

Летняя международная XXXI молодежная Школа-конференция по
параллельному программированию


ЭКСПЕРТНАЯ СИСТЕМА ВОПРОСОВ И ОТВЕТОВ НА ОСНОВЕ БАЗЫ ЗНАНИЙ

Исполнитель проекта: Жигмытов Т.А

Руководитель проекта: Толенбеков Е.К

Дата доклада: 10.07.2018

ПЛАН ДОКЛАДА:

1. Постановка задачи
 2. Идея решения
 3. Реализация
 4. Тестирование
 5. Заключение
- 

ТЕРМИНЫ

Стоп-слово – это **слова**, знаки, символы, которые самостоятельно не несут никакой смысловой нагрузки и просто игнорируются поисковыми системами при осуществлении ранжирования или индексации сайтов.

Токенизация - разбиение текста на слова (и не-слова, те знаки препинания, границы абзацев и т.п). Полезность её в машинном обучении - прямое донесение до сетки факта, что человек (чьим действиям её надо научить подражать) воспринимает текст как поток слов, а не поток букв.

Стемминг — это процесс нахождения основы слова для заданного исходного слова. Основа слова не обязательно совпадает с морфологическим корнем слова.

Стеммер Портера - Алгоритм не использует баз основ слов, а лишь, применяя последовательно ряд правил, отсекает окончания и суффиксы, основываясь на особенностях языка, в связи с чем работает быстро, но не всегда безошибочно.

ПОСТАНОВКА ЗАДАЧИ

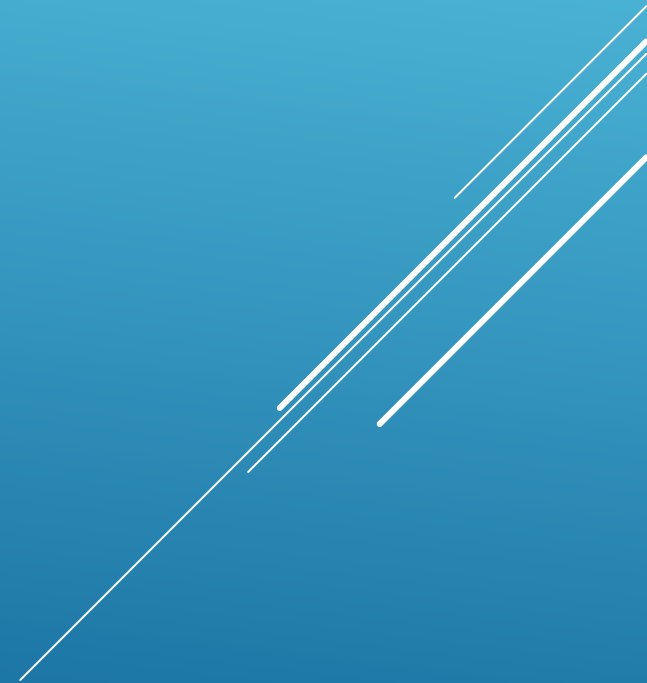
Задачей является разработка экспертной системы вопросов и ответов на основе базы знаний

Требования к решению задачи:


1. Изучить и внедрить в систему простую обработку текста
2. Настроить отправку данных на сервер
3. Создание минимального интерфейса для взаимодействия с системой
4. Развернуть систему на веб-сервере()

ТЕСТИРОВАНИЕ

Тестирование работы экспертной системы не отобразить в графиках, поэтому я использовал nrm пакет localhost для создания точки доступа внутри Wi-fi сети.



РЕАЛИЗАЦИЯ

1. Изучение библиотеки nltk(Natural language toolkit)
 2. Использование nltk для выявления стоп-слов, проведения токенизации и стемминга
 3. Развернуть и настроить маршрутизацию с помощью Python библиотеки Flask
 4. Тестирование отправки запросов
 5. Написание минимального интерфейса для взаимодействия
 6. Развернуть npm localhost для отображения
- 

ЗАКЛЮЧЕНИЕ

Результаты:

Получена простейшая экспертная система, которая проводит начальную очистку текста, представленная в виде веб-интерфейса с рабочей отправкой данных сервер. Также изучены основы работы Python и Flask.

Направление развития проекта:

Подключение корпусов определенной предметной области.

Добавление семантического анализа текста. Создание полноценной системы на русском языке на основе данного прототипа. Выявления особенностей казахского языка и основе корпуса, созданного экспертами. разработка бота для казахского языка. И последним этапом сделать систему кроссплатформенной.

СПАСИБО ЗА ВНИМАНИЕ!

