

1. StackOverflow map
2. Event extraction 2

1. StackOverflow Map

Цель:

Отрисовать граф тегов StackOverflow.

Задачи:

- Разобраться с работой алгоритма tSNE (читать статьи и документацию на английском).
- Научиться укладывать граф в несколько десятков тысяч вершин меньше чем за сутки.
- Сделать веб-интерфейс для просмотра графа.
- Научиться обновлять граф свежими данными с data.stackexchange.com

1. StackOverflow Map

high energy experiment
(hep-ex)

high energy phenomenology
(hep-ph)

nuclear experiment
(nucl-ex)

high energy lattice
(hep-lat)

high energy theory
(hep-th)

general relativity
quantum cosmology
(gr-gc)

astrophysics
(astro-ph)

mathematical physics
(math-ph)

1. StackOverflow Map

С чем придется работать:

- Python
- Возможно C++
- Немножко JavaScript

2. Event extraction from natural texts 2

Цель:

Доработать модель извлечения событий из текстов. Внедрить её в существующую систему сбора данных.

Задачи:

- Изучить существующее решение.
- Попробовать использовать другие библиотеки разбора и анализа текста.
- Сделать возможность дообучения модели.
- Сделать поддержку нескольких языков.

2. Event extraction from natural texts

Текст:

Apple today announced [Swift Playgrounds for the iPad](#), a new project that aims to teach kids to code in Swift.

Из статьи на techcrunch.com

Событие:

Subject: Apple Inc.

Action: announce

Object: Swift Playgrounds

2. Event extraction from natural texts

С чем придется работать

- Python
- JavaScript
- Natural Language Processing
- SQL

Контакты

arkady.kalakutsky@jetbrains.com

Аркадий Калакуцкий