

Используемые метрики:

- likelihood ratio
- jaccard

Общий вид датасета.

СУД	АРЕСТОВАТЬ	ПОДПОЛКОВНИК	ЗАМНАЧАЛЬНИК	ОТДЕЛ
СУД	ВЕСТИ	ПЕРЕГОВОРЫ	ПОСТАВКА	БУДУЩИЙ
СУД	ВЗЯТЬ	РЕАЛИЗАЦИЯ	ЗАКОН	ОБЯЗАТЕЛЬНЫЙ
СУД	ВЫКУПИТЬ	ГОСУДАРСТВО	АКЦИЯ	РАНЕЕ
СУД	ГОВОРИТЬ	ПРЕСС-СЕКРЕТАРЬ	ЕВГЕНИЙ	ВРЕМЕНИТЬ

Золотой стандарт подбирался на основе примера решения к задаче **collocations_court**.

Топ 10 likelihood ratio.

```
('ПРЕКРАТИТЬ', 'ПРОИЗВОДСТВО'),  
( 'ОСВОБОДИТЬ', 'ЗАЛОГ'),  
( 'ЛИШЕНИЕ', 'СВОБОДА'),  
( 'ПРОЙТИ', 'ПРЕНИЕ'),  
( 'ПРИЙТИ', 'ВЫВОД'),  
( 'МЕРА', 'ПРЕСЕЧЕНИЕ'),  
( 'ПЕРЕПЕЧАТКА', 'СТАТЬЯ'),  
( 'ЗАБАСТОВКА', 'НЕЗАКОННЫЙ'),  
( 'ИЗБРАТЬ', 'МЕРА'),  
( 'СОДЕРЖАНИЕ', 'СТРАЖА')
```

Топ-10 jaccard.

```
('ОСВОБОДИТЬ', 'ЗАЛОГ'),  
( 'ЛИШЕНИЕ', 'СВОБОДА'),  
( 'МЕРА', 'ПРЕСЕЧЕНИЕ'),  
( 'ПРОЙТИ', 'ПРЕНИЕ'),  
( 'ВЫДАТЬ', 'САНКЦИЯ'),  
( 'ПРЕКРАТИТЬ', 'ПРОИЗВОДСТВО'),  
( 'ПРИЙТИ', 'ВЫВОД'),  
( 'ЗАБАСТОВКА', 'НЕЗАКОННЫЙ'),  
( 'ИЗБРАТЬ', 'МЕРА'),  
( 'ПЕРЕПЕЧАТКА', 'СТАТЬЯ')
```

Сравнение "золотого стандарта" с порядковыми номерами этих же коллокаций в списках(по метрикам)

```
gold [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]  
pmi [1089, 1400, 1979, 1927, 2000, 2095, 1440, 1753, 1966, 1654]  
jaccard [2, 4, 26, 12, 18, 34, 11, 14, 31, 20]
```

Корреляции.

PMI & jaccard

```
SpearmanrResult(correlation=0.8424242424242423,  
pvalue=0.0022200312259168411)
```

```
golden standart & jaccard SpearmanrResult(correlation=0.53939393939393931,  
pvalue=0.10759318778241481)
```

PMI & golden standart

```
SpearmanrResult(correlation=0.23636363636363633, pvalue=0.51088531751520017)
```

Выводы.

Мы видим сильную корреляцию между PMI и jaccard → ранг коллокаций довольно неплохо совпадает в обоих списках.

Между Золотым стандартом и PMI самая низкая корреляция.

Между jaccard и "золотым стандартом" корреляция присутствует небольшая корреляция тоже.

Всего 2 коллокации из "золотого стандарта" встречаются в топ-10 списка jaccard и ни одной не встречается в PMI.