

# **Документация к приложению «BindWord XP» для анализа текстовых данных (NLP)**

## **Основной интерфейс**

Главное окно (MainWindow) — это центр работы пользователя с программой. Здесь можно ввести текст, запустить нужный тип анализа и посмотреть результаты в виде таблиц, графиков и текстов.

## **Панель управления (верхняя часть экрана)**

### **1. Файл**

Меню для загрузки и сохранения текстов. Функции: открытие текстового документа, автоматическое помещение его в поле ввода, а также завершение работы приложения.

### **2. Сохранить**

Сохраняет результаты анализа в выбранный формат: текст (.txt), таблицу (.csv) или структуру данных (.json).

### **3. Проверка**

Кнопка запуска предварительной проверки текста. Может использоваться перед запуском более сложного анализа.

### **4. Слова**

Запускает анализ слов: подсчёт общего количества слов, определение самых часто встречающихся, вывод статистики.

### **5. Векторы**

Представляет каждое слово или предложение в виде числового вектора. Это нужно, чтобы программа могла «понимать» смысл текста на математическом уровне. Методы, например: Word2Vec,

TF-IDF. Векторы можно использовать для кластеризации, сравнения текстов и визуализации.

## **6. Комбинации**

Анализирует, какие слова часто встречаются рядом друг с другом. Например, «искусственный интеллект» — это биграмма. Программа ищет такие пары и тройки слов (n-граммы), подсчитывает их частотность и показывает, в каком контексте они чаще всего встречаются.

## **7. Тематический анализ**

Позволяет определить, о чём идет речь в тексте. С помощью алгоритмов (например, LDA или NMF) выявляются скрытые темы, которым принадлежат слова. Это полезно для анализа больших объёмов текста — например, новостей или отзывов.

## **8. Распознавание именованных сущностей (NER)**

Определяет и выделяет имена людей, организации, даты, локации и другие важные элементы в тексте. Отображает их отдельно, что удобно для поиска конкретных фактов.

## **9. Переключатель модели (Русская/Английская)**

Позволяет выбрать, на каком языке будет происходить анализ. Для корректной работы с русским и английским текстом используются разные языковые модели.

## **10. Кнопка Старт**

Запускает выбранные анализы в зависимости от установленных галочек и загруженного текста.

## **Центральное текстовое поле (слева)**

Здесь вводится текст для анализа. Поле может быть заполнено вручную или автоматически при открытии файла.

## **Правый блок интерфейса**

В правой части окна расположены вкладки и текстовые поля, предназначенные для отображения результатов анализа текста. Каждая вкладка представляет собой отдельную категорию данных — например, таблица со словами, найденные комбинации, тематические блоки, именованные сущности и т.д. Пользователь может свободно переключаться между этими вкладками, чтобы просматривать различные аспекты анализа. Под вкладками находятся текстовые поля, где отображается более подробная информация: список результатов, выбранных фильтров или ключевых слов.

## **Нижний блок — графики и визуализация**

Во вкладках отображаются:

- графики частот слов и тем;
- список слов и их статистика;
- визуализация сочетаний слов (n-грамм);
- пространственное расположение векторов (например, с помощью t-SNE);
- выявленные темы и связи между ними.

## **Строка состояния (внизу)**

Показывает общее количество слов в тексте и время, которое занял последний анализ. Это помогает оценить производительность приложения.