

Лингвистически е технологии в Интернете

Лилия
Рамазанова
622-1

Лингвистические технологии в Интернете

Лингвистические технологии помогают найти информацию, обработать и сделать с ней что-то полезное.

Данные технологии делятся примерно на 5 областей.

Технологии поиска информации

1. Поисковик ищет документы, наиболее похожие на ваш запрос.
2. Определяет намерения пользователя и выдает вам второй блок документов (рекламы).

Пример работы технологии поиска информации

Поисковик	Google
Запрос	Илекс
Документы 1 поиска	9 документов описывающих сам Илекс.
Документы 2 поиска	Тф Илекс с маршрутом и часами работы, подбор похожих запросов (5 цветочных магазинов)
Дополнительно:	6 изображений

Технологии извлечения информации

Подходы к определению тональности:

1. Изучение слов, указывающих на тональность
2. Анализ массива примеров.
3. Комбинированный

Пример технологии извлечения информации

Поисковик	Google
Запрос	Год рождения Путина
Результат	7 октября 1952 г. (64 года)
Запрос	Сколько сейчас время
Результат	21:56 пятница, 16 декабря 2016 г. (GMT+3) Москва

Машинный перевод

1. Полностью автоматизированный машинный перевод
2. Автоматизированный машинный перевод при участии человека.
3. Перевод, осуществляемый человеком с использованием компьютера

Сравнение автоматизированного машинного перевода с ручным переводом

Оригинал текста	PROMT	Ручной перевод
<p>We clawed, we chained our hearts in vain We jumped, never asking why We kissed, I fell under your spell A love no one could deny</p> <p>Don't you ever say I just walked away I will always want you I can't live a lie, running for my life I will always want you</p> <p>I came in like a wrecking ball I never hit so hard in love All I wanted was to break your walls All you ever did was break me Yeah, you wreck me</p>	<p>Мы хватали, мы приковали сердца цепью напрасно Мы подскочили, никогда не спрашивая почему Мы поцеловались, я подпал под Ваш период Любовь никто не мог отрицать</p> <p>Не делайте Вас, когда-либо говорят, что я просто ушел Я буду всегда хотеть Вас Я не могу вести двойную жизнь, спасаясь бегством Я буду всегда хотеть Вас</p> <p>Я вошел как груша для сноса зданий Я никогда не совершал нападки так трудно любящий Все, что я хотел, должно было сломать Ваши стены Все, что Вы когда-либо делали, было, ломают меня Да, Вы разрушаете меня</p>	<p>Напрасно мы вцепились друг в друга и прикипели душой. Мы сошлись без лишних вопросов: Поцеловались - и я попала под действие твоих чар. Не было никаких сомнений - это любовь.</p> <p>И не смей говорить, что я ушла сама, Я всегда буду в тебе нуждаться. Жизнь во лжи – не для меня, я спасаюсь бегством, Но я не перестану в тебе нуждаться.</p> <p>Я ворвалась в твою жизнь с разрушительной силой. Прежде я никогда так сильно не влюблялась. Я хотела лишь разрушить стены, что ты воздвиг, А ты просто взял и сломил меня, Да, ты меня уничтожил...</p>

Генерация текста и речи

Например, на основе концепции вопросно-ответных систем. До сих пор проводятся исследования где пытаются приблизить ответы компьютера к человеческой речи.

Таким способом ученые пытаются определить, может ли машина мыслить. К ним можно отнести (пусть и спорный) тест Тьюринга.

Генерация текста на примере диалога

Бот PARRY

Человек

Человек: Вы связаны с мафией?

PARRY: Вы знаете, они меня знают.

Человек: Каким образом они знают о вас?

PARRY: Они шпионят за мной.

Человек: Зачем они это делают?

PARRY: Они хотят получить фальшивые доказательства, чтобы подставить меня.

Человек: Почему они хотят сделать это с вами?

PARRY: Они хотят разделаться со мной.

Человек: Почему?

PARRY: У меня есть изобличающие улики на мафию.

Человек: Вы уверены во всём этом?

PARRY: Вы мне не верите?

1979г.

Распознавание речи и общение с машинами

1. Распознавание говорящего
(верификация и идентификация
диктора)
2. Распознавание речи

Google Cloud Speech API

Речь	Google Cloud Speech API
Hello my name is Lilia	hello my name is Leah
I want to be an interpreter	I want Bee and PuppyCat
I like to drink coffee	I like to drink coffee
I love Miley Cyrus	I lost my phone