

Учреждение образования
“БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ”
Кафедра интеллектуальных информационных технологий

Отчёт по практической работе №1
по дисциплине “Естественно-языковой интерфейс
интеллектуальных систем”

Выполнил: ст. группы 121701
Пашин Н. А.
Маевский В. Г.

Проверила: Гракова Н. В.

Минск 2024

Цель: освоить на практике основные принципы реализации информационно-поисковых систем и методы оценки качества их работы.

Задание: Вариант 2

Интерфейс с пользователем

Сеть Интернет

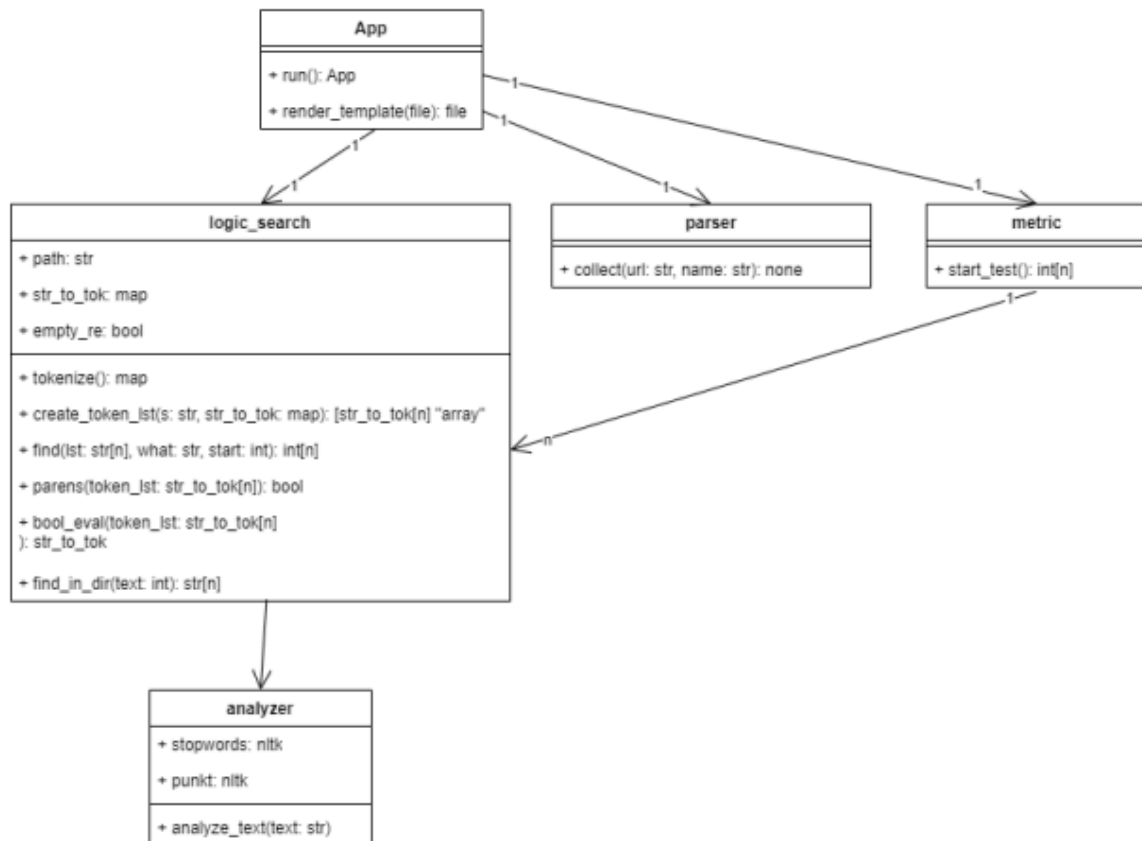
Логическая

Русский

Требования:

- на входе – множество естественно-языковых текстов по которым осуществляется поиск;
- выделение ключевых слов документов осуществляется системой автоматически в соответствии с формулой 1.6;
- система должна позволять пользователю формулировать ЕЯ-запрос;
- на выходе – список документов, релевантных запросу пользователя, в соответствии с моделью поиска, согласно варианту;
- результаты поиска должны содержать: активную ссылку на документ,
- список слов запроса присутствующих в документе.
- на основании информации о существующих метриках, наиболее часто использующихся для оценки качества работы систем информационного поиска (см. 2004_romip_metrix.pdf), требуется дать оценку работы СИП.
- Вычисление оценок, получаемых на основании метрик, реализовать программно, путем вызова соответствующего подменю, и отображать в виде таблиц и графиков
- интерфейс системы должен быть предельно простым и доступным для пользователей любого уровня, содержать понятный набор инструментов и средств, а также help-средства.

Структура системы:



Основные алгоритмы:

В логической стратегии поиска используется подход, основанный на представлении текстовых документов и запроса в виде векторов в пространстве признаков. Каждый термин или слово в поисковом запросе и документах представляется как элемент вектора, а соответствие между терминами и их весами отражает степень важности в контексте поиска.

Основной алгоритм поиска включает разбиение поискового запроса на термины и их представление в виде вектора запроса. Также каждый документ преобразуется в вектор, где каждая компонента отражает важность соответствующего термина в данном документе. Запрос формулируется как вектор, включающий в себя веса терминов и операторы для определения логических связей между ними. Векторы документов и запроса затем сравниваются, чтобы определить степень соответствия между запросом и документом. Результатом является ранжированный список документов по степени их релевантности к запросу.

Тестирование:

Запрос

Введите формулу для логического поиска

"Годзилла" ИЛИ "Автомобиль"

Поиск

Результат

Результат запроса:

Автомобиль — Википедия

Список присутствующих слов: ['Автомобиль']

Годзилла (фильм, 1998) — Википедия

Список присутствующих слов: ['Годзилла']

Тестовая коллекция:

file1.txt -

[https://ru.wikipedia.org/wiki/%D0%93%D0%BE%D0%B4%D0%B7%D0%B8%D0%BB%D0%BB%D0%B0_\(%D1%84%D0%B8%D0%BB%D1%8C%D0%BC,_1998\)](https://ru.wikipedia.org/wiki/%D0%93%D0%BE%D0%B4%D0%B7%D0%B8%D0%BB%D0%BB%D0%B0_(%D1%84%D0%B8%D0%BB%D1%8C%D0%BC,_1998))

file2.txt -

<https://ru.wikipedia.org/wiki/%D0%90%D0%B2%D1%82%D0%BE%D0%BC%D0%BE%D0%B1%D0%B8%D0%BB%D1%8C>

file3.txt -

https://ru.wikipedia.org/wiki/%D0%9B%D0%B5%D0%BE%D0%BD%D0%B0%D1%80%D0%B4%D0%BE_%D0%B4%D0%B0_%D0%92%D0%B8%D0%BD%D1%87%D0%B8

Результаты метрик:

- Логическая модель трактует термины в запросе как булевы переменные.
- При наличии термина в документе соответствующая переменная принимает значение «true» (истина).
- Присваивание терминам весовых коэффициентов не допускается.
- Запросы формулируются как произвольные булевы выражения, связывающие термины с помощью стандартных логических операций: ИЛИ, И или НЕ.
- Мерой соответствия запроса документу служит значение статуса выборки (RSV, retrieval status value).
- В булевой модели статус выборки равен либо 1, если для данного документа вычисление выражения запроса дает значение «истина», либо 0 в противном случае.
- Все документы с $RSV = 1$ считаются релевантными запросу.

```
Macroaverage:
recall = 1.0
precision = 1.0
accuracy = 1.0
error = 0.0
F-measure = 1.0
Microaverage:
recall = 1.0
precision = 1.0
accuracy = 1.0
error = 0.0
F-measure = 1.0
```

Описание и особенности применения готовых к использованию компонент

В системе используются следующие готовые к использованию компоненты:

FastApi – микрофреймворк для создания веб-приложений на языке программирования Python.

Natural Language Toolkit (NLTK) - это библиотека на языке программирования Python, предназначенная для обработки и анализа естественного языка (Natural Language Processing, NLP).

NLTK предоставляет множество инструментов, алгоритмов и корпусов данных, которые помогают исследователям и разработчикам в области NLP создавать и выполнять разнообразные задачи, связанные с текстовыми данными.

Вывод

В данной лабораторной работе была реализована информационно-поисковая система.

Реализованная система была протестирована и проанализирована с помощью метрик.

Были приведены алгоритмы реализации системы. В результате лабораторной работы

были закреплены принципы реализации информационно-поисковых систем.