Проект “ANTIDICT”

База морфологических структур, допустимых в современном русском языке

Студенты:
Алексей Доркин
Владислава Смирнова
Антон Вахранев

Куратор: Варвара Магомедова

Что мы успели сделать по намеренным искажениям

– Определились с формальными критериями

– Составили и предобработали выборку

– Обучили модель находить намеренные искажения

– Расширили обучающую выборку за счет слов, которые выявил fasttext, и обучили модель с новой выборкой

Тетрадка по намеренным искажениям:
https://github.com/wksmirnowa/antidict/blob/master/Jupyter%20Notebooks/deforming_words_classification.ipynb

Примеры намеренных искажений, выявленных с помощью fasttext и линейной регрессии

ПРЫХОДИ	йооолочки	вэнтыляцыю
ИрландЭц	слуцицца	Павярнула
шикаладной	шоколядного	паркувалися
кыевлянами	бутылошного	Птючка
маааааинькая	офысный	женсссчины
Шшшшшмитана	сднемражденя	звИзды
питоНцы	Прахадите	рябитёнок
гуманитариеФЪ	мэлицэонэры	смартфончЕГ
шпашыба	фодкафф	пейсательныцы
ПЕАНЕР	сабакы	
антиресуюцца		
ржамши		

Что мы успели сделать по экспрессивным формам

– Определились с формальными критериями

– Составили обучающую выборку

Что мы успели сделать по экспрессивным формам

Примеры экспрессивных слов, которые мы вручную нашли в датасете:

справулечек
любопытнеько
мегадоклад
хняшечка
особнячково
совятки
кисуля
скрытненько
сламечка
гринписьки
пачпортинка
сякуха
сверхвипов
евреюшка
замерзайкой
леляночка
заколдовка
зачинателец
галюники

Что у нас в процессе

– Разрабатываем схемы подсчета inter-annotator agreement: Krippendorff's alpha

– Собираем слова для расширения выборки по намеренным искажениям

– Обучаем модель находить экспрессивные формы

– Пишем финальную статью

Что нам можно сделать в будущем

– Опробовать другие модели и классификаторы на англицизмах, намеренных искажениях и экспрессивных формах

– Поработать со словами с креативной морфологией (примеры: *Психургия, Масквaляндия, натарфаретил, пьятнеца, Ляпатуська*)

Ссылки

Тетрадка по намеренным искажениям:
https://github.com/wksmirnowa/antidict/blob/master/Jupyter%20Notebooks/deforming_words_classification.ipynb
ТЗ:
https://docs.google.com/document/d/1fpn72q8bKqhFnCaTmbqGZEtWS6WcXIP9WC1VpviAGEc/edit
Trello: https://trello.com/b/XEnCTnHj/antidict
Репозиторий: https://github.com/wksmirnowa/antidict