

12.04.2018

Саша рассказал о завершении лингвистических работе над word2vec моделями для всех подкорпусов из киберленинки.

Настя рассказала о своей работе над readability. Нужно подумать над скриптом, который будет считать над всем корпусом TTR и readability.

Аня рассказала о посчитанных мерах для каждого подкорпуса и текущей и грядущей работе над выступлением в семинаре в Вороново.

Курсовая!

Над чем работаем дальше? Имеем два варианта: 1) составлять словарь биграмм и коллокаций; 2) писать базовую курсовую.

Две статьи? Синтаксическую и про коллокации. Точно последний вариант!

Как работать с синтаксической разметкой? Точно нужно что-то сделать с генетивными цепочками. Посмотреть таблицу Насти, где указаны потенциальные ошибки, в том числе синт-ие.

Рабочие названия статей:

“Извлечение устойчивых сочетаний/коллокаций из текстов научного стиля”.

“Диагностирование синтаксических ошибок в студенческих академических текстах”.