

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ))

АННОТАЦИЯ К
РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Анализ больших текстовых данных и текстовый поиск

Направление подготовки: 09.03.01 – Информатика и вычислительная техника

Направленность (профиль): IT-сервисы и технологии обработки данных на транспорте

Форма обучения: Очная

Общие сведения о дисциплине (модуле).

Целями освоения учебной дисциплины являются формирование у обучающихся системных знаний и практических навыков в области обработки, анализа и поиска больших текстовых данных. В рамках курса рассматриваются современные методы и алгоритмы обработки естественного языка (NLP), извлечения информации из неструктурированных текстов, построения поисковых систем и рекомендательных сервисов. Особое внимание уделяется интеграции результатов текстового анализа в бизнес-процессы компании, формулировке практических рекомендаций и адаптации аналитических моделей под потребности заказчика.

Задачами освоения дисциплины являются:

1. формирование теоретических знаний о методах обработки естественного языка, алгоритмах токенизации, лемматизации, векторизации текстов и построении поисковых индексов;
2. освоение практических навыков предобработки и анализа больших текстовых массивов с использованием библиотек машинного обучения;

3. приобретение умений реализовывать алгоритмы машинного обучения для задач классификации, кластеризации, тематического моделирования и извлечения сущностей из текстов;

4. формирование компетенций по адаптации аналитических моделей под бизнес-потребности, преобразованию результатов анализа в бизнес-инсайты и практические рекомендации;

5. освоение инструментов создания аналитических отчетов и визуализации результатов текстового анализа для различных групп заинтересованных сторон.

Дисциплина предназначена для формирования компетенций, необходимых для решения профессиональных задач, связанных с обработкой больших текстовых данных, построением поисковых и рекомендательных систем, а также интеграцией аналитических решений в бизнес-процессы организации.

Общая трудоемкость дисциплины (модуля) составляет 7 з.е. (252 академических часа(ов)).