**Министерство науки и высшего образования Российской Федерации**

**Федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«Московский государственный технический университет имени Н.Э. Баумана**

**(национальный исследовательский университет)»**

**(МГТУ им. Н.Э. Баумана)**

ФАКУЛЬТЕТ **ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ**

КАФЕДРА **КОМПЬЮТЕРНЫЕ СИСТЕМЫ И СЕТИ (ИУ6)**

НАПРАВЛЕНИЕ ПОДГОТОВКИ 09.03.01 Информатика и вычислительная техника

Программная система семантического поиска в текстах

Техническое задание на курсовую работу

по дисциплине Технология разработки программных систем

Листов 8

Студент \_гр. ИУ6-55Б**\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_А. А. Оленев\_\_\_\_**

(Группа) (Подпись, дата) (И.О. Фамилия)

Руководитель курсовой работы,

(Уч. степень, уч. звание, должность)  **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_Е. Ю. Гаврилова\_\_**

(Подпись, дата) (И.О. Фамилия)

Москва, 2025

# 1 Введение

Настоящее техническое задание распространяется на разработку программной системы семантического поиска по текстовым документам (ПССПТД), предназначенной для выполнения глубокого смыслового анализа и поиска информации в неструктурированных текстовых данных.

Современный этап развития информационных технологий характеризуется экспоненциальным ростом объемов текстовой информации, хранящейся в электронном виде. Традиционные системы поиска, основанные на ключевых словах, зачастую неспособны эффективно решать задачи поиска по смыслу, учитывать синонимию, контекст использования терминов и семантические связи между понятиями. Это приводит к низкой релевантности результатов и потере значимой информации.

Разрабатываемая система предназначена для преодоления указанных ограничений за счет применения методов обработки естественного языка (NLP) и семантического анализа. Ключевым отличием системы является способность понимать смысловое содержание запроса и документов, что позволяет находить информацию на концептуальном уровне, а не на уровне простого совпадения слов.

# 2 Основания для разработки

Программная система семантического поиска по текстовым документам разрабатывается на основании учебного плана кафедры ИУ6 «Компьютерные системы и сети» факультета «Информатика и системы управления» МГТУ им. Н. Э. Баумана.

# 3 Назначение разработки

Основное назначение программной системы семантического поиска по текстовым документам заключается в обеспечении полнотекстового поиска информации на основе смыслового содержания документов, а не исключительно по ключевым словам. Система предназначена для обработки неструктурированных текстовых данных и предоставления релевантных результатов, учитывающих семантические связи и контекст.

# 4 Требования к программному изделию

## 4.1 Требования к функциональным характеристикам

## 4.1.1 Выполняемые функции

## 4.1.1.1 Для пользователя:

* формирование поискового запроса на естественном языке;
* фильтрация результатов поиска по типам документов, дате создания/изменения, источнику и другим метаданным;
* просмотр найденного текстового документа с выделением семантически близких к запросу фрагментов;
* просмотр списка документов, семантически близких к текущему;
* просмотр, сохранение и повторное использование предыдущих поисковых запросов и их результатов;
* формирование системы рекомендаций документов на основе анализа истории поиска и просмотра;
* изменение содержащихся в системе документов;
* добавление новых документов в систему.

## 4.1.2 Исходные данные:

* пользовательские запросы, вводимые на естественном языке;
* текстовые документы для выполнения поиска (форматы: TXT, PDF, DOCX, HTML, MD);
* метаданные документов (название, автор, дата создания/изменения, тип документа).

## 4.1.3 Результаты:

* ранжированный список документов, релевантных смыслу поискового запроса, отсортированный по степени семантического сходства;
* представление результатов поиска для каждого документа, включающее название документа, источник, дата создания/изменения, фрагмент текста с подсветкой ключевых слов и семантически близких к запросу элементов, ссылка на полный текст документа, список семантически близких документов (аналоги) для выбранного документа из результатов поиска, персонализированные рекомендации;
* возможность вывода результатов (списка документов и их метаданных) в машиночитаемых форматах для последующего анализа (например, CSV, JSON).

## 4.2 Требования к надежности

4.2.1 Предусмотреть контроль вводимой информации.

4.2.2 Предусмотреть блокировку некорректных действий пользователя.

4.2.3 Обеспечить целостность информации в базе данных.

## 4.3 Условия эксплуатации

4.3.1 Условия эксплуатации в соответствии с СанПиН 2.2.2/2.4.1340-03.

## 4.4 Требования к составу и параметрам технических средств

4.4.1 Программное обеспечение должно функционировать на IBM-совместимых персональных компьютерах.

4.4.2 Минимальная конфигурация технических средств:

4.4.2.1 Тип процессора Intel Core i3 и выше

## 4.4.2.2 Объем ОЗУ 100 Мб.

## 4.4.2.3 Объем свободного дискового пространства 500 Мб.

## 4.5 Требования к информационной и программной совместимости

4.5.1 Программное обеспечение должно работать под управлением операционных систем семейств WIN32, WIN64 и Unix - подобные

4.5.2 Входные данные представлены в следующем формате: база данных с документами и поисковый запрос пользователя на русском языке.

4.5.3 Результаты должны быть представлены в следующем формате: список релевантных текстовых документов.

4.5.4 Программа должна разрабатываться на языке программирования Python c использованием фреймворка PyQt5.

# 5 Требования к программной документации

5.1 Разрабатываемые программные модули должны быть самодокументированы, т.е. тексты программ должны содержать все необходимые комментарии.

5.2 Разрабатываемое программное обеспечение должно включать справочную систему.

5.3 В состав сопровождающей документации должны входить:

5.3.1 Расчетно-пояснительная записка на 25-30 листах формата А4 (без приложений 5.3.2, 5.3.3 и 5.3.4).

5.3.2 Техническое задание (Приложение А).

5.3.3 Руководство пользователя (Приложение Б).

5.4 Графическая часть должна быть включена в расчетно-пояснительную записку в качестве иллюстраций:

5.4.1 Диаграмма вариантов использования (объектный подход).

5.4.2 Концептуальная модель предметной области (объектный подход).

5.4.3 Диаграммы (схемы) компонентов структур данных.

5.4.4 Схема структурная программного обеспечения.

5.4.5 Граф состояний интерфейса.

5.4.6 Схемы алгоритмов модулей (подпрограмм).

5.4.7 Диаграммы компоновки программных компонентов.

5.4.8 Таблицы тестов.

# 6 Стадии и этапы разработки

| **Этап** | **Содержание этапа** | **Сроки и**  **объем** | **Представляемые результаты** | |
| --- | --- | --- | --- | --- |
| **Спецификации и программный продукт** | **Документы** |
|  | Выбор темы, составление задания, решение организационных вопросов | 1..2 недели  (10 %) | - | **Заполненный бланк задания на курсовую работу – вывешивается на сайт кафедры для получения утверждающей подписи заведующего кафедрой** |
|  | Анализ предметной области, разработка ТЗ. Исследование методов решения, выбор основных проектных решений | 3..4 недели | Результаты декомпозиции предметной области.  Эскизный проект: интерфейс, схемы, возможно, часть программы (выбранные готовые решения). | Фрагмент расчетно-пояснительной записки с обоснованием выбора средств и подходов к разработке |
|  | **Сдача ТЗ** | **4 неделя**  **(25 %)** | **-** | **Техническое задание – утверждается руководителем** |
|  | Проектирование и реализация основных компонентов – ядра программы | 5..7 недели | Технический проект основной части: структура программы, алгоритмы программ, описания структур данных, диаграмма классов – в зависимости от выбранной технологии разработки.  Программный продукт, реализующий основные функции (демонстрируется руководителю) | Фрагмент расчетно-пояснительной записки с обоснованием разработанных спецификаций  Тексты части программного продукта, реализующего основные функции. |
|  | **Сдача прототипа программного продукта** | **7 неделя**  **(50 %)** | **Прототип программного продукта – демонстрируется руководителю** |  |
|  | Разработка компонентов, обеспечивающих функциональную полноту | 8..10 | Рабочий проект программы.  Готовая программа | Черновик расчетно-пояснительной записки.  Тексты программного продукта. |
|  | **Сдача программного продукта** | **11 неделя**  **(75 %)** | **Готовая программа – оценивается руководителем в баллах** | **-** |
|  | Тестирование программы и подготовка документации | 12..14 | Тесты и результаты тестирования. | РПЗ и Руководство пользователя. |
| **9.** | **Оформление и сдача документации** | **14 неделя**  **(90 %)** | **–** | **Расчетно-пояснительная записка и Руководство пользователя – проверяются и подписываются руководителем** |
| 10. | Защита курсовой работы | 15..16 недели  (100%) | – | Доклад (3-5 минут). Защита курсовой работы.  Подписанная документация – вывешивается на сайт кафедры |

# 7 Порядок контроля и приемки

## 7.1 Порядок контроля

Контроль выполнения осуществляется руководителем еженедельно.

## 7.2 Порядок защиты

Защита осуществляется комиссии преподавателей кафедры.

## 7.3 Срок защиты

Срок защиты: 15-16 недели.

# 8 Примечание

В процессе выполнения работы возможно уточнение отдельных требований технического задания по взаимному согласованию руководителя и исполнителя.