

Тональность текста

Асирян Александр

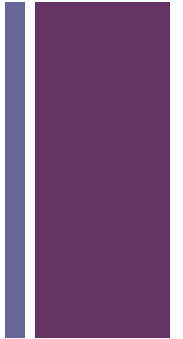


Задача определения тональности текста



- Определение тональности всего текста
- Нахождение фрагментов текста, выражающих оценку
- Определение оцениваемых объектов и их тональности
- Классификация только на положительное и отрицательное
- Ранжирование положительных и отрицательных оценок, нейтральные оценки

Обзор инструментов анализа тональности



- Texterra API
- Eureka Engine
- AlchemyAPI

Texterra API



- Разработана в ИСП РАН
- Решает задачи морфологического, синтаксического и тонального анализа, исправление опечаток, выделение именованных сущностей и т. д.
- Реализована на Java
- Использует Apache OpenNLP и Yandex MyStem
- Определяет тональность с помощью SVM
- Rest API + Web Demo

Texterra API

Тональность



Определяет три вида тональности сообщения (позитивную, негативную и нейтральную)

Русский язык:

- Присутствие эмоциональной окраски
 $F1 = 0.86$
- Полярность эмоциональной окраски
 $F1 = 0.91$

Английский язык

- Присутствие эмоциональной окраски
 $F1 = 0.99$
- Полярность эмоциональной окраски
 $F1 = 0.79$

Texterra API Интерфейс



Next Snippet	Finances	Internet	Health	Customize
--------------	----------	----------	--------	-----------

В Барселоне проходит акция в поддержку единства Испании.

В Барселоне проходит демонстрация против отделения Каталонии от Испании, сообщает Би-би-си.

Verdict

Text is Negative

Domain

Detected Domain

General

Specify Domain

.....

Texterra API Интерфейс



Next Snippet

Finances

Internet

Health

Customize

«Дипан» — это прекрасный образец хорошей современной европейской драмы. Неспешная поначалу история постепенно перерастает в подлинный грандшторм чувств, где острые приправы и острый же нож будут словно такие современные, и в тоже время как древнейшие из детищ человеческой культуры, что живёт в вечном колебании маятника между насилием и любовью.

Verdict

Text is Positive

Domain

Detected Domain

General

Specify Domain

Autodetect

Eureka Engine



- Определение языка
- Определение тональности (объектное / безобъектное)
- Классификации текстов
- Определение и классификация именованных сущностей
- Морфологический анализ
- Rest API + Web Demo

Eureka Engine

Тональность



- Определяет три вида тональности сообщения
- Тональность относительно заданного объекта тональности как в пределах одного предложения, так и усредненная по всему документу
- Объектом может быть как отдельное слово, так и словосочетание
- Средняя точность по трем видам тональности около 86%

Eureka Engine Интерфейс



Объектный анализ

Тональность документа:

позитив 0

негатив 0.18

В Барселоне проходит акция в поддержку единства Испании. В Барселоне проходит демонстрация против отделения Каталонии от Испании, сообщает Би-би-си.

Безобъектный анализ

Тональность документа:

позитив 1

негатив 0

«Дипан» — это прекрасный образец хорошей современной европейской драмы. Неспешная поначалу история постепенно перерастает в подлинный грандшторм чувств, где острые приправы и острый же нож будут словно такие современные и в тоже время как древнейшие из детищ человеческой культуры, что живёт в вечном колебании маятника между насилием и любовью.

AlchemyAPI



- Приобретена IBM
- Извлечение ключевых слов, именованных сущностей, отношений и т. д.
- Объектный и безобъектный анализ тональности
- Использует машинное обучение, глубинное обучение
- Rest API + Web Demo

AlchemyAPI Интерфейс

Безобъектный / объектный анализ

Text

URL

В Барселоне проходит акция В поддержку единства Испании.

В Барселоне проходит демонстрация против отделения Каталонии от Испании, сообщает Би-би-си.

Overall Sentiment

Positive 0.00

Targeted Sentiment

Барселона

>

!

 Cannot locate keyphrase

Overall Sentiment

Positive 0.00

Targeted Sentiment

Барселоне

>

Барселоне

Positive 0.99