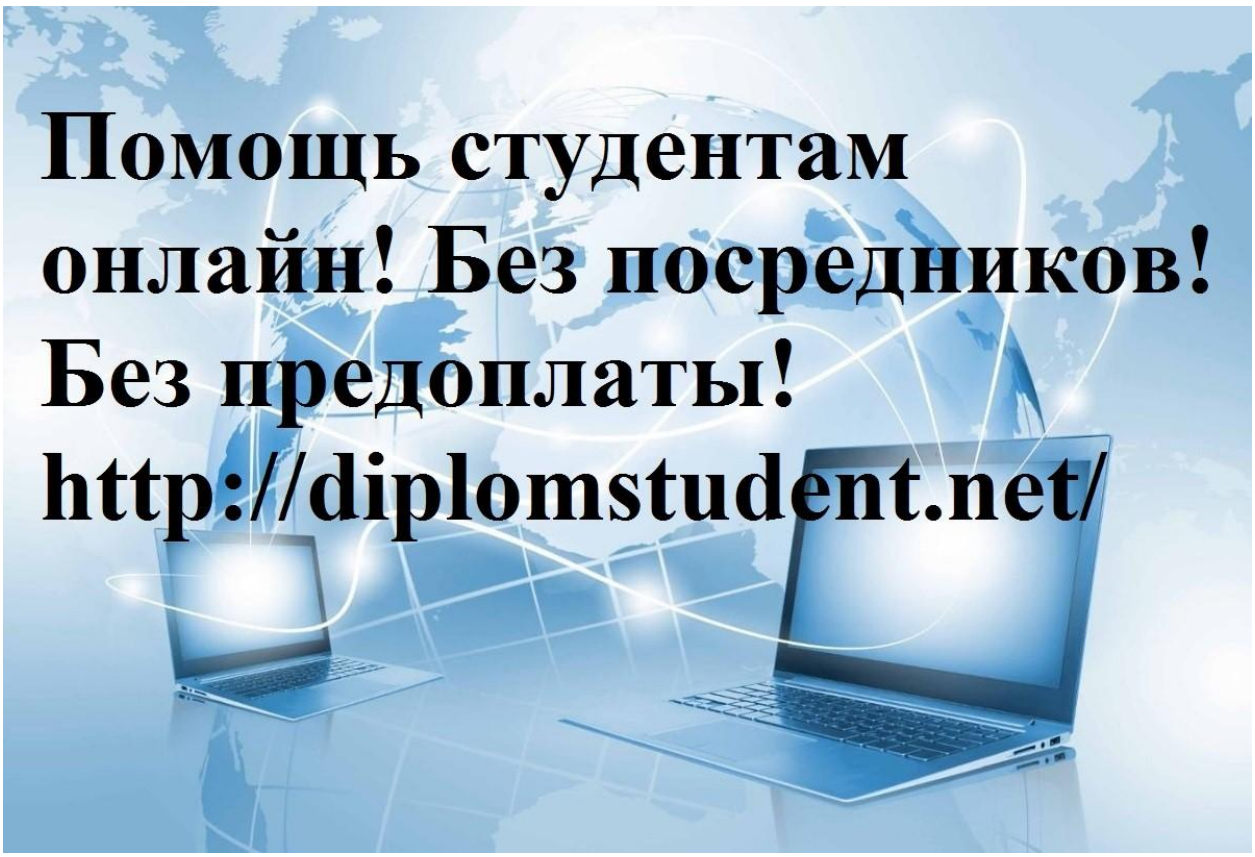

Практическая работа 1. Анализ характеристик поисковых систем

Примером сервисов, позволяющих работать с Большими данными, являются поисковые машины. Они аккумулируют на своих серверах миллиарды веб-страниц, содержащих текстовую информацию. Данная лабораторная работа преследует цель выявления наилучшего и наихудшего поискового ресурса, который следует использовать для поиска информации в конкретной области знаний. Отчет по лабораторной работе должен состоять из таблицы, аналогичной представленной ниже.

В отчете необходимо указать ФИО студента, выполнившего работу, указать формулировку запроса, по которому проводился поиск информации в поисковых системах. Затем необходимо полностью заполнить таблицу и сделать вывод о качестве поиска в каждой из представленных в таблице поисковых систем.

№	адрес	Всего проиндексировано документов	Всего найдено по запросу	Релевантность
1	https://yandex.ru/			
2	https://duckduckgo.com			
3	https://www.yahoo.com/			
4	https://www.google.ru/	2 млрд.	4,6 млн.	8 из 10
5	https://www.rambler.ru/			
6	https://www.bing.com			

В графу адрес можно внести не менее 6 поисковых систем, в таблице выше представлены программные продукты – лидеры данного рынка в РФ.

В столбце «***Всего проиндексировано документов***» необходимо внести результаты поиска по короткому слову, встречающемуся на каждой странице русскоязычного текста

(в примере показан результат поиска в Гугле по предлогу «и» – для получения данных о других системах *необходимо ввести то же слово* в остальные поисковые системы).

Для заполнения столбца «***Всего найдено по запросу***» также следует одну и ту же фразу (не менее 3-х слов) ввести во все представленные поисковые системы (из данных таблицы следует, что по запросу «информационные технологии в общественном питании» Гугл выдал более 12 млн. результатов).

Для заполнения графы «***Релевантность***» следует открыть первые 10 страниц в выдаче и проанализировать, сколько из них соответствуют цели поиска, а сколько из них попали в результирующий список по ошибке (в таблице указано, что изначальным ожиданиям пользователя, искавшего данные по запросу «информационные технологии в социальной сфере» соответствовали 8 из 10 первых предоставленных поисковой системой ссылок).

После заполнения таблицы необходимо сделать обоснованные выводы – о наилучшем и наихудшем поисковом ресурсе.

Решение:

Для выполнения задания в адрес поисковой строки набираем запрос: короткий запрос-роботизация и словосочетание роботизация для бизнеса, данные оформим в виде таблицы 1.

Таблица 1 – Анализ поисковых систем по запросу: роботизация и роботизация для бизнеса

№	адрес	Всего проиндексировано документов	Всего найдено по запросу	Релевантность
1	https://yandex.ru/	4 тыс.	25 тыс.	10 из 10
2	https:// www. mail.ru/	31 тыс.	246 тыс.	10 из 10
3	https://www.yahoo.com/	1650 тыс.	688тыс.	9 из 10
4	https://www.google.ru/	552 тыс.	739 тыс.	10 из 10
5	https://www.rambler.ru/	32 тыс.	311 тыс.	10 из 10
6	https://www.bing.com	1650 тыс.	5180 тыс.	10 из 10

Анализ поисковых систем, представленный в таблице 1 показывает, что по запросу роботизация для бизнеса, лучшей поисковой системой следует считать https://www.bing.com, так как по данному запросу поисковая система выдала больше всего результатов при очень высокой релевантности 10 из 10. Второе место в рейтинге можно отдать поисковой системе https://www.google.ru/, также высокая релевантность и большие количество найденных источников -739 тыс. Третье место по количеству найденных документов принадлежит поисковой системе – https://www.yahoo.com/, хотя по релевантности показатель ниже, чем у двух предыдущих поисковых систем. По другим поисковым системам релевантность составляет 10 из 10. Меньшее количество запросов наблюдается в поисковой системе- https://yandex.ru/.

Вывод: лучшая поисковая система по исследуемому запросу- https://www.bing.com