

CMU SPHINX

Подготовил студент группы КМБО-02-13

Голубев М.А.

Что такое Sphinx

- CMU Sphinx — крупнейший на данный момент проект по распознаванию человеческой речи с бесплатной лицензией
- В инструментарий входят:
 - Pocketsphinx — небольшая быстрая программа, обрабатывающая звук, акустические модели, грамматики и словари
 - библиотека Sphinxbase, необходимая для работы Pocketsphinx
 - Sphinx4 — гибкая библиотека для распознавания
 - Sphinxtrain — программа для обучения акустическим моделям

Термины

- Акустическая модель — отвечает за сопоставление звуку произнесенной фонемы. Позволяет оценить распознавание речевого сегмента с точки зрения схожести на звуковом уровне. Для каждого звука изначально строится сложная статистическая модель, которая описывает произнесение этого звука в речи

Термины

- Словарь — это файл, в котором сопоставлены лексемы и фонемы (слово и его транскрипция). Необходим для преобразования фонем, распознанных акустической моделью, в лексемы

- Примеры:

калькулятор — k a ll k u ll ja t ay r

компьютер — k a m pp j ju tt i r

разделить — r ay z dd i ll ii tt

умный — uu m n ay j

дом — d oo m

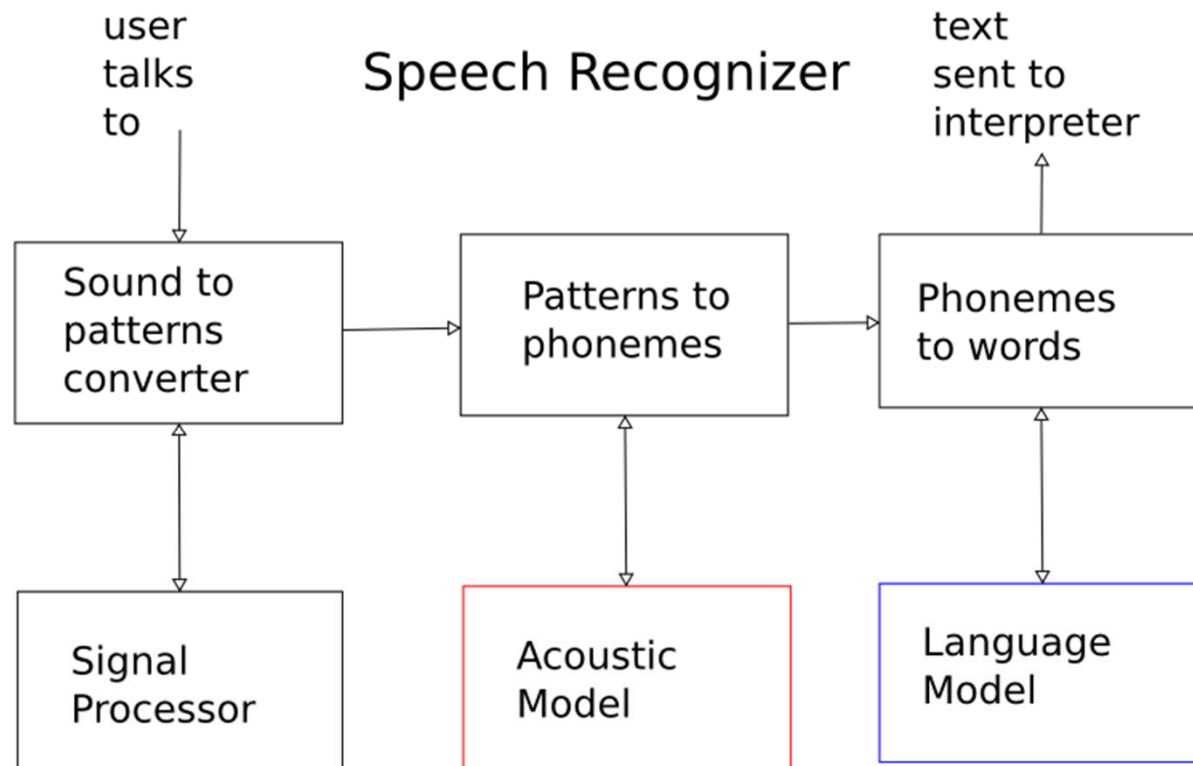
Термины

- Грамматика — формальные правила, которые описывают простые правила построения предложений. Лексемы, полученные на предыдущем шаге, пытаются сопоставиться с грамматикой

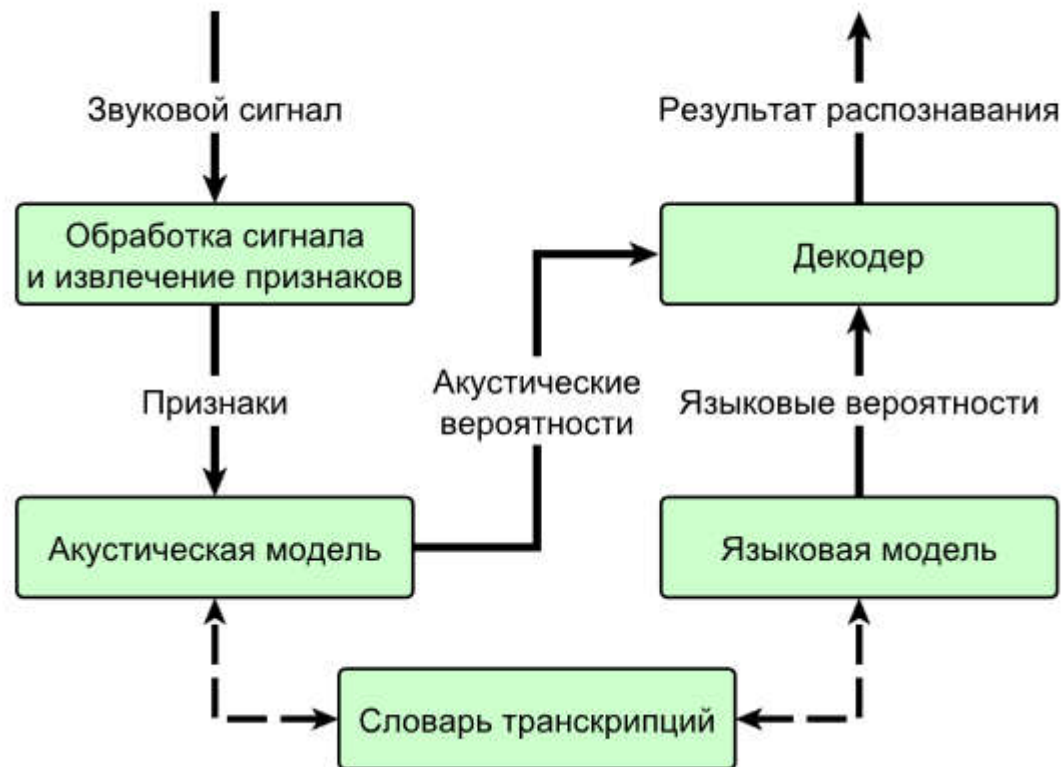
Термины

- Языковая модель — это статистическая модель языка. Она описывает вероятности слов и их комбинаций. Таким образом распознавание лексем — это максимизация правдоподобности распознанной фразы

Термины

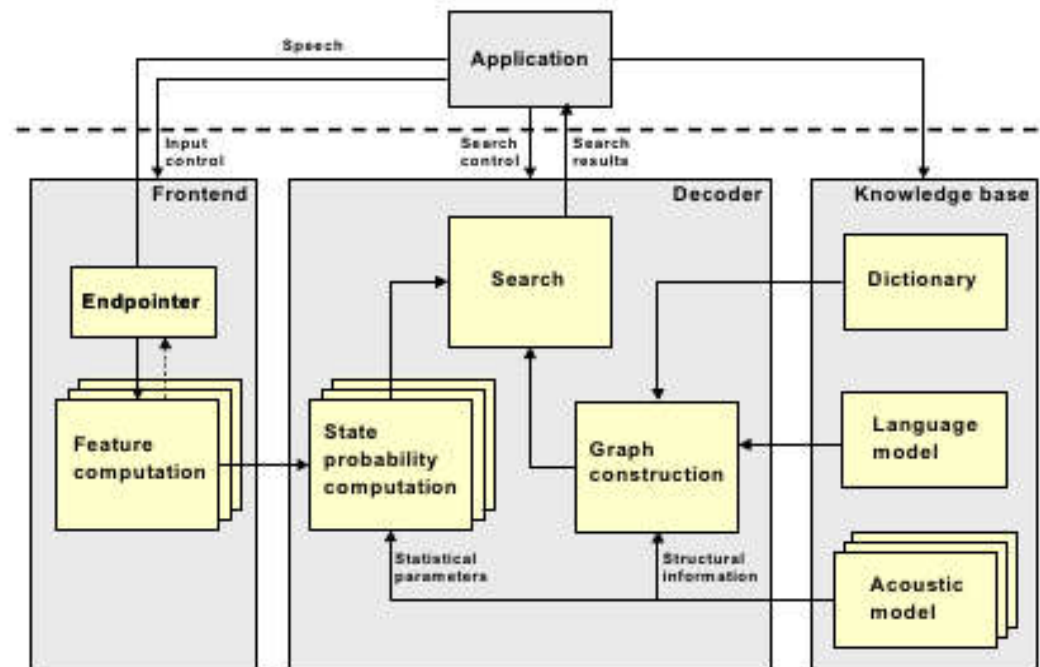


Как это работает



Как это работает

- Архитектура Sphinx



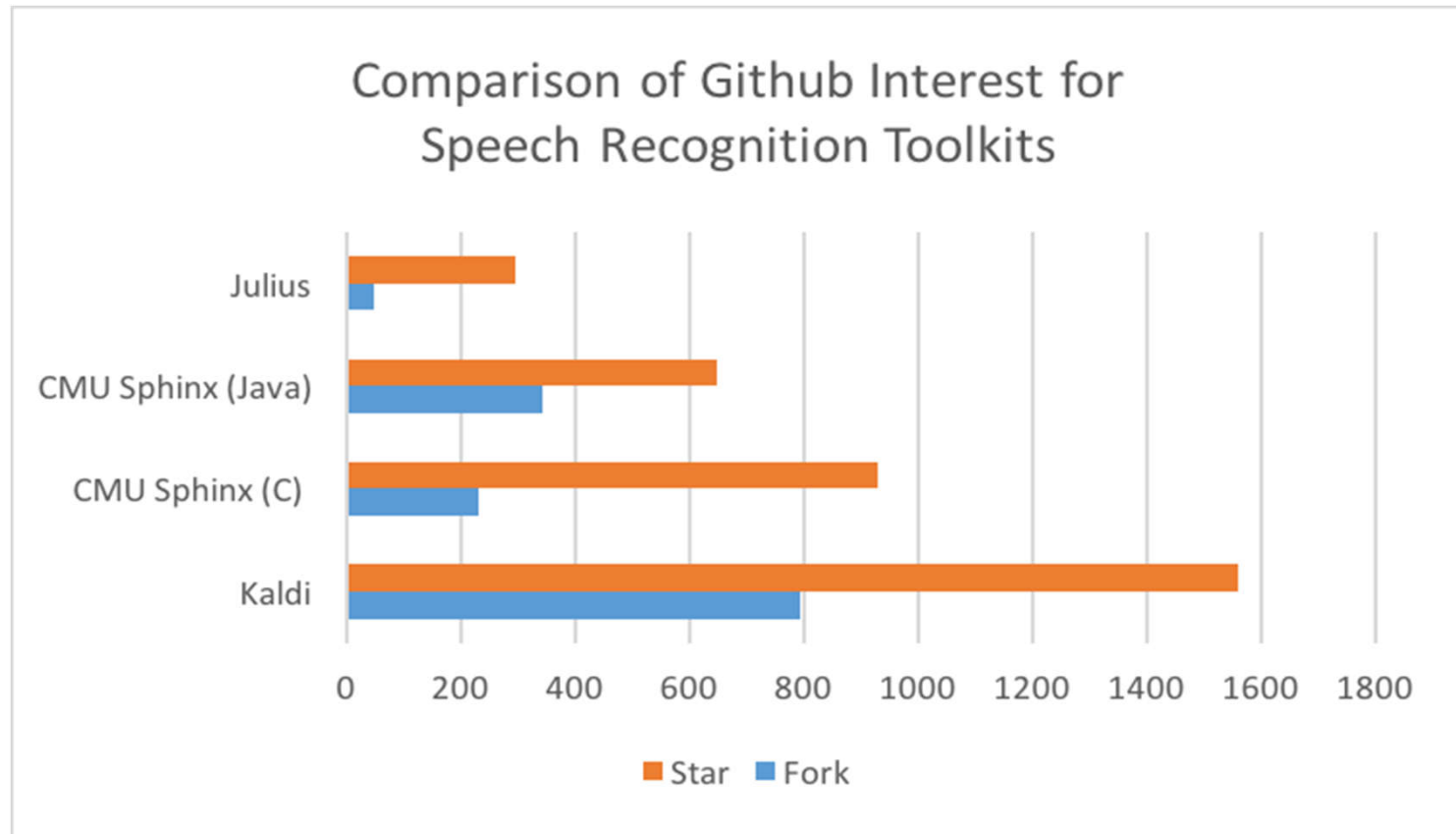
Другие проекты

- Kaldi
- HTK
- Julius
- ISIP
- Google Speech Recognition API
- Yandex Speech Kit

Сравнение проектов

Toolkit	Programming languages	Development activity	Tutorials and examples	Community	Trained models
CMU Sphinx	Java, C, Python, others	+++	+++	+++	English plus 10 other languages
Kaldi	C++, Python	+++	++	+++	Subset of English
HTK	C, Python	++	+++	++	None
Julius	C, Python	++	++	+	Japanese
ISIP	C++	++	++	+	Digits only

Сравнение проектов



Применение в «умном доме»

- Умный дом — система домашних устройств, способных выполнять действия и решать определенные повседневные задачи по командам человека без его участия



Применение в «умном доме»

- Приложения с использованием Sphinx могут реализовывать систему управления умным домом с голосовым интерфейсом



Применение в «умном доме» с использованием собственной грамматики

- Качаем акустическую модель с репозитория проекта CMU Sphinx
- В файл словаря добавляем слова, которые входят в команды
- В файле грамматики описываем предложения, которые будут распознаваться
- ???
- PROFIT!

Особенности использования собственной грамматики

Плюсы:

- Выдает хорошие результаты на малом количестве команд
- Разработчик может создавать удобные ему конструкции

Минусы:

- Говорить нужно членораздельно
- Предложения должны строго соответствовать грамматике

Пример грамматики в формате JSGF

```
grammar exampleGrammar;  
public <exampleGrammar> = (<object><action>);  
<object> = (Компьютер | Свет | Чайник);  
<action> = (Включить | Выключить);
```

Применение в «умном доме» с использованием языковой модели

- Качаем акустическую и языковую модели с репозитория проекта CMU Sphinx
- В файл словаря добавляем слова, которые входят в команды
- ???
- PROFIT!

Особенности использования языковой модели

Плюсы:

- Предложения могут строиться с ошибками
- Пользователь не ограничен малым набором предложений

Минусы:

- На задачах, где не требуется большее количество команд, распознает фразы хуже, чем приложение с использованием грамматики

Заключение

- Sphinx — мощное средство для задач распознавания речи в оффлайне
- Open-source, постоянно развивается. Активно используется в коммерческих и некоммерческих проектах (Chatter, Misterhouse, Nightingale)
- Можно использовать на разных платформах (Rocketsphinx) и писать приложения на разных языках программирования
- Не вызывает особых сложностей в использовании

ИСТОЧНИКИ

- Меденников И.П. «Методы, алгоритмы и ПС распознавания русской телефонной спонтанной речи», 2016
- Федяев О. И., Бакаленко В. С., Савкова Д. Г. «Речевой интерфейс для интеллектуализации ввода исходного кода программ», 2015
- <https://habrahabr.ru/post/237589/>
- <https://habrahabr.ru/post/167479/>
- <http://cmusphinx.sourceforge.net/>
- <https://svds.com/open-source-toolkits-speech-recognition/>