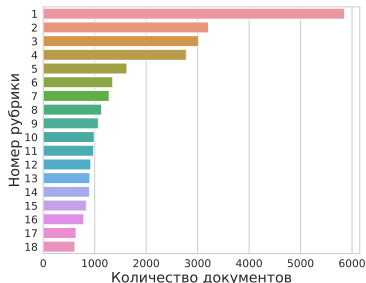
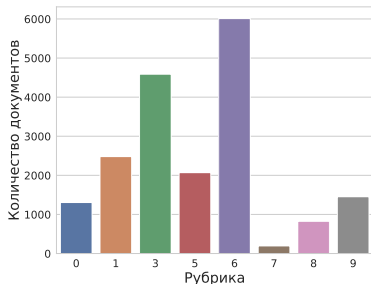


Использование мультимодальной тематической модели

В качестве модальностей используются **100** языков, а также научные рубрики ГРНТИ и УДК. В качестве предобработки используется **ВРЕ-токенизация** с 2 тыс. токенов на отдельный язык.



Средняя частота УДК: 0.4332

Средняя частота ГРНТИ: 0.5364