

Встреча 15.02.2018

Организационные вопросы: обсуждение результатов недели: код для работы с парсингом архива киберленинки Саши (ближайшая задача - готовый корпус по лингвистике), итоги начала работы с синтаксической разметкой Ани и Маши. Ане нужно выделить от однограмм до шестиграмм внутри каждого домена. + коллокации без учета синтаксиса. Настя собрала из КРУТ, финских и американских собраний ошибок типологию ошибок.

Есть минимальная задача - побочный продукт: некоторые словари на основе всего корпуса или по поддоменам. Это словарь научных выражений, который можно потом будет создать.

Поставили вопрос о том, как будет выглядеть корпусной менеджер. Пока ответа нет (мы откладываем, но помним и думаем).

Списки коллокаций нужны для решения проблем: сможем что-то предлагать (а можно на основе частотности)