**Анализ предметной области**

**1. Понятия**

**Классификация текстов** — задача компьютерной лингвистики, заключающаяся в отнесении документа к одной из нескольких категорий на основании содержания документа. Классификация текстов применяется, в том числе, для:

* разделения веб страниц и сайтов по тематическим каталогам;
* борьбы со спамом;
* определение языка текста;
* показа более релевантной рекламы.

**Машинное обучение** — класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач. Для построения таких методов используются средства математической статистики, численных методов, математического анализа, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме.

**Нейронная сеть** (также искусственная нейронная сеть, ИНС) — математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма.

**TensorFlow** — открытая программная библиотека для машинного обучения, разработанная компанией Google для решения задач построения и тренировки нейронной сети с целью автоматического нахождения и классификации образов, достигая качества человеческого восприятия. Применяется как для исследований, так и для разработки собственных продуктов Google. Основной API для работы с библиотекой реализован для Python, также существуют реализации для R, C Sharp, C++, Haskell, Java, Go и Swift.

**2. Классификация текстов**

Цель создаваемой модели - классификации входного текста по категориям:

* Входные данные - Текст
* Выходные данные – Категория классификации

**3. Постановка задачи**

Основной целью создания ПО «Классификатор текстов» является автоматизации процедуры классификации текстов.

В соответствии с поставленной целью, создаваемая система должна уметь решать следующие задачи:

* Считывать входные данные в качестве текста
* Предоставлять пользователю доступ к конфигурации
* Выводить результаты о классификации входных данных

**Техническое задание**

**1 Общие сведения**

**1.1 Полное наименование системы и ее условное обозначение**

Полное наименование системы – «Программное обеспечение классификации текстов». Краткое наименование - ПО «Классификатор текстов».

**1.2 Наименование предприятия (объединения) разработчика, заказчика (пользователя) разработки и их реквизиты**

Заказчик системы: ООО «Текст».

Разработчик системы: Глухов Иван Дмитриевич, студент группы AI.MGIMO/19 МГИМО.

**1.3 Перечень документов, на основании которых создается разработка**

Данная система создается на основании технического задания заказчика и заявки заказчика.

**1.4 Плановые сроки начала и окончания работы по созданию системы**

Начало работ по созданию системы: \_\_\_\_\_\_\_\_\_

Окончание работ по созданию системы: \_\_\_\_\_\_\_\_\_

**1.5 Порядок оформления и предъявления заказчику результатов работ по созданию разработки программного комплекса**

Результаты работ по созданию работы оформляются в исполняемого файла, реализующего функции ПО, а также документов «Руководство по эксплуатации» и «Руководство программиста».

**2 Назначение и цели создания ПО «Классификатор текстов»**

**2.1 Назначение ПО «Классификатор текстов»**

Система разрабатывается для и предназначена для автоматизированной классификации текстов с использованием библиотеки **TensorFlow**.

**2.2 Цели и задачи создания ПО «Классификатор текстов»**

Основной целью создания ПО «Классификатор текстов» предоставление функционала по автоматизированной классификации текстов с использованием нейронной сети.

В соответствии с поставленной целью, создаваемая система должна уметь решать следующие задачи:

* считывание текста в качестве входных данных;
* обработку полученных данных в удобном для ПО виде;
* предоставлять пользователю возможности по конфигурации;
* формирование классификации текста в качестве выходных данных.

**3 Характеристика объекта автоматизации**

В связи с увеличением объема задач по работе с текстом и продвинутым технологиям искусственного интеллекта и машинного обучения возникает необходимость в создании и использовании инструмента, позволяющего выполнять классификацию текста в автоматизированной режиме.

**4 Требования к системе**

**4.1 Требования к системе в целом**

**4.1.1 Требования к структуре и функционированию ПО «Классификатор текстов»**

**4.1.1.1 Перечень подсистем**

Вся система должна состоять из 2 подсистем:

* подсистема работы с текстом
* подсистема нейронной сети

**4.1.1.2 Требования к функционированию ПО «Классификатор текстов»**

Общение пользователя с системой должно происходить в консольном терминале посредством ввода пользователем системных команд.

Для корректного функционирования ПО «Классификатор текстов» необходимо наличие отдельного компьютера, подключенного к ИБП, причем функционировать система должна 24 часа в сутки. Также необходим доступ к сети Интернет.

**4.1.2 Требования к численности и квалификации работников и режиму его работы**

С системой работает один человек, который должен обладать знаниями и навыками работы с персональным компьютером, компонентами ПК, иметь знания по технике безопасности при работе с электроустановками, иметь навыки работы с данной системой.

Режим работы персонала определяется предприятием, эксплуатирующим систему, и должен соответствовать требованиям по эргономике и безопасности труда, определенных в стандартах.

**4.1.3 Требования к безопасности**

Требования по безопасности должны соответствовать разделу 2 ГОСТ 24.104-85. Неправильные действия персонала не должны приводить к аварийной ситуации. Требования по безопасности средств вычислительной техники должны соответствовать ГОСТ 25.861-83. Все внешние элементы технических средств, находящиеся под напряжением, должны иметь защиту от случайного прикосновения, а сами технические средства иметь защитное заземление в соответствии с ГОСТ 12.1.030-81 и “Правилами устройства электроустановок”.

**4.1.4 Требования к эксплуатации, техническому обслуживанию, и ремонту**

Пользователь, работающий с системой, должен обладать знаниями и навыками работы с персональным компьютером, ИБП, должен знать правила техники безопасности при работе с электроустановками, иметь навыки работы с данной системой.

Режим эксплуатации системы круглосуточный, с учетом специфики работы.

Для размещения комплекса технических средств и персонала необходима площадь не менее 2 кв.м.

Регламент обслуживания технических средств предусматривает проведение профилактических обследований и ремонтов; замену вышедших из строя деталей.

**4.1.5 Требования по сохранности информации при авариях**

В случае аварий информация должна сохраняться. С этой целью используется система создания резервных копий, обеспечение безопасности хранимых данных от несанкционированного доступа.

**4.2 Требования к функциям и задачам ПО «Классификатор текстов»**

В состав разрабатываемой системы будут входить следующие подсистемы:

* *подсистема работы с текстом. Основные функции – обработка, форматирование и анализ входных данных в виде текста;*
* *подсистема нейронной сети. Основные функции – инициализация, конфигурация и тренировка нейронной сети с возможностью сохранения результатов.*

**4.3 Требования к видам обеспечения**

**4.3.1 Требования к математическому обеспечению**

Для реализации функций системы будут использоваться алгоритмы машинного обучения.

**4.3.2 Требования к информационному обеспечению**

Входными данными будут являться текстовые данные.

Выходными данными являются классификатор текста.

**4.3.3 Требования к программному обеспечению**

Для успешного проектирования и эксплуатации системы необходимы следующие минимальные требования к программному обеспечению:

- операционная система Windows 10 и выше;

- Python 3.6 и выше

- TensorFlow

**4.3.4 Требования к техническому обеспечению**

Для успешного функционирования системы необходимы следующие минимальные требования к техническому обеспечению:

- процессор Intel Core i3 с тактовой частотой не ниже 1,5 ГГц;

- 1024 Мбайт оперативной памяти;

- жесткий диск емкостью не менее 80 Гб;

- монитор 19'';

- клавиатура стандартная русифицированная;

- мышь;

- сетевая карта.

**5 Состав и содержание работ по созданию системы**

Разработка ПО «Классификатор текстов» должна содержать следующие этапы:

* Обзор и анализ существующих методов и систем;
* разработка общесистемных решений;
* разработка информационного обеспечения;
* разработка математического обеспечения;
* разработка программного обеспечения;
* разработка технического обеспечения;
* разработка организационного обеспечения;
* разработка документов.

Эти работы должны быть выполнены в срок с \_\_\_\_\_\_\_\_\_\_\_\_\_\_ по\_\_\_\_\_\_\_\_\_\_. По окончании всех видов работ предъявляется пояснительная записка к дипломному проекту, содержащая разделы: обзор и анализ существующих систем, общесистемные решения, информационное обеспечение, математическое обеспечение, техническое обеспечение, программное обеспечение, организационное обеспечение.

К записке должны быть приложены настоящее техническое задание, текст программы, графические материалы.

**6 Порядок контроля и приемки разработки**

Контроль и приемка работы должна происходить на контрольном примере, обеспечивающем просмотр работоспособности всех функций разработки.

Готовая разработка представляется Государственной аттестационной комиссии. Работа сопровождается пояснительной запиской и графическим материалом.

**7 Требования к документированию**

Схемы алгоритмов, программ данных и систем, условные обозначения и правила выполнения регламентируются ГОСТ 19.701-90.

Общие требования к текстовым документам регламентируются ГОСТ 2.105-95.

Техническое задание на создание автоматизированной системы выполняется на основании ГОСТ 34.602-89.