

Коллокации отчет

Для каждого языка pm2, mutual_info, log_dice дают примерно одинаковые результаты. При этом mutual_info3 и t_score иногда совпадают, иногда отличаются, но никогда не совпадают с остальными тремя. Судя по английскому языку t_score показывает биграммы из стоп-слов, в целом показывая именно устойчивые выражения. Pmi, mutual_info и log_dice сильно зависят от редких слов, поэтому и показывают почти идентичные результаты. Mutual_info3 на мой взгляд сработал лучше всего, потому что вывел действительно интересные словосочетания (судя по английскому языку). Например «high school», «united states» или «inclusive democracy».