

Педагогика, информатика и обработка естественного языка

Михаил Пожидаев

5 сентября 2019 г.

Обработка материалов для педагогических приложений

1. Обработка текстов на основе грамматик.
2. Исследование применения sentence embeddings для русского языка.
3. Восстановление структур PDF-документов.

Обработка текстов на основе грамматик

Задача: повысить качество извлечения информации из текста на русском языке с применением теории формальных грамматик и утилиты Томита-парсер компании Яндекс.

В ходе выполнения работы исполнитель:

- ▶ получит опыт применения теории формальных грамматик на практике;
- ▶ ознакомится с особенностями автоматизированной обработки текста на русском языке;
- ▶ ознакомится с инструментарием, предлагаемым компанией Яндекс для обработки русского языка.

Исследование применения sentence embeddings для русского языка

Задача: проанализировать возможности и ограничения существующих инструментов построения sentence embeddings для русского языка.

В ходе выполнения работы исполнитель:

- ▶ познакомится с применением машинного обучения в обработке русского языка;
- ▶ познакомится с инструментариями компании Facebook в сфере обработки текста на естественном языке;
- ▶ получит опыт применения языка Python для решения задач, связанных с машинным обучением.

Восстановление структур PDF-документов

Задача: провести построение и обучение нейросети для массовой обработки статей в фонде научной библиотеки ТГУ.

В ходе выполнения работы исполнитель:

- ▶ получит опыт применения TensorFlow на практике;
- ▶ сделает вклад в развитие родного ТГУ.

Спасибо за внимание!

Веб-сайт: <http://marigostra.ru/>

E-mail: msh@luwrain.org