Доклад «Инспектор для коллекции» на семинаре НУГ

Иван Торубаров, Екатерина Герасименко

Сейчас для пользователей доступен инспектор, считающий некоторые характеристики по одному тексту (<http://linghub.ru/inspector/>). Для исследований, в частности, для сравнения множества текстов из разных коллекций, было бы хорошо иметь подобный инструмент для обработки коллекции текстов. И мы решили его создать.

Три проблемы:

1) У каких метрик абсолютное значение по всем текстам, а у каких – среднее (+ std)?

2) Какой POS-тэггер выбрать, чтобы было быстро, но качественно?

3) Как сделать удобную кнопку в brat?

Мы решили первую, разбив все метрики на две категории. Некоторые метрики попали в обе, что значит, что было решено выводить оба значения. Когда будет решен вопрос скорости и удобства пользования (кнопка + несколько параметров, возможно), инспектор коллекции будет запущен для работы.