

Практическая работа 2. Анализ частотности терминов в книжной продукции

Помимо поиска в интернет публикациях, на сегодняшний день возможно осуществление анализа частотности употребления отдельных слов и фраз, названий, имен и фамилий известных людей и т.п. в книжных коллекциях на основных языках мира. Для этого можно воспользоваться сервисом Ngrams (<https://books.google.com/ngrams>), работающим на материалах проекта Google Books, на сегодняшний день аккумулировавшего более четверти из когда-либо изданных человечеством книг.

Google Books Ngram Viewer

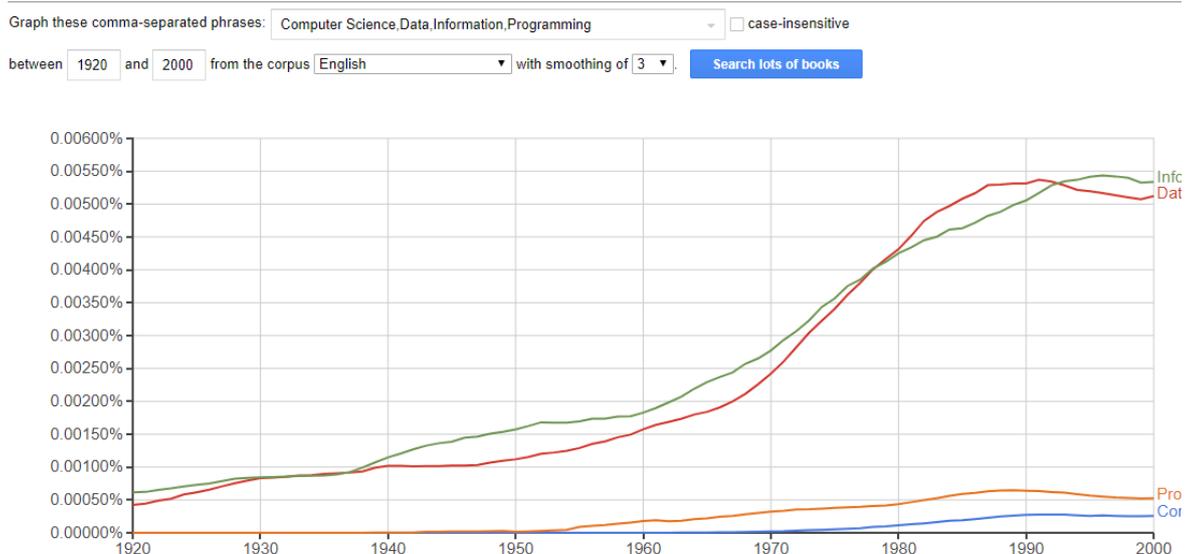


Рис. 1. Динамика популярности терминов Computer Science, Data, Information, Programming

Данные графика показывают, что за последние 100 лет трижды происходила смена приоритетов в использовании терминов Data и Information, два других термина - Computer Science и Programming, являются более узкоспециализированными, поэтому обладают меньшей относительной популярностью среди авторов книг и частоты их упоминания колеблется менее резко.

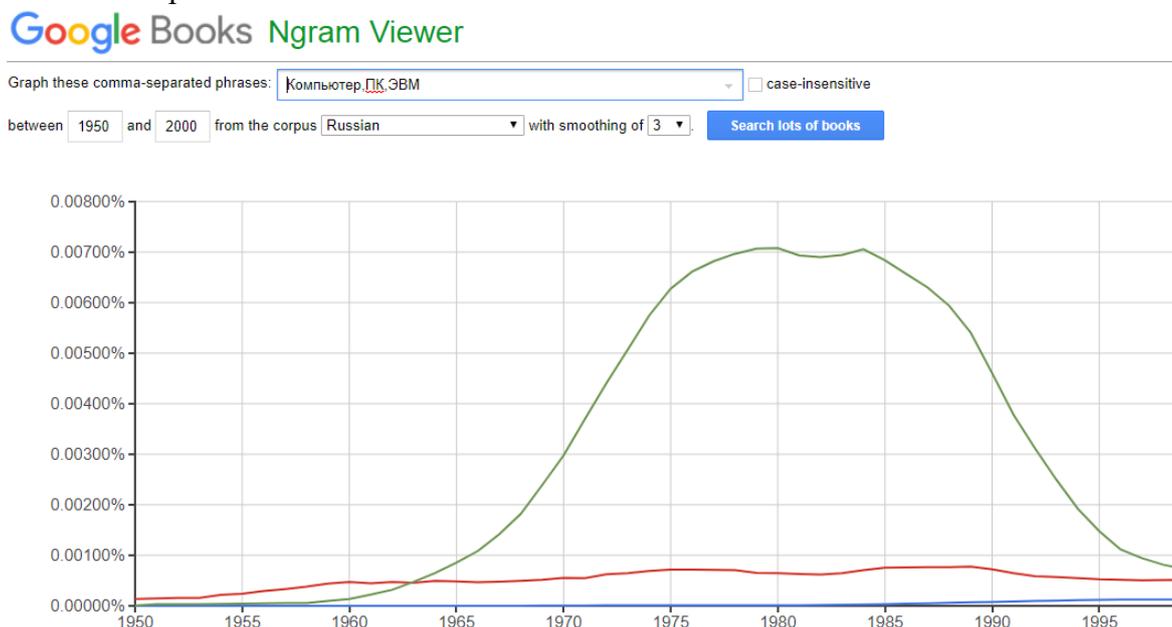


Рис.2. Динамика популярности названий вычислительной техники на русском языке.

Такого рода исследования можно выполнять не только на английском языке, но и, например, на русском. Для качественного выполнения данной работы следует поэкспериментировать с несколькими (минимум 5-7 группами) совокупностями синонимов или антонимов, обозначающих один и тот же объект, предмет, процесс и т.п. в различных областях деятельности. Также можно произвести поиск ФИО ученых, либо общественных деятелей, внесших значимый вклад в развитие какой-либо отрасли науки, культуры и т.д.

Решение:

Данные рисунка 1 показывают, что популярность искусственного интеллекта начала появляться после 1960 года, после 2000 года график резко пошел вверх и к 2019 году находится на самых высоких значениях за анализируемый период.

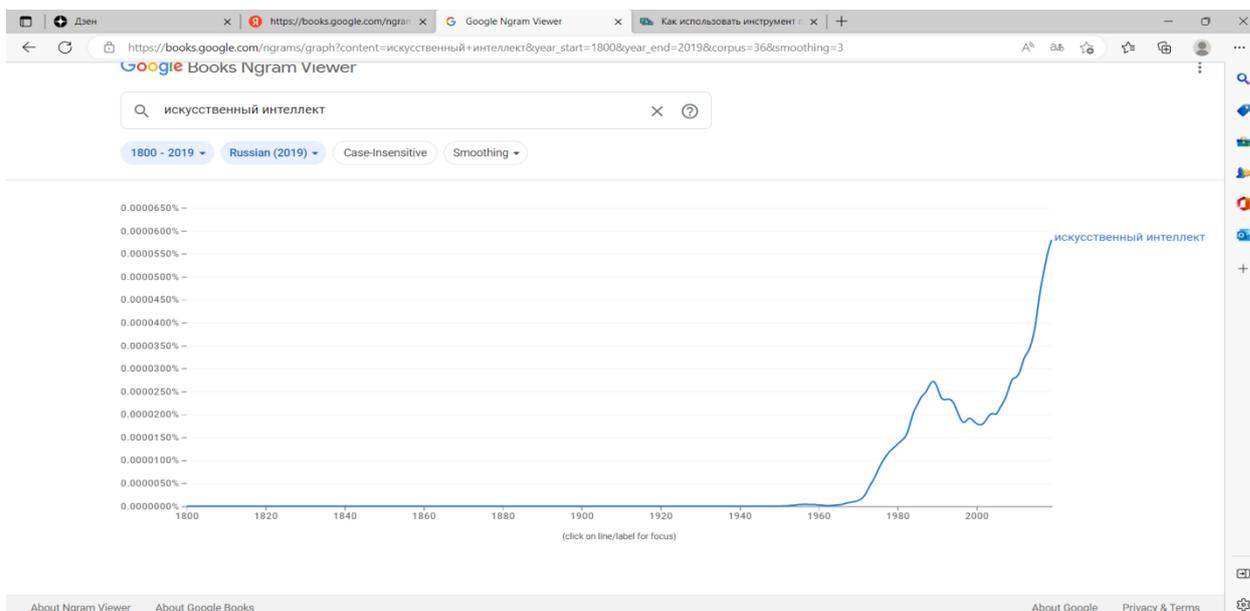


Рисунок 1- Динамика популярности термина «Искусственный интеллект» с 1800-2019гг.

По запросу –роботизация наблюдается резкое движение вверх к 1987 году- был достигнут пик, далее началось резкое снижение количества напоминаний по данному слову. Очевидно, что после 2000года на тему роботизации написано было не так много книг , рисунок 2.

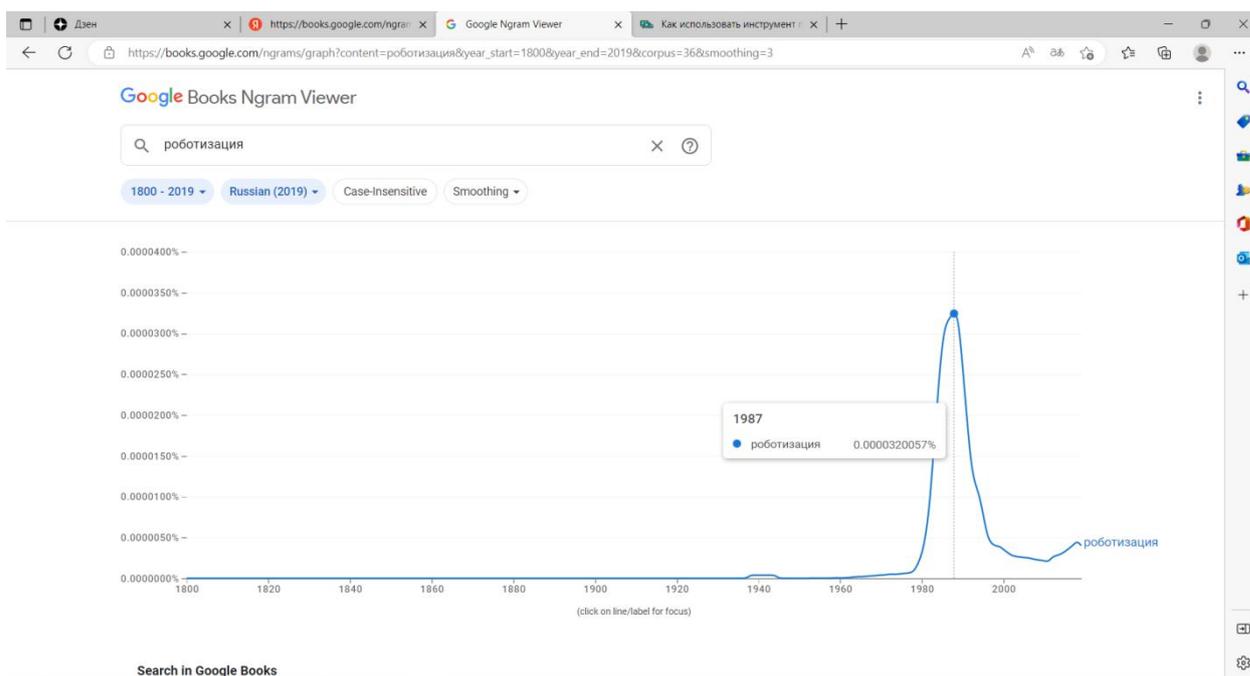


Рисунок 2- Динамика популярности термина «Роботизация» с 1800-2019гг.

Нельзя сказать, что о роботизации после 2000 года перестали говорить, но очевидно, что количество написанных книг на эту тему стало меньше.

По слову- блокчейн видно существенный рост популярности запроса после 2012года, фаза пика роста –самая высокая точка- в 2019 году, рисунок 3.

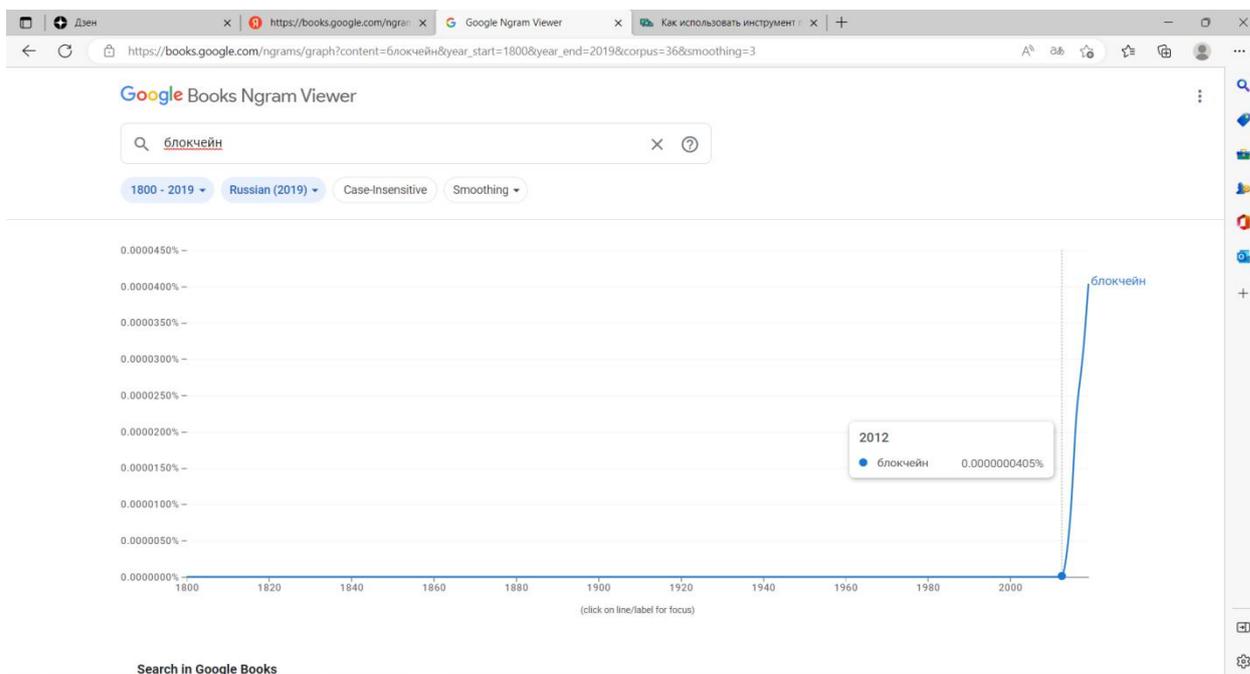


Рисунок 3- Динамика популярности термина «Блокчейн» с 1800-2019гг.

Пик популярности термина «большие данные» пришелся на 2016год, далее начался спад, рисунок 4.

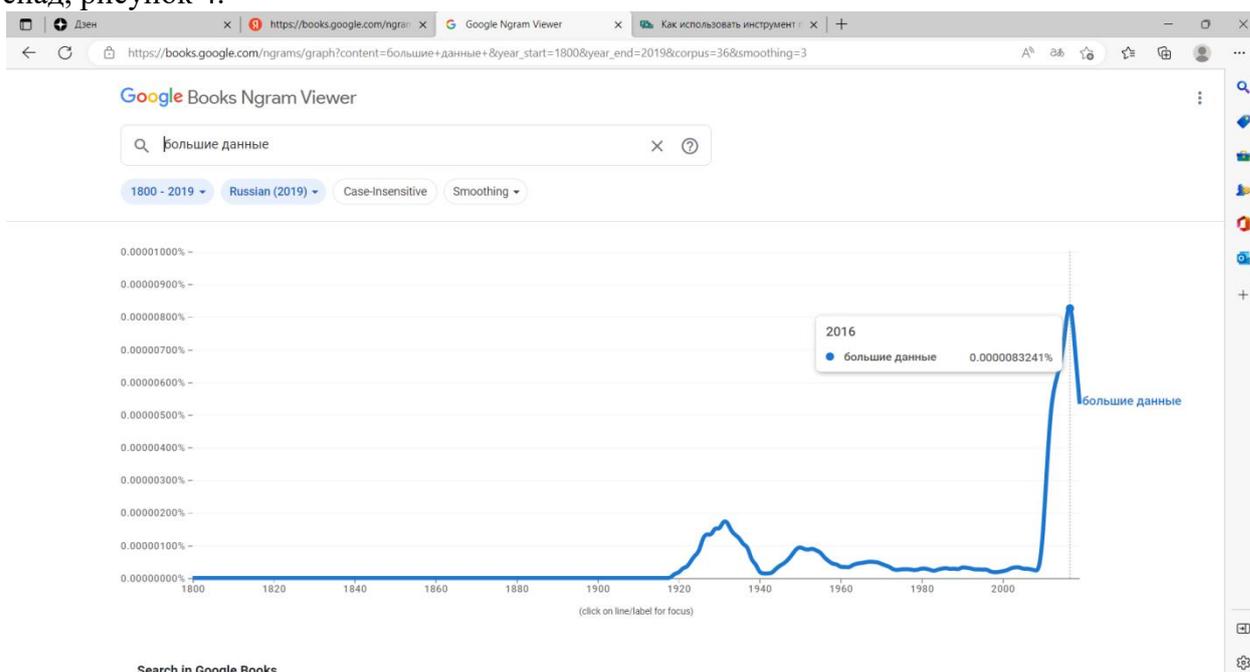


Рисунок 4- Динамика популярности термина «Большие данные» с 1800-2019гг.

Термин- цифровая экономика после 2000 до 2004 года получил популярность и количество упоминаний этого термина начало расти, однако, с 2004 по 2012 год видна динамика снижения, а после начался стремительный рост, который был достигнут в 2019 году, рисунок 5.

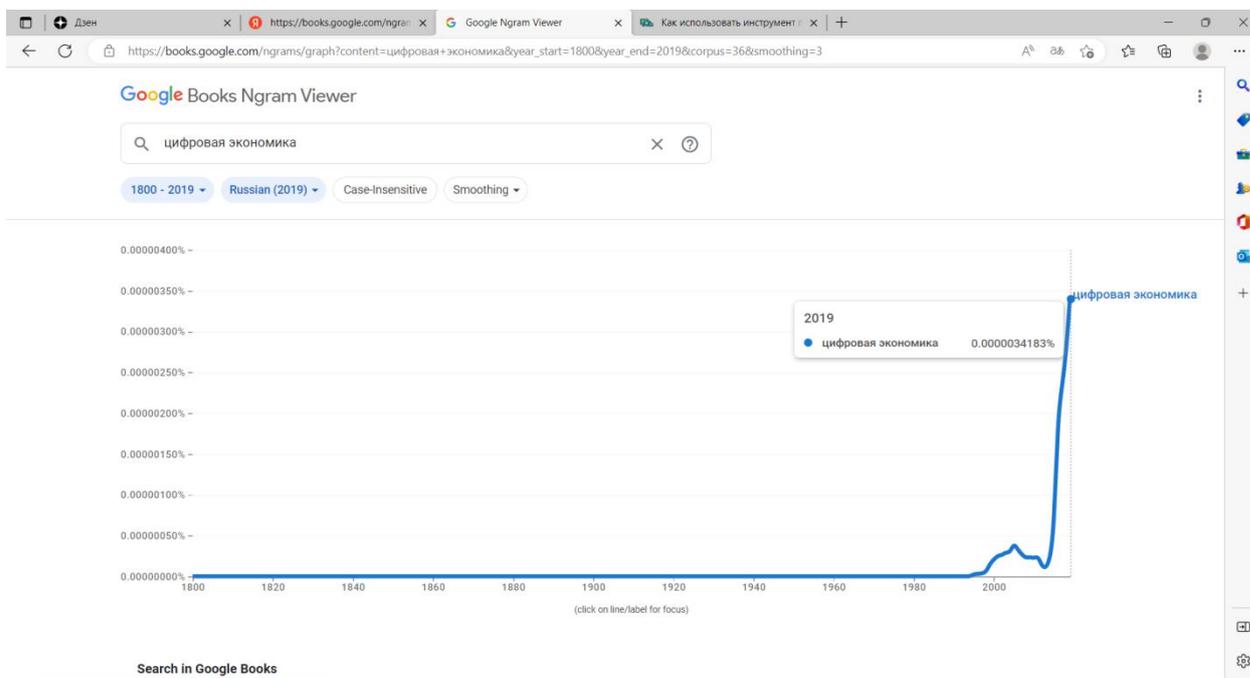


Рисунок 5- Динамика популярности термина «Цифровая экономика» с 1800-2019гг.

Таким образом, анализ частотности употребления терминов в книжной продукции на ресурсе (<https://books.google.com/ngrams>) показал, что термины связанные с понятием искусственного интеллекта и цифровой экономики приобрели большую популярность после 2012 года, к настоящему времени эти термины достигли максимальных значений. То есть количество упоминаний этих слов в книжных коллекциях увеличилось.