

РАБОТА С ТЕКСТОВЫМИ ДАННЫМИ

- › Предсказать рейтинг записи в блоге
- › Определить эмоциональный окрас комментария
- › Определить тематику научной статьи
- › Сгруппировать новости по сюжетам
- › Найти слова, похожие по смыслу на данное

ЗАДАЧИ АНАЛИЗА ТЕКСТОВ

- › Выделить все упоминания имён в тексте
- › Построить краткую аннотацию текста
- › Построить модель, отвечающую на вопросы
- › Сгенерировать новый текст, похожий на заданный набор текстов

ОСНОВНАЯ ЗАДАЧА

- › Дан текстовый признак
- › Преобразовать текст в матрицу «объекты-признаки»
- › Описать текст произвольной длины фиксированным числом признаков

- Пример: определение тематики текста (датасет 20newsgroups)

- Пример: определение тематики текста (датасет 20newsgroups)

- › Пример: определение тематики текста (датасет 20newsgroups)
- › Порядок слов не так важен для решения задачи
- › Мешок слов — представление текста в виде счётчиков вхождений

- Основная цель при работе с текстами —
признаковое описание