

Нормализация данных при помощи LLM

Паспорт №963/ЛКП-2334-2025 от 12.02.2025

на основании заявки № 963/ЛКП-2334-2025 от 12.02.2025

Статус документа: Паспорт утвержден

Полное название проекта

Нормализация данных при помощи LLM

Краткое название проекта

Безопасность компьютерных систем

Нормализация данных при помощи LLM

Название для диплома студента

Безопасность компьютерных систем

Нормализация необработанных текстовых данных при помощи LLM

Цель

Нормализация необработанных текстовых данных, в виде товарных записей, при помощи большой языковой модели (LLM). Под нормализацией понимается разбиение записей на компоненты, относящихся к определенному атрибуту (характерному свойству объекта)

Результат (продукт)

Набор проанализированных идей по способу решения задачи, набор скриптов, Jupiter ноутбуки (в основе решение на локальной open source LLM).

Безопасность компьютерных систем

Набор проанализированных идей по способу решения задачи, набор скриптов, Jupiter ноутбуки (в основе решение на локальной open source LLM).

Критерии приемки результата / продукта

РОС (proof of concept). Код опубликован и протестирован, определены метрики эффективности алгоритмов. Решение обладает ценностью для заказчика. Исследованы несколько больших языковых моделей, выбрана оптимальная LLM, обеспечивающая лучшую эффективность (качество/скорость/энерго и ресурсопотребление). Определены технические требования и ограничения решения.

Безопасность компьютерных систем

РОС (proof of concept). Код опубликован и протестирован, определены метрики эффективности алгоритмов. Решение обладает ценностью для заказчика. Исследованы несколько больших языковых моделей, выбрана оптимальная LLM, обеспечивающая лучшую эффективность (качество/скорость/энерго и ресурсопотребление). Определены технические требования и ограничения решения.

Описание проекта

Нормализация данных при помощи LLM.

Пример исходной записи №1 (входные данные): Автошина 235/70R16 Cordiant Snow Cross 2 SUV шип

Пример нормализованной записи №1: Вид продукции - Автошина; Ширина профиля покрышки -235; Высота профиля покрышки, % - 70; Диаметр посадочной полки обода - R16; Тип шины - Неизвестно; Коэффициент нагрузки шины автомобильной - Неизвестно; Индекс скорости шины автомобильной - Неизвестно; Норма слойности шины - Неизвестно; Фирма производитель: Cordiant Snow Cross; Наличие шипов у автошин – шипованная/

Пример исходной записи №2: Гайка M10 ГОСТ 5915-70

Пример нормализованной записи №2:

Вид продукции - Гайка; Тип гайки - Неизвестно; Исполнение - Неизвестно; Диаметр резьбы, мм - M10; Шаг резьбы, мм - Неизвестно; Класс прочности - Неизвестно; Условное обозначение группы материалов - Неизвестно; Металлы и сплавы - Неизвестно; Покрытие изделия - Неизвестно; Толщина покрытия, мкм - Неизвестно; Стандарт - ГОСТ 5915-70

Нормализация данных при помощи LLM.

Пример исходной записи №1 (входные данные): Автошина 235/70R16 Cordiant Snow Cross 2 SUV шип

Пример нормализованной записи №1: Вид продукции - Автошина; Ширина профиля покрышки -235; Высота профиля покрышки, % - 70; Диаметр посадочной полки обода - R16; Тип шины - Неизвестно; Коэффициент нагрузки шины автомобильной - Неизвестно; Индекс скорости шины автомобильной - Неизвестно; Норма слойности шины - Неизвестно; Фирма производитель: Cordiant Snow Cross; Наличие шипов у автошин – шипованная/

Пример исходной записи №2: Гайка M10 ГОСТ 5915-70

Пример нормализованной записи №2:

Вид продукции - Гайка; Тип гайки - Неизвестно; Исполнение - Неизвестно; Диаметр резьбы, мм - M10; Шаг резьбы, мм - Неизвестно; Класс прочности - Неизвестно; Условное обозначение группы материалов - Неизвестно; Металлы и сплавы - Неизвестно; Покрытие изделия - Неизвестно; Толщина покрытия, мкм - Неизвестно; Стандарт - ГОСТ 5915-70

Максимальное количество экземпляров проекта

2

Название организации

ООО "САЙБЕРЛИМФА"

None

Заказчик

Черноскутов Дмитрий

+7(922)229-28-45

chernoskutov1989dv@mail.ru

Участники проекта от университета

Безопасность компьютерных систем

- Овечкина Елена Владимировна

Главный ООП

elena.ovechkina@urfu.ru

- Овечкина Елена Владимировна

Главный куратор

elena.ovechkina@urfu.ru

Образовательные программы

Направление: 10.03.01 Информационная безопасность

Программа: 10.03.01/33.01 Безопасность компьютерных систем

Семестр: Весенний, курс: 2, 3

Куратор от профильной организации

Адрес для прохождения практической подготовки в форме проектного обучения

Свердловская область, Екатеринбург, Мира, 19

Тип и сложность проекта

Безопасность компьютерных систем

Тип: прикладной (практико-ориентированный), уровень сложности: Тип В. Известные методы и инструменты, уникальный результат, маловероятна финансовая отдача

Количество команд и студентов

Количество команд

2

Количество студентов в команде

Безопасность компьютерных систем

5

Требования к ролям

Безопасность компьютерных систем

- Разработчик x 4
- Руководитель x 1

Список формируемых компетенций

Безопасность компьютерных систем

Универсальные компетенции

- системное и критическое мышление,
- разработка и реализация проектов,
- командная работа и лидерство,
- коммуникация,
- самоорганизация и саморазвитие

Общепрофессиональные компетенции

- Способен применять информационно-коммуникационные технологии, программные средства системного и прикладного назначения, в том числе отечественного производства, для решения задач профессиональной деятельности;
- Способен использовать необходимые математические методы для решения задач профессиональной деятельности;
- Способен использовать языки программирования и технологии разработки программных средств для решения задач профессиональной деятельности;
- Способен осуществлять подбор, изучение и обобщение научно-технической литературы, нормативных и методических документов в целях решения задач профессиональной деятельности;
- Способен разрабатывать и реализовывать политики управления доступом в компьютерных системах;
- Способен обеспечивать защиту информации при работе с базами данных, при передаче по компьютерным сетям.

Сроки реализации

01.02.2025 - 31.08.2025

Выделенные заказчиком ресурсы

-

Приложения

- Описание_ПП_Нормализация_данных_с_LLM_2025 (на сайт УРФУ).docx

Документ сформировал Овечкина Елена Владимировна