

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САНКТ-ПЕТЕРБУРГСКИЙ
ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО»

Институт компьютерных наук и технологий
Высшая школа интеллектуальных систем и суперкомпьютерных технологий

Дисциплина «Цифровые ресурсы в научных исследованиях»

ОТЧЕТ

по лабораторной работе №1

на тему

«Первичный полнотекстовый поиск источников информации в глобальной сети интернет»

Выполнил:

Студент группы 3540901/02001
Дроздов Н.Д.

«....» 2020г.,
(Подпись)

Проверил:

Бендерская Е.Н.

«....» 2020г.,
(Подпись)

Санкт-Петербург
2020 г.

Содержание

1	Лабораторная работа №1	2
1.1	Цель работы	2
1.2	Программа работы	2
1.3	Ход работы	3
1.3.1	Задание 1	3
1.3.2	Задание 2	3
1.3.3	Задание 3	4
1.3.4	Задание 4	4
1.3.5	Задание 5	4
1.3.6	Задание 6	5
1.3.7	Задание 7	5
1.3.8	Задание 8	6
1.4	Вывод	6
1.5	Список литературы	7

Лабораторная работа №1

1.1 Цель работы

Дать ответы на поставленные вопросы.

1.2 Программа работы

1. Сформулировать основные ключевые слова и словосочетания для выполнения первичного информационного запроса на русском языке по теме аналитического отчета. Записать ключевые слова и словосочетания в порядке релевантности. Выполнить поиск по всем ключевым словам и словосочетаниям, используемым как в одном поисковом запросе, так и по-отдельности, а провести поиск непосредственно по названию темы аналитического отчета. Провести первичную оценку результатов поиска путем выставления оценки от 1 до 10 степени релевантности полученных результатов для каждого запроса и представить их в отчете.
2. Оценить качество и количество представленных ссылок для лучшего варианта поискового запроса - оценка степени удовлетворенности качеством представленных источников от 1 (совсем не соответствует ожиданиям) до 10 (именно то, что ожидалось найти) и оценка достаточности представленных источников от 1 (очень мало) до 10 (очень много), и на этом основании выполнить уточнение информационного запроса – дополнение ключевых слов (исключение ключевых слов), расширение ключевых словосочетаний синонимами и т.д. Провести первичную оценку результатов поиска путем выставления оценки от 1 до 10 степени релевантности полученных результатов для каждого запроса и представить их в отчете.
3. Сравнить набор ключевых слов из лучших найденных источников (представить в отчете) и использованных при выполнении поискового запроса. Сделать вывод о необходимости коррекции состава ключевых слов или формулировки общего поискового запроса.
4. Выполнить п. 1-3 в еще одной (другой) поисковой системе. Сравнить полученные результаты и сформулировать выводы.
5. Изучить возможности поисковых систем по работе с расширенным поиском. Использовать расширенный поиск – привести примеры выполненного расширенного поиска и оценить качество полученных результатов по сравнению с простым поисковым запросом. Сделать выводы.
6. Выбрать из найденных источников наиболее релевантный и осуществить переработку первичного текста во вторичный, путем выполнения аннотирования с критической оценкой первоисточника и реферирования. Составить описательную аннотацию к тексту (описательная аннотация излагает, «о чем» написан первоисточник и лишь называет основные моменты содержания), составить реферативную аннотацию к тексту (реферативная аннотация, отражая основные вопросы содержания, в предельно сжатом виде передает также выводы по каждому из затронутых вопросов и по материалу в целом), составить аннотацию-резюме к тексту (аннотация-резюме характеризуется четкой передачей главного тематического содержания в предельно сжатом виде). Составить информативный реферат текста (не более 1/2 - 3/4 страницы)
7. Выполнить п. 1-6 для информационного запроса на английском языке.
8. Представить отчет о выполненной работе.

1.3 Ход работы

1.3.1 Задание 1

Сформулировать основные ключевые слова и словосочетания для выполнения первичного информационного запроса на русском языке по теме аналитического отчета. Записать ключевые слова и словосочетания в порядке релевантности. Выполнить поиск по всем ключевым словам и словосочетаниям, используемым как в одном поисковом запросе, так и по-отдельности, а провести поиск непосредственно по названию темы аналитического отчета. Провести первичную оценку результатов поиска путем выставления оценки от 1 до 10 степени релевантности полученных результатов для каждого запроса и представить их в отчете.

Основные ключевые слова и словосочетания для выполнения первичного информационного запроса на русском языке по теме "Большие данные и искусственный интеллект: новое и хорошо забытое старое" являются:

- Большие данные и искусственный интеллект;
- Интеллектуальный анализ данных;
- Обработка больших данных;
- Методы интеллектуального анализа данных;
- Задачи интеллектуального анализа данных;
- Области применения методов и технологий интеллектуального анализа данных;
- Области применения методов и технологий обработки больших данных;
- Большие данные;
- Искусственный интеллект;
- Машинное обучение.

Оценка результатов поиска:

Большие данные и искусственный интеллект	8
Интеллектуальный анализ данных	10
Обработка больших данных	9
Методы интеллектуального анализа данных	9
Задачи интеллектуального анализа данных	5
Области применения методов и технологий интеллектуального анализа данных	8
Области применения методов и технологий обработки больших данных	6
Большие данные	9
Искусственный интеллект	7
Машинное обучение	7

1.3.2 Задание 2

Оценить качество и количество представленных ссылок для лучшего варианта поискового запроса - оценка степени удовлетворенности качеством представленных источников от 1 (совсем не соответствует ожиданиям) до 10 (именно то, что ожидалось найти) и оценка достаточности представленных источников от 1 (очень мало) до 10 (очень много), и на этом основании выполнить уточнение информационного запроса – дополнение ключевых слов (исключение ключевых слов), расширение ключевых словосочетаний синонимами и т.д. Провести первичную оценку результатов поиска путем выставления оценки от 1 до 10 степени релевантности полученных результатов для каждого запроса и представить их в отчете.

По итогам оценки результатов поиска можно выделить наилучший вариант поискового запроса - Интеллектуальный анализ данных. Проведем оценку качества и количества представленных ссылок для данного поискового запроса:

Интеллектуальный анализ данных	
Качество	10
Количество	11665 (7)

После уточнения информационного запроса я могу сказать, что лучшим запросом является оригинал, потому что с помощью этого запроса можно добиться наилучшего результата при поиске. Были предложены следующие варианты поискового запроса:

- Интеллектуальный анализ больших данных;
- Интеллектуальная обработка больших данных;
- Искусственный интеллект в области обработки больших данных.

Оценка предложенных поисковых запросов:

Интеллектуальный анализ больших данных	6
Интеллектуальная обработка больших данных	6
Искусственный интеллект в области обработки больших данных	5

1.3.3 Задание 3

Сравнить набор ключевых слов из лучших найденных источников (представить в отчете) и использованных при выполнении поискового запроса. Сделать вывод о необходимости коррекции состава ключевых слов или формулировки общего поискового запроса.

После сравнения набора ключевых слов из лучших найденных источников (например [1] и [2]) с использованными при выполнении поискового запроса, можно заметить, что чаще всего первые ключевые слова полностью совпадают с текстом поискового запроса.

1.3.4 Задание 4

Выполнить п. 1-3 в еще одной (другой) поисковой системе. Сравнить полученные результаты и сформулировать выводы

В качестве второй поисковой системы было предложено использовать научную электронную библиотеку eLibrary [4].

Оценка результатов поиска:

Большие данные и искусственный интеллект	5
Интеллектуальный анализ данных	10
Обработка больших данных	7
Методы интеллектуального анализа данных	6
Задачи интеллектуального анализа данных	4
Области применения методов и технологий интеллектуального анализа данных	6
Области применения методов и технологий обработки больших данных	6
Большие данные	7
Искусственный интеллект	7
Машинное обучение	7

Можно заметить, что во второй поисковой системе запрос "Интеллектуальный анализ данных" также имеет наивысшую оценку.

Проведем оценку качества и количества представленных ссылок для выбранного поискового запроса:

Интеллектуальный анализ данных	
Качество	9
Количество	617243 (5)

Уточнение поискового запроса не зависит от поисковой системы, поэтому, я считаю, что наилучшим поисковым запросом из предложенных является "Интеллектуальный анализ данных".

1.3.5 Задание 5

Изучить возможности поисковых систем по работе с расширенным поиском. Использовать расширенный поиск – привести примеры выполненного расширенного поиска и оценить качество полученных результатов по сравнению с простым поисковым запросом. Сделать выводы

Анализ поисковых запросов проводился в научной электронной библиотеке «КИБЕРЛЕНИНКА». В ней присутствует возможность фильтрации результатов при поиске. Перечень доступных фильтров:

- Фильтр по году;
- Фильтр по терму OECD;
- Фильтр по научным базам;
- Фильтр по журналам.

Для выполнения заданий я использовал фильтр по терму OECD, а именно "Компьютерные и информационные науки". Можно было бы также использовать фильтр по году и поставить 2019-2020 года, так как в последние годы искусственный интеллект развивается все больше и больше, а значит качество научных публикаций на эту тему также растет. Используя фильтрацию можно получить результат более точный или более свежий, что сильно упрощает поиск.

1.3.6 Задание 6

Выбрать из найденных источников наиболее релевантный и осуществить переработку первичного текста во вторичный, путем выполнения аннотирования с критической оценкой первоисточника и реферирования. Составить описательную аннотацию к тексту (описательная аннотация излагает, «о чем» написан первоисточник и лишь называет основные моменты содержания), составить реферативную аннотацию к тексту (реферативная аннотация, отражая основные вопросы содержания, в предельно сжатом виде передает также выводы по каждому из затронутых вопросов и по материалу в целом), составить аннотацию-резюме к тексту (аннотация-резюме характеризуется четкой передачей главного тематического содержания в предельно сжатом виде). Составить информативный реферат текста (не более 1/2 - 3/4 страницы)

В качестве наиболее релевантного источника было использовать научную статью [1].

В данной статье излагаются особенности применения технологии интеллектуального анализа данных, описываются основные инструменты реализации данного метода анализа и представляются направления их развития.

В начале статьи делают утверждение о том, что нынешние учетные системы, которые используются на каждом предприятии, плохо приспособлены для принятия решения. В связи с этим, для анализа данных, которые накоплены в данных системах, есть необходимость использовать интеллектуальный анализ данных.

После чего автор статьи объясняет что такое интеллектуальный анализ данных и описывает, грубо говоря, стадии жизненного цикла анализа.

В статье также затронута тема реализации интеллектуального анализа. Представлены примеры механизмов, серверов, инструментов и платформ.

В конце статьи автор делает выводы по статье, говоря об направлениях развития. Он выделяет три направления, немного описывая каждое из них.

1.3.7 Задание 7

Выполнить п. 1-6 для информационного запроса на английском языке

В качестве поисковой системы было предложено использовать бесплатную поисковую систему Google Scholar [6].

Основные ключевые слова и словосочетания для выполнения первичного информационного запроса на английском языке по теме "Большие данные и искусственный интеллект: новое и хорошо забытое старое" являются:

- Big data and artificial intelligence;
- Data mining;
- Big data processing;
- Big data;
- Artificial intelligence;
- Machine learning;
- Information analysis.

Оценка результатов поиска:

Big data and artificial intelligence	8
Data mining	10
Big data processing	9
Big data	9
Artificial intelligence	5
Machine learning	7
Information analysis	7

По итогам оценки результатов поиска можно выделить наилучший вариант поискового запроса - Data mining.

Проведем оценку качества и количества представленных ссылок для данного поискового запроса:

Data mining

Качество 10

Количество примерно 3750000 (7)

Если сравнить ключевые слова в публикациях, полученных в результате поиска, с выбранным поисковым запросом, можно сделать вывод, что чаще всего выбранная тема попадает в первые ключевые слова. Вследствие чего можно сделать вывод, что Data mining является хорошим поисковым запросом.

В Google Scholar также имеется возможность фильтрации и сортировки с целью получения конкретного результата. Основными фильтрами и сортировками являются:

- Фильтр по году;
- Сортировка по дате;
- Сортировка по релевантности;

Также стоит отметить функцию "Похожие запросы" которая может сузить спектр поискового запроса.

В качестве наиболее релевантного источника было использовать научную статью [5].

В предложенной статье автор рассказывает про Big Data и развитие сбора данных в современном мире.

В этой статье представлена теорема НАСЕ, которая характеризует особенности революции больших данных, и предлагает модель обработки больших данных с точки зрения интеллектуального анализа данных.

Эта модель, управляемая данными, включает в себя объединение источников информации по запросу, интеллектуальный анализ и анализ, моделирование интересов пользователей, а также соображения безопасности и конфиденциальности.

1.3.8 Задание 8

Представить отчет о выполненной работе

1.4 Вывод

В ходе проделанной лабораторной работы были составлены ключевые слова, основные поисковые запросы на русском и английском языках на тему "Большие данные и искусственный интеллект: новое и хорошо забытое старое". Были проанализированы различные поисковые системы с составленными поисковыми запросами. Стоит отметить возможность фильтрации в поисковых системах, что сильно ускоряет и упрощает поиск необходимой информации.

Благодаря ключевым словам в научных публикациях, можно видоизменять свой поисковый запрос, что также может привести к упрощению поиска нужной информации.

1.5 Список литературы

- [1] ТЕХНОЛОГИИ РЕАЛИЗАЦИИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ [Научная статья/Электронный ресурс]. — URL: <https://cyberleninka.ru/article/n/tehnologii-realizatsii-intellektualnogo-analiza-dannyh-1>
- [2] ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И ОБЛАЧНЫЕ ВЫЧИСЛЕНИЯ [Научная статья/Электронный ресурс]. — URL: <https://cyberleninka.ru/article/n/intellektualnyy-analiz-dannyh-i-oblachnye-vychisleniya>
- [3] НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА «КИБЕРЛЕНИНКА» [Электронный ресурс]. — URL: <https://cyberleninka.ru/>
- [4] НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА «elibrary.ru» [Электронный ресурс]. — URL: <https://elibrary.ru/>
- [5] Data mining with big data [Научная статья/Электронный ресурс]. — URL: <https://ieeexplore.ieee.org/abstract/document/6547630>
- [6] Google Scholar [Электронный ресурс]. — URL: <https://scholar.google.com/>