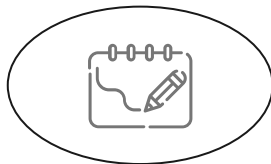


ОПРЕДЕЛЕНИЕ ТОНАЛЬНОСТИ

НЕСТРУКТУРИРОВАННОГО ТЕКСТА

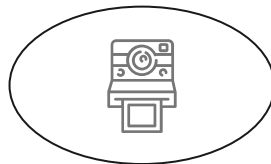
Студент: Казаева Т.А, группа ИУ7-56Б
Научный руководитель: Кострицкий А.С

АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ



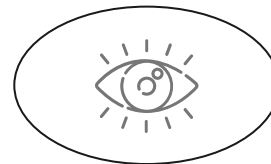
ИНТЕРНЕТ - МАГАЗИНЫ

Получают множество
отзывов на товары каждый
день.



ЛИЧНЫЕ БЛОГИ

Получение личной
характеристики автора,
мониторинг репутации
различных компаний.



НОВОСТНЫЕ ТЕКСТЫ

Выделение мнений об
отношениях между
странами, ситуаций,
третьих лиц

АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

Тональностью текста принято называть эмоциональную оценку, выраженную в неструктурированном тексте по отношению к сущности, и определяемую тональностью составляющих ее лексических единиц и правил их сочетания.

ТОНАЛЬНОСТЬ ТЕКСТА

ПОСТАНОВКА ЗАДАЧИ

АНАЛИЗ

Анализ существующих работ посвященных определению тональности текста.

КЛАССИФИКАЦИЯ

Классификация методов и выделение их основных особенностей

ОЦЕНКА

Выделение достоинств и недостатков методов. Подбор прикладных задач.

КЛАССИФИКАЦИЯ МЕТОДОВ

ИНЖЕНЕРНО - ЛИНГВИСТИЧЕСКИЙ ПОДХОД

Заключается в генерации правил, на основе которых будет определяться тональность текста.

Для реализации таких методов создаются словари оценочной лексики.

Разрабатываются алгоритмы применения лингвистических правил.

КЛАССИФИКАЦИЯ МЕТОДОВ

ПОДХОД НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

Машинное обучение с учителем. Отбирается некоторое количество текстов для обучения и на их основе обучается классификатор.

Машинное обучение без учителя. Подход основан на идее, что наибольший вес в тексте имеют самые часто встречаемые термины.

МЕТОД	ДОСТОИНСТВА	НЕДОСТАТКИ
Метод на основе шаблонов и правил	результат предсказуем	составление словаря оценочной лексики и правил оценки вручную дорого
	возможен более глубокий анализ тональности на уровне высказывания	при узком диапазоне слов в словаре оценочной лексики метод дает неточные результаты
Метод на основе машинного обучения	высокие показатели эффективности	создание обучающей коллекции – это трудоемкий процесс
		алгоритмы плохо переносятся на другую предметную область

ПРИКЛАДНЫЕ ЗАДАЧИ

МЕТОД НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

предлагается использовать для
текстов, в которых чаще всего
оцениваются сущности одной
предметной
области, но в разных ее аспектах. К
таким текстам относятся личные
блоги
и новостные тексты

ИНЖЕНЕРНО – ЛИНГВИСТИЧЕСКИЙ МЕТОД

предлагается использовать для
тех случаев, когда заранее
определены сущности оценивания:
например,
отзывы на определенную вещь в
интернет – магазинах.