

19.04.2018

Текст курсовой!

Встреча была посвящена в основном решению задачи с предложением замен и синтаксису. Два модуля в нем: поиск ошибочных коллокаций и предложение замен. Алгоритм для второй готов и работает хорошо. Первый модуль работает не очень хорошо: много надежд на семинар в Вороново, где могут подсказать, как решить эту проблему. Спорный вопрос: нужно ли вообще использовать машинное обучение для первой задачи? Мало данных, сама задача сложная для МО (т.к. не на чем обучить и подтекст алгоритм не считает). Можно работать вместо этого с низкочастотными величинами и с частотами в целом (странности в распределении). Можно смотреть на совпадение биграмм (или n-грамм) с биграммами из эталонного корпуса.

Маша: есть готовый код, ловящий генитивные цепочки, но не без проблем. Надо искать подряд идущие генитивные цепочки из существительных. (а ещё у нас нет задачи делать 100% попадание, мы можем давать “мягкие” советы - наше дело обратить внимание на вопросы стиля).

Нужно еще подумать о синтаксических ошибках (отдельная встреча). Например, сочинительные группы, плеоназмы, деепричастия (но кореферентность сложно контролировать; можно ловить по косвеннопадежным формам “я” (“меня”, “мной” и т.д.)), анафоры (например, к чему относится в тексте “это”, но как это формализовать? По последнему слову в предыдущем предложении?).

Настя: надо узнать О. Ляшевской про ридабилити (формулы и параметры).