У РКИ уже готова часть работы: они нашли списки литературы крупнейших вузов. Они нашли самые упоминаемые тексты, они работают с вордовскими файлами. Они их чистят (от цитат классиков и списков литературы, например). У них есть три учебника по политологии и три по международным отношениям.

РКИ интересуют только политология и международные отношения. Возможно, они подключатся на ещё какие-то темы.

РКИ интересует частотность: так как лексические минимумы строятся по тематическим группам и по частотности. Союзные средства, вводные слова, предложные конструкции.

Может быть перекос в отношении некоторых слов из-за маленького размера корпуса. Нужно сделать общенаучный корпус, относительного которого сравнивать малый корпус РКИ. Так можно выявлять и отделять частотные общенаучные слова и какие относятся к отдельным доменам.

Внутри каждой дисциплины нужно следить за тем, что включаем. Нельзя включать только одну область дисциплины: нельзя только одно банковское дело добавлять. РКИ предлагает страховое дело.

Нужно посмотреть поднауки домена в киберленинке (и не только) и оценить их. Важно оценить жанры: ревью (рецензия, обзорная статья), научная статья; должно быть равновесное количество статей разного жанра. Нужно оценить по авторитетности и по импакт-фактору: «Белый список журналов». Однако нельзя ограничиваться журналами, нужно будет искать вручную авторитетные учебники и монографии.

Копотев и Киселева против переводных учебников и статей из-за непредсказуемости качества.

Монографии и статьи (по политологии и международным отношениям) мы получим от РКИ.

2 задачи: а) выяснить у РКИ какие журналы и статьи брать по каким годам. Какие монографии искать?

б) как автоматизировать удаление: что именно удалять. Кроме того, есть работники, которые делали КРУТ. Они могут поделиться тем, как они удаляли цитаты и т. п. Лучше удалять больше, чем меньше (лучше убрать Вольтера и авторский текст, чем не убрать хоть что-то из Вольтера). Источники не обязательно выкидывать, нужно складывать их в отдельный файл, откуда можно добывать наиболее авторитетных и частотных авторов.