

Практическое задание №2. Осень 2024

Постановка задачи

Целью работы является разработка метода, позволяющего выделять припевы из текстов песен. В качестве набора данных используются тексты треков из Яндекс.Музыки.

- В качестве входа метода будут даны строки песни (уже разбитые в виде списка), а также название трека (необходимость его использования остаётся на ваше собственное усмотрение).
- Необходимо вернуть список пар индексов для каждого припева в формате **(начало, конец)**, где **начало** и **конец** — индексы строк текста.

Некоторые тонкости выделения припевов:

- **Минимальная длина припева** составляет не менее двух строк (пример 2).
- **Чёткие повторяющиеся припевы**: если припев повторяется без изменений, каждый повтор нужно выделить как отдельный элемент (пример 2).
- **Частично совпадающие припевы**: если припевы или их части совпадают лишь частично (например, первые две строки совпадают, а следующие две отличаются), то такие припевы не разбиваются и их следует считать одним цельным припевом (пример 3).
- **Постприпевы и предприпевы не должны учитываться**: обычно это несколько строчек из реального припева, но не использующие его целиком (пример 4).
- **Не всегда первая строка разных припевов полностью совпадает**, следует учитывать это при выделении припевов (пример 5).

Пример 1:	Пример 2:
<p>Тёмный, мрачный коридор Я на цыпочках, как вор Пробираюсь, чуть дыша Чтобы не вспугнуть Тех, кто спит уже давно Тех, кому не всё равно В чью я комнату тайком Желаю заглянуть Чтобы увидеть</p> <p>Как бессонница в час ночной Меняет, нелюдимая, облик твой Чьих невольница ты идей? Зачем тебе охотиться на людей?</p> <p>Крестик на моей груди На него ты погляди Что в тебе способен он Резко изменить? Много книжек я читал Много фокусов видал Свою тайну от меня Не пытайся скрыть! Я это видел!</p> <p>Как бессонница в час ночной Меняет, нелюдимая, облик твой Чьих невольница ты идей? Зачем тебе охотиться на людей?</p> <p>Очень жаль, что ты тогда Мне поверить не смогла В то, что новый твой приятель Не такой, как все!</p>	<p>На твоих диванах сходим с ума мы Мы как две мадамы после нирваны Мы так летали, что не заметили твою мать Её, и она сказала, что я просто</p> <p>Мама Люба, давай, давай, давай Мама Люба, давай, давай, давай</p> <p>Мама Люба, давай, давай, давай Мама Люба, давай, давай, давай</p> <p>Мама Люба, давай, давай, давай Мама Люба, давай, давай, давай</p> <p>Наша-наша кроватька делала шик-шик Я твоё пианино, а ты мой настройщик Мы так летали, что не заметили твою мать Ага, и она сказала, что я просто</p> <p>(У) Мама Люба, давай, давай, давай Мама Люба, давай, давай, давай</p> <p>Мама Люба, давай, давай, давай Мама Люба, давай, давай, давай</p> <p>Мама Люба, давай, давай, давай Мама Люба, давай, давай, давай</p> <p>Мама Люба, давай, давай, давай Мама Люба, давай, давай, давай</p> <p>Мама Люба, давай, давай, давай (У) Мама Люба, давай, давай, давай</p>

<p style="text-align: center;">Пример 3:</p> <p>Я падаю вниз, я срываюсь со всех краёв Хотел выше птиц, но не попадаю в дверной проём Я опустился... Куда бы с тобою не смог Я распустился, как ядовитый цветок</p> <p>Я так хотел, чтоб ты Была счастлива со мной Но я — отрицательный герой Ведь я так хотел, чтоб ты Была счастлива со мной Но я — отрицательный герой</p> <p>Я не хочу, чтоб ты сомневалась в себе Я заплачу за все твои слёзы, поверь Я хуже тебя, мои обещания дешевле твоих, а-а Ты меня не прощай, я шум... И я стих</p> <p>Я так хотел, чтоб ты Была счастлива со мной Но я — отрицательный герой Ведь я так хотел, чтоб ты Была счастлива со мной Но я — отрицательный герой, ооа</p> <p>Я так хотел, чтоб ты Была счастлива со мной Но я — отрицательный герой Я так хотел, чтоб ты Была счастлива со мной Но я — отрицательный герой, ооао</p>	<p style="text-align: center;">Пример 4:</p> <p>Кто, не знаю, распускает слухи зря Что живу я без печали и забот Что на свете всех удачливее я И всегда, и во всём мне везёт</p> <p>Так же, как все, как все, как все Я по земле хожу, хожу И у судьбы, как все, как все Счастья себе прошу Счастья себе прошу</p> <p>Вы не верьте, что живу я, как в раю И обходит стороной меня беда Точно так же я под вечер устаю И грущу, и реву иногда</p> <p>Так же, как все, как все, как все Я по земле хожу, хожу И у судьбы, как все, как все Счастья себе прошу Счастья себе прошу</p> <p>Жизнь меня порой колотит и трясёт Но от бед известно средство мне одно В горький час, когда смертельно не везёт Говорю, что везёт всё равно</p> <p>Так же, как все, как все, как все Я по земле хожу, хожу И у судьбы, как все, как все Счастья себе прошу</p> <p>И у судьбы, как все, как все Счастья себе прошу Счастья себе прошу</p>
<p style="text-align: center;">Пример 5:</p> <p>Сколько сказано слов о любви, сколько спето Эта тема задета каждым поэтом Но я, не желая им уподобляться Предпочитал стихам совокупляться Но какое-то чувство в момент нашей встречи Озарило глаза мои нежно-зелёным Пошлость потеряла дар речи Наполнив мне лёгкие чем-то влюблённым</p> <p>Сердце в груди как только начало биться Стремиться быть с тобой рядом Ощущая тепло любимого взгляда Лучшее, что могло со мною случиться</p> <p>Я устал от сучек, лжи, запаха псины Ты - счастливейший случай быть самым красивым Так устав от зимы, я нашёл в тебе лето Я вышел из тьмы – значит, я иду к свету В этом мире жестоком, где нежность как слабость Где, чтобы выжить, ты вынужден биться Я набрался неслыханной наглости Я позволил себе влюбиться</p> <p>Сердце в груди продолжает биться Стремиться быть с тобой рядом Ощущая тепло влюблённого взгляда Лучшее, что могло со мною случиться</p> <p>Сквозь покрытые дымом конопляные степи Я прошёл, переплыл океаны спиртного Искушён всеми бесами, но обошёл их А любовь оказалась сильнее их намного И словно голодный, капризный мальчишка Узнавший, как просто унять этот голод Нашёл в тебе повод быть самым счастливым Если для этого вообще нужен повод Сердце в груди стучит словно молот И голым засыпать с тобой рядом Ощущая тепло влюблённого взгляда Лучшее мне и не надо иного</p> <p>Сердце в груди стучит словно птица Стремиться быть с тобой рядом Ощущая тепло любимого взгляда Лучшее, что могло со мною случиться Лучшее, что могло со мною случиться</p>	

Решение задачи

Тестирование

На личной странице (practicum.tpc.ispras.ru/submissions/chorus) находится статистика со всеми результатами в т.ч. результатами последнего тестирования (дата, метрики качества).

На странице practicum.tpc.ispras.ru/results доступны результаты всех участников. Таблица обновляется раз в неделю по средам в 00:00.

Практические аспекты

Решения должны быть написаны на языке Python (версия 3.10.7). Можно использовать все стандартные библиотеки, а также:

- pandas==1.5.1
- pymystem3==0.2.0
- regex==2023.10.3
- gensim==4.2.0
- tensorflow==2.10.0
- tensorflow-addons==0.18.0
- scikit-learn==1.1.3
- onnxruntime==1.13.1
- onnxruntime-tools==1.7.0
- torch==2.4.1+cpu
- torchvision==0.19.1+cpu
- torchaudio==2.4.1+cpu
- nltk==3.9.1 и все его пакеты
- numpy==1.26.4
- Levenshtain==0.26.0

В случае необходимости использования дополнительных библиотек, сообщите об этом организаторам практикума (библиотеки будут добавлены для всех студентов). Доступ в интернет на проверяющей машине закрыт.

Загрузка решения

Загружаемый файл должен представлять собой zip архив с любым именем. Архив должен обязательно содержать:

- Решение в файле `solution.py`. В файле должен быть класс `Solution`, содержащий метод `detect(self, tracks: List[Tuple[List[str], str]]) -> List[Set[Tuple[int, int]]]`, который получает на вход список пар строк текста, название трека и возвращает список множеств границ припевов. Границы припева задаются как `[start, end]` (внимание, правая граница **включена**).
- Описание применяемых методов в файле `description.txt`. Пожалуйста, напишите подробное описание, какие методы использовались.
- Все используемые ресурсы, необходимые для корректной работы метода.
- Код, позволяющий выполнить обучение модели (при использовании машинного обучения) и инструкции по его запуску. Инструкции должны быть достаточно подробными для возможности воспроизведения модели без дополнительных консультаций с автором метода. Отсутствие таких инструкций будет приравниваться к невозможности воспроизведения. При решении можно использовать дополнительные ресурсы для обучения, но они должны быть доступны организаторам в момент воспроизведения решения.

Пример решения, возвращающего пустые результаты:

```
from typing import List, Tuple, Set

class Solution:
    def detect(self, tracks: List[Tuple[List[str], str]]) -> List[Set[Tuple[int, int]]]:
        return [set() for lines, title in tracks]
```

Примеры треков с размеченными припевами доступны по [ссылке](#). Примеры данных даются исключительно в ознакомительных целях для выполнения данного задания. Использование их для других целей запрещено.

Ограничения

- Каждую неделю можно послать не более 10 решений.
- Внимание! Итоговое тестирование будет проводиться на последнем загруженном решении.
- Время тестирования одного решения не должно превышать 30 минут.
- Размер загружаемого архива не должен превышать 100Мб.
- На проверяющей машине доступно 16 Гб оперативной памяти и 8 CPU.

Внимание: при использовании PyTorch или Tensorflow (или другого фреймворка машинного обучения, использующего многопоточность) явным образом установите ограничение на максимальное количество потоков.

- [set num threads](#) для PyTorch
- [set inter op parallelism threads](#) для Tensorflow

Оценка качества

Для оценки задания используется micro-averaged F_1 -мера:

$$P = \frac{|predicted \cap expected|}{|predicted|}, \quad R = \frac{|predicted \cap expected|}{|expected|}, \quad F_1 = \frac{2PR}{P + R},$$

где *predicted*, *expected* – множества предсказанных и ожидаемых границ припевов.