

Государственное бюджетное нетиповое образовательное учреждение
«Санкт-Петербургский городской Дворец творчества юных»

Всероссийский школьный конкурс по статистике «Тренд»

Номинация 1: «Учимся собирать статистические данные»

Тема исследования: «Технологизация русского языка: историческая перспектива и современность».

Работу выполнили учащиеся Юношеского клуба общественных наук (9 класс):

Донская Валентина

Фасхутдинов Ренат

Научный руководитель: Педагог дополнительного образования

Кулешова Юлия Александровна

Наставник: Педагог дополнительного образования

Бергштрейсер Кристина Евгеньевна

2025-2026 гг.

Аннотация

В условиях глобализации, внедрения в нашу жизнь новых технологий и, в том числе, популяризации интернета, русский язык находится в процессе постоянного изменения [2]. Было решено изучить, то, как слова, связанные с техническим прогрессом, используются в языке. Для этого были собраны актуальные текстовые массивы (базы данных), в которых мы определили топ «технологичных» слов. После чего изучен исторический контекст и современное использование тех слов, которые сегодня и сейчас в топе употребления.

После изучения задания номинации «Учимся собирать статистические данные», была сформулирована цель проведенного исследования - статистическое описание исторической перспективы и современного контекста использования слов, отражающих технологизацию русского языка. Для достижения этой цели был разработан план последовательных действий (задачи): изучение литературы на тему технологизацию русского языка для накопления теоретической базы по выбранной теме исследования; сбор базы данных новостных статей для статистически обоснованного выбора слов, которые станут ключевыми для дальнейшего исследования; изучение статистических методов анализа текста; изучение архитектуры и функциональных возможностей веб-ресурса «Национальный корпус русского языка»; создание базы текстов новостных статей, поисковых запросов и академических текстов; изучение исторической перспективы и контекстов современного использования выбранных ключевых слов в русском языке. Более подробное описание логики исследования представлено в основной части работы.

Объектом исследования выбраны три ключевых слова, семантически связанные с технологиями: «фильм», «сайт», «производство». Предметом исследования является изучение влияния технологизацию на русский язык: историческая перспектива и современные коллокации ключевых слов.

Обзор литературы

Язык - зеркало общества, он отражает в себе нормы и ценности, политическую и экономическую, социальную ситуацию. На язык влияет массовая культура, научные открытия, сфера образования, государственная языковая политика, лингвистические моды, педагогические подходы, технологический прогресс и многое другое [1].

Благодаря технологическому прогрессу в лексикон русского языка добавляются термины, связанные с программированием, технологиями, например, связанные с появлением новых устройств, программного обеспечения и интернет-сервисов.

Не менее важным фактором является популяризация социальных сетей, так как в них массово используют сокращенные формы слов, некоторые пренебрежения и упрощения

норм русского языка, также символы и эмодзи. Эта тенденция меняет структуру предложений, влияет на грамматические нормы [3].

Русский язык — это система, которой свойственно постоянное, динамичное изменение. Основной причиной изменений является глобализация, которая помогает разным культурам взаимодействовать, что приводит к массовому заимствованию слов из других языков, особенно из английского. Сегодня этот процесс поддерживается благодаря популяризации интернета и развитию технологий [2]. Описанная тенденция сопровождается некоторыми затруднениями - заимствованные слова не всегда адаптируются к лексическим и грамматическим особенностям языка [3].

По мнению Оразкулыева Б. еще один фактор, способствующий изменению языка это программы автоматического перевода текста. Часто они не могут учитывать особенности контекста и искажают структуру предложений. Второй проблемой, связанной с такими сервисами, является стирание разницы между стандартным и разговорным стилями речи, что также ведет к искажению смысла. В сумме эти проблемы оказывают непредсказуемое влияние на нормы русского языка [3].

Ход исследования: материалы и методы

В рамках выбранной номинации была поставлена задача изучить как технологический прогресс отражается в русском языке, в том числе проверив гипотезу, что с появлением новшеств соответствующие термины включаются в повседневный язык, причём, чем шире применяется то или иное техническое достижение, тем чаще соответствующий термин упоминается в книгах, газетах, социальных сетях и т.д. Для более подробного ознакомления с визуализацией исследования был создан чат-бот @TrendUKON_bot (приложение А).

Для того, чтобы выбрать ключевые слова для анализа, был создан массив данных, в который вошли 800 новостных статей с сайта информационного агентства ТАСС по четырем направлениям (спорт, культура, экономика, наука) по 200 статей в каждом. Направления выбраны таким образом, чтобы затрагивать разные сферы общественной жизни общества.

СМИ — это каналы распространения информации широкой аудитории. Соответственно, новостные статьи пишутся языком доступным, понятным и распространенным среди населения РФ. ТАСС является ведущим государственным информационным агентством России, ежедневно выпускающим тысячи сообщений.

Для того, чтобы проанализировать собранную базу и определить список наиболее часто используемых слов, команда воспользовалась языком программирования «R» (бесплатной средой с открытым исходным кодом, созданной специально для

статистического анализа данных).

Для создания частотного списка была проведена предварительная обработка текстовой базы данных, а именно токенизация (разбиение новостных статей на отдельные слова, которые станут единицей подсчета), первичная обработка и чистка текста (удаление цифр, слов, написанных латиницей, слов входящих в список стоп-слов и слова короче двух символов), лемматизация (приведение слов к базовой форме). Далее был составлен частотный список слов.

Был получен частотный список из 19 448 слов. В ходе поиска слов, связанных с технологическим прогрессом, были отобраны следующие ключевые слова для анализа: сайт (292 упоминания), фильм (260 упоминаний), производство (118 упоминаний). Каждое из этих слов обозначает область, которая была создана или трансформирована благодаря техническому прогрессу, и служит примером его влияния на современный язык и мир в целом.

Следующий этап исследования – изучение исторической перспективы применения ключевых слов в русском языке. Для его реализации были изучены данные «Национального корпуса русского языка». Для анализа исторической динамики использования ключевых слов в русском языке были взяты основной корпус, газетный и корпус социальных сетей. Проанализированы графики, демонстрирующие частоту встречаемости слова на миллион словоформ (количество употреблений слова за дату (например, за год), поделенное на объем корпуса за эту дату и умноженное на 1 миллион.). Также проанализирован прирост (в %) частоты встречаемости слова на миллион словоформ. Также изучены коллокации ключевых слов в основном корпусе.

«Национальный корпус русского языка» представляет собой коллекции текстов на русском языке общим объемом более 2 млрд слов, оснащенную лингвистической разметкой и инструментами поиска.

Основной корпус НКРЯ - корпус русских письменных прозаических текстов, созданных после 1700 года. Он отражает русский язык Нового времени, отсчёт истории которого начинается именно с Петровской эпохи. Корпус социальных сетей НКРЯ - корпус, который отражает живые языковые изменения за пределами литературного языка. Такие языковые явления почти не фиксируются в основном и газетном корпусе. Газетный (центральный) корпус НКРЯ - корпус открыт в 2010 г. и охватывает статьи из средств массовой информации начиная с 1983 г. до 2021 г. Значительные объёмы текстов СМИ, доступные в электронном виде и представляющие большой интерес для мониторинга языковых изменений «в режиме реального времени».

Ресурс «Яндекс Вордстат» позволил оценить объемы поисковых запросов, по

ключевым словам, что демонстрирует распространенность применения слова в современном интернет-пространстве.

Далее была собрана вторая база текстовых данных на основе ключевых слов, с целью изучения их использования в современном русском языке. Для изучения современного контекста использования ключевых слов для сравнения были выбраны три сферы (типа) текстовых данных: текст топ-20 интернет ресурсов по результатам поисковых запросов в системе Яндекс (первыми поисковая система выдает самые популярные результаты запросов), топ-30 новостных статей по поиску через ключевое слово, топ-20 академических текстов с сайта научной электронной библиотеки Киберленинка (статьи отбирались по следующим критериям: написана не ранее 2021 года, фильтр по научным базам «ВАК»).

Данные были сохранены в виде отдельных файлов в формате xls с соответствующими названиями, например, «поиск фильм», то есть это таблица данных из 20 строк, полученная в результате поискового запроса по ключевому слову «фильм» со следующими столбцами: ссылка на интернет-ресурс, текст со страницы ресурса. Мы получили 9 файлов для сравнительного анализа.

Для анализа второй базы текстовых данных также был применен инструмент R. Ниже представлена логика анализа.

В начале были скачаны и установлены пакеты для анализа. Далее загружены данные в формате таблицы. При помощи функции (концепции) `corpus` мы преобразовали первоначальную таблицу в формат, подходящий для работы в R и первоначальной обработки текста, а именно: понижение регистра букв, удаление пунктуации и цифр, слов, содержащих латинские буквы, стоп-слов русского языка, а также повторных пробелов. В результате получаем корпус, очищенный от лишнего шума. Конечно, идеально вычистить текст механически достаточно сложно, но базовая обработка текста для дальнейшего анализа необходима.

Были определены и созданы функции, то есть механизмы для дальнейшего анализа, а именно: функция для создания частотного списка, функция расчёта энтропии (меру неопределенности, степень разнообразия/вариативности текстового набора), функция для создания облаков слов, функция анализа биграмм (поиск повторяющихся словосочетаний и создание из них частотного списка), функция визуализации семантической сети (сеть биграмм с тремя wybranными формами ключевого слова, дополненная частотой встречаемости каждого словосочетания).

Для анализа использовалась нормализованная энтропия, то есть энтропия с учетом уникальных слов в тексте.

Результаты: сайт

Сайт — это набор связанных веб-страниц, размещённых в интернете под общим доменным именем, предназначенных для предоставления информации, взаимодействия с пользователями или выполнения определённых функций. Слово «сайт» образовано от англ. Website, где web – это паутина, сеть, а site – место. Термин появился в 1990-х годах с появлением первых веб-страниц. В русском языке слово «сайт» появилось в конце 1990-х с распространением интернета в России [IV].

Согласно данным «Яндекс Вордстат» в период с 06.10.2025 по 06.11.2025 слово «сайт» искали 107 826 601 раз. В период с 16.11.2025 по 16.12.2025 слово «сайт» искали 109 141 318 раз. Больше всего слово «сайт» искали в центральной части России (больше 37 тысяч раз), Москве и Московской области (больше 24 тысяч раз), Поволжье (больше 18 тысяч раз). Далее идет Северо-запад (больше 13 тысяч раз) и замыкает пятерку лидеров Сибирь (больше 10 тысяч раз)

В основном корпусе НКРЯ график демонстрирует распределение результатов поиска ключевого слова («сайт») по датам (частота встречаемости ключевого слова на миллион словоформ) в период с 1853 года по 2022 год. Во второй половине 19-го века слово «сайт» имело иное значение, относительно современного понятия. В тот исторический период, заимствованное из французского языка слово было связано с обозначением местоположения, участка земли или места жительства, например, «удачный сайт для постройки дома». Так же в корпусе НКРЯ в этот период упоминается город «Сайто», и большая часть цитат связана с японской культурой и религией. График показывает динамику роста частоты использования ключевого слова «сайт» начиная с 1997 года уже в более привычном нам значении. Мы можем связать это с распространением и популяризацией интернета, то есть появлением первых сайтов. Затем начинается спад в употреблении ключевого слова с 2016, можно предположить, что это связано с увеличением общего числа статистической информации и пополнением базы НКРЯ.

Облако слов, построенное на основе частотного списка колокоатов по основному корпусу НКРЯ демонстрирует самые популярные семантические словосочетания ключевого слова со словами «официальный», «интернет», «зайти», «создать», «найти», «сделать» и т.д. То есть слово сайт в русском языке чаще всего употребляется с соответствующими глаголами. Частотный список по базе новостных сообщений демонстрирует наибольшую частоту встречаемости у источника новостей, даты и страны - «Тасс», «года», «Россия», «ноября».

Для ключевого слова «сайт», чаще всего употребляют слово «сайт» (301 повторения) и другие формы этого слова «сайта» (274 повторения), «сайтов» (96 повторений), значит

пользователи чаще всего использовали ключевые слова в процессе поиска веб-страниц. Также часто повторяются слова «это» (196 повторений) и «например» (93 повторения).

В ходе анализа биграмм были визуализированы семантические связи с ключевыми словами в трех формах «сайт», «сайта», «сайты». Чаще всего встречаются словосочетания «Создать сайт», то есть большая часть результатов поиска браузер по ключевому слову, связано с созданием сайтов. Также популярно словосочетание «Веб сайт», «доменное имя», «интернет магазин».

Если анализировать биграммы с ключевым словом, то самыми популярными являются «создать сайт» (52 повторения), «веб сайт» (34 повторения) и «сайт это» (23 повторения). То есть пользователи чаще всего используют ключевое слово, когда ищут инструкции по созданию сайтов, когда хотят растолковать значение слова «сайт», а также, как устойчивую связку «веб сайт».

Новостной блок

Входящие в топ частотного списка слова демонстрируют, что к окончанию 2025 года новости публикуют статьи про принадлежность к организации, учреждению, проекту или же указывают на владельца или оператора сайта (сайте ТАСС, сайте НАТО, сайте фонда, сайте гидрометцентра, сайте Кремля, сайте правительства, сайте законодательного, сайте национального), о принадлежности к событию/акции или указывают на сайт, посвященный конкретному мероприятию (сайте выставки, сайте Геодиктанта).

Для анализа семантических сетей было взято две формы ключевого слова: сайт и сайте. Обе формы являются изолированными и никак друг с другом не связаны. Однако мы можем увидеть самую популярную пару слов: сайте ТАСС, но это объясняется тем, что мы собрали БАЗУ 1, используя сайт агентства ТАСС.

Поисковый блок

Входящие в топ частотного списка слова показывают, что к окончанию 2025 года чаще всего в поисковых сервисах ищут следующую информацию со словом «сайт»: о действии и процессе, указание на создание сайта (создать сайт, создания сайта), об определении и классификации типа сайта (веб сайт, корпоративный сайт, сайт визитка), о принадлежности и владении сайтом (ваш сайт, свой сайт), о действии с сайтом, его состоянии и проблемах (блокировка сайта, сайта шаг).

Для анализа семантических сетей было взято три формы ключевого слова: сайты, сайт и сайта. Первая форма является изолированной, а вот вторая и третья связаны словом «веб». (18 ноября 2025 года произошел глобальный сбой интернета, который затронул и российский сегмент. Причиной стал сбой в работе компании Cloudflare, через сеть

которой проходит около 20% всего веб-трафика. Из-за сбоя пользователи по всему миру, в том числе в России, столкнулись с проблемами доступа к многим популярным сайтам и сервисам.)

Академический блок

Входящие в топ частотного списка слова демонстрируют, что к окончанию 2025 года чаще всего в сервисе Киберленинка со словом «сайт» встречаются статьи: о терминах, связанных с версткой, технической организацией и версиями сайта (грид сайта, сайт специфик, версии сайта, страницы сайта), об уточнении статуса, формата или принадлежности сайта (корпоративный сайт, официальный сайт, веб-сайта).

Для анализа семантических сетей было взято три формы ключевого слова: сайта, сайты и сайт. В данном случае все формы слова связаны между собой. Связаны первая и вторая формы словом «веб». Первая и третья формы связаны словами «работы», «оценки». Также мы можем увидеть, что все три формы связаны словом «библиотеки».

Результаты: производство

В этимологических словарях русского языка слово «производство» отдельно не рассматривается, но можно предположить следующий смысл его двух частей. Первая – это префикс про- вперед, дальше, действующий в интересах, замещающий. Вторая часть - изводство – водить – вести: у слова праславянские истоки, несущие значение – двигать по поверхности, держать, действовать и в этом смысле – прилагать усилия [I]. Производство, процесс создания благ с использованием факторов производства для непосредственного (предметы конечного потребления) и опосредованного средства производства, удовлетворения человеческих потребностей [III].

Согласно данным «ЯндексВордстат» в период с 06.10.2025 по 06.11.2025 слово «производство» искали 9 806 393 раза. В период с 16.11.2025 по 16.12.2025 слово «производство» искали 10 284 510 раз. Больше всего слово «производство» искали в центральной части России (больше 3 тысяч раз), Поволжье и Московской области (больше 1.8 тысяч раз), Москве и Сибири (больше 1 тысячи раз).

В основном корпусе НКРЯ график демонстрирует распределение результатов поиска ключевого слова («производство») по датам (частота встречаемости ключевого слова на миллионов словоформ) в период с 1682 года по 2022 год. График показывает динамику роста частоты использования слова «производство» до 2004 года. На графике видны максимальные значения, превышающие 100 упоминаний в 2004 году и 2014 году. Это может быть связано с тем, что 2004 году в России произошел рост промышленного производства на 6,1% по сравнению с 2003 годом. Далее демонстрируется спад, который может быть связан с увеличением объема базы НКРЯ, поэтому на каждый миллион слово

«производство» встречается реже.

В газетном центральном корпусе график демонстрирует распределение результатов поиска ключевого слова («производство») по датам. На графике видны максимальные значения, превышающие 1400 упоминаний в 1980 году, 1983 и 1986 году. Это может быть связано с тем, что прирост производительности труда снизился почти в два раза и приблизился к нулевой отметке, что освещали в газетах. Однако, стоит отметить, что в 1983 году произошло повышение темпов роста в промышленности и других отраслях народного хозяйства. Страна произвела промышленной продукции на 4% больше, чем год назад, производительность труда возросла на 3,5%. А в 1986 объем производимой продукции промышленности вырос по сравнению с предыдущим годом на 39 миллиардов рублей и достигла 840 миллиардов рублей.

Топ колоколатов по ключевому слову «производство» состоит из слов, связанных с процессом производства, указывает на то, что именно производится или в какой отрасли (вино производство, документации производство, гальваническом производстве, производстве печатных), способ организации производства (товарное производство, масштабное производство), принцип, философию или метод ведения производства (бережливое производство, быстро реагирующее производство, упрощенное производство), правовыми процедурами, а не с физическим созданием продукта (исполнительное производство, исполнительном производстве, исковое производство)

Новостной блок

Входящие в топ частотного списка базы данных новостных сообщений по ключевому слову «производство» слова демонстрируют, что к окончанию 2025 года новости чаще всего публикуют информацию о местоположении производств (днецкого производства, американского производства, немецкого производства), а также новости, касающиеся судебного производства (уголовного производства).

Для анализа семантических сетей было взято три формы ключевого слова: «произшеств», «производства» и «производство». Первая форма является изолированной, а вот вторая и третья связаны

Семантическая сеть: Ключевые слова - industrial

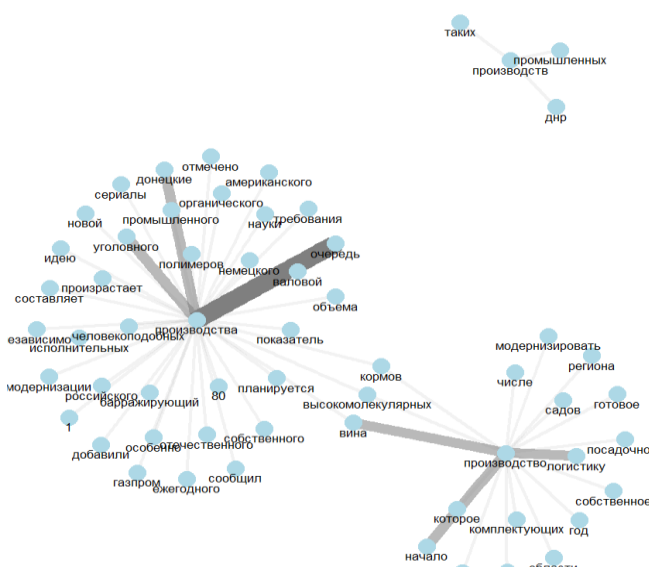


рис. 1. Визуализация семантических сетей по ключевому слову "производство"

словом «кормов», «вина», «высокомолекулярных». Пояснение: (С 1 ноября 2025 года вступил в силу закон, о новом этапе производства, а именно маркировки кормов для животных «Честный знак».), (По данным Международной организации виноградарства и виноделия, опубликованным в ноябре 2025 года, по итогам 2025 года Россия произведет 5,1 миллиона гектолитров вина и останется шестой в Европе и одиннадцатой в мире по его выпуску.), 21 ноября 2025 года на флагманском заводе группы компаний «Полипласт» — «Полипласт Новомосковск» — запустили вторую очередь единственного в России комплекса по производству высокомолекулярных полимеров.

Поисковый блок

Входящие в топ частотного списка слова демонстрируют, что к окончанию 2025 года чаще всего в поисковых сервисах ищут следующую информацию со словом «производство»: о типе, свойстве или сфере производства (рыночное

	word frequency	type
производства	производства	295 industrial
мм	мм	165 industrial
производство	производство	150 industrial
это	это	122 industrial
процесс	процесс	75 industrial
изделия	изделия	71 industrial
процесса	процесса	63 industrial
высота	высота	60 industrial
длина	длина	60 industrial
ширина	ширина	60 industrial
продукции	продукции	59 industrial

рис. 2. Частотный список по ключевому слову "производство" - поисковый блок

производство, серийное / массовое / единичное производство, промышленное / домашнее производство, крупносерийное / мелкосерийное производство), о продукте или результате деятельности (производство товаров), о рамках определенного типа производства (единичном производстве).

Для анализа семантических сетей было взято три формы ключевого слова: «производств», «производства» и «производство». Вторая и третья форма связаны словами «любое», «крупносерийное». (слово «любое» не может быть описано в связке со словами «производство», «производства» так как не имеет логического обоснования, однако мы можем описать связку со словом «крупносерийное», по данным отчета S&P Global, в ноябре 2025 года произошло самое быстрое с апреля 2022 года падение крупносерийного производства в России, что объясняется сокращением заказов и задержками в цепочке поставок)

Академический блок

Входящие в топ частотного списка слова демонстрируют, что к окончанию 2025 года чаще всего в сервисе Киберленинка со словом “производство” встречаются статьи: о методологии, подходе или принципе организации производства (бережливое производство, быстрореагирующее производство, упрощенное производство, масштабное производство), о производстве в значении процедуры, юридического процесса (исполнительное производство, исковое производство, упрощенное

производство), о назначении продукта с точки зрения экономики (товарное производство), о конкретной отрасли, технологии или продукта (гальваническом производстве, печатном производстве).

Для анализа семантических сетей было взято три формы ключевого слова: производств, производства и производство. Первая форма является изолированной, а вот вторая и третья связаны словом “электроэнергии”, “энергии”, “сыров”. (Производство электроэнергии в России сократилось в январе-сентябре 2025 года на 1,3 %), (По данным отчета Аналитического центра Milknews, за январь — сентябрь 2025 года производство сыров и сырных продуктов в России снизилось на 1,3% по сравнению с аналогичным периодом 2024 года — до 782 тыс. т. При этом производство сыров увеличилось на 0,8% — до 644,6 тыс. т, а выпуск сырных продуктов снизился на 10,2% — до 137,4 тыс. т.)

Результаты: фильм

Древнеанглийское – film (кожа). Слово появилось в русском языке после революции 1917 г. и некоторое время использовалось в форме «фильма». В английском языке, где возникло данное слово и понятие, фильмом первоначально называли тонкую кожицу, пленку. После создания кинематографа слово стало обозначать одно из самых замечательных изобретений XX в. – киноленту [11].

Согласно данным «Яндекс Вордстат» в период с 06.10.2025 по 06.11.2025 слово «фильм» искали 225 633 022 раза. В период с 16.11.2025 по 16.12.2025 слово «фильм» искали 235 446 976 раз. Больше всего слово «фильм» искали в центральной части России (больше 70 тысяч раз), Москве (больше 40 тысяч раз) и Московской области и Поволжье (больше 30 тысяч раз).

График результатов поиска по датам по основному корпусу НКРЯ (частота на миллион словоформ) демонстрирует распределение результатов поиска ключевого слова («фильм») в период с 1682 года по 2022 год. Видна динамика роста частоты использования слова «фильм» до 2012 года. Далее демонстрируется спад, который может быть связан с увеличением объема и обогащением базы НКРЯ, поэтому на каждый миллион слово «фильм» встречается реже.

При рассмотрении аналогичного графика без эффекта сглаживания мы видим максимальные значения, превышающие 500 упоминаний в 1972 году, 1976 году и 2010 году. Так как количество текстов и их жанровое разнообразие для каждого года в НКРЯ неравномерно подобные пики могут указывать на включение в корпус большого количества специализированных текстов (например, научные статьи или монографии по киноведению, журнальные статьи и кинокритика) на заданную тему.

Также в 1972 году вышло несколько популярных советских кинолент («А зори здесь

тихие», «Большая перемена»), которые могли повлиять на число упоминаний ключевого слова. 1 января 1976 года был впервые показан фильм Эльдара Рязанова «Ирония судьбы, или с лёгким паром!», ставший всенародным хитом и новогодней традицией.

Пик 2010 года может быть связан также с ростом кинопроката и активной деятельностью «Федерального фонда социальной и экономической поддержки отечественной кинематографии» - государственной организации, созданной для поддержки и развития отечественного кино.

Топ колоколатов по ключевому слову «фильм» состоит из слов, связанных с процессом просмотра кинофильма, жанрами (самые популярные - документальный и художественный), а также страной производства (американский и французский фильм).

Новостной блок

Входящие в топ частотного списка слова демонстрируют, что к окончанию 2025 года новости публикуют информацию о лучших фильмах года, в том числе отечественных. Топ частотного списка биграмм демонстрирует упоминания известных людей, умерших в начале ноября 2025: Павел Никонов и Владимир Симонов. В топе биграмм с ключевыми словами достаточно мало повторений, однако упоминаются документальный и спортивный жанры фильмов.



рис. 3. Облако слов по ключевому слову "фильм" - новостной блок

Для анализа семантических сетей было взято три формы ключевого слова: фильме, фильма и фильм. Первая форма является изолированной, а вот вторая и третья связаны словом «Алиса» (в ноябре 2025 года вышел российский фильм «Алиса в стране чудес»).

Форма слова «фильм» в употреблении чаще всего связана с именами («Ольги», «Татьяны» и т.д.) и деятельностью людей, работавших над фильмом («откроет», «создал», «отметил», «сняв» и т.д.). Чаще всего эта форма употребляется со словом «лучший». Форма слова «фильма» больше связана с названиями или тематикой киноленты («Нюрнберг», «Онегин» и т.д.) и базовыми слова описывающими «жизненный» процесс фильма («съемки», «показ», «сборы» и т.д.). Чаще всего эта форма употребляется со словом «документального».

Поисковый блок

Поисковик выдает самыми первыми страницы, которые чаще всего запрашивают люди. По ключевому слову «фильм» чаще всего люди ищут платформы для бесплатного

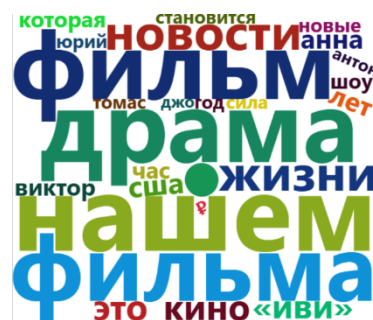


рис. 4. Облако слов по блоку "фильм" - поисковый блок

онлайн просмотра и выбирают фильмы по оценкам и отзывам Кинопоиска. Топ биграмм подтверждает предыдущий вывод, добавляя к нему поиск топовых фильмов, то есть рейтинговых списков кинолент. Топ биграмм с ключевыми словами демонстрирует в основном жанровую тематику («драма», «триллер», «мультфильм» и т.д.).

Для анализа семантических сетей было взято три формы ключевого слова: фильмов, фильма и фильм. Вторая форма является изолированной. Две другие связаны жанром «ужасы». Больше всего вариантов словосочетаний с формой «фильм», в том числе много запросов по годам выхода кинолент, то есть использования числительных с этой формой слова. Эта форма чаще всего употребляется со словами «документальный» и «драма». Форма «фильмов» связана с поисковыми запросами: топ, лучшие, самых.

Академический блок

Топ частотного списка этого блока в основном наполнен несодержательными словами, однако встречаются слова «режиссер», «вестник» и «истории». Топ биграмм позволяет уточнить, что один из популярных источников информации в данном блоке -

это Научный информационно-аналитический журнал «Вестник ВГИК». Также описывается «отечественное кино» и упоминается «политическая память». Топ биграмм с ключевыми формами слова как и в предыдущих блоках демонстрирует палитру жанров и типов: биографический, зарубежный, отечественный, документальный.

Топ-15 биграмм с ключевыми словами:

	word1	word2	n
1	биографический	фильм	9
2	любимый	фильм	9
3	структура	фильма	8
4	versus	фильм	7
5	биографического	фильма	7
6	зарубежные	фильмы	7
7	мартьянова	фильм	7
8	отечественные	фильмы	7

рис. 5 Частотный список биграмм по ключевому слову "фильм" - академический блок

Для анализа семантических сетей было взято три формы ключевого слова: фильмы, фильма и фильм. Первые две формы связаны словом «производство», вторая и третья форма связаны словом «смотрит». Слово «фильмы» из всей семантических связей чаще всего употребляется в нашей базе со словом «отечественный». Форма «фильм» имеет несколько наиболее часто встречаемых сочетаний «любимый», «сходства», «документальный», «предпочтения».

Сравнение энтропии

База	Общее количество слов	Количество уникальных слов
академ фильм	61885	22572
академ сайт	49556	14718
академ производство	36183	13026
поиск фильм	34905	11244
поиск сайт	27563	7667
поиск производство	26516	8826
новости сайт	9310	4353
новости фильм	8811	4196
новости производства	8581	4061

рис. 6. Таблица количества слов в текстовых массивах

Вторая база текстовых данных характеризуется следующими показателями объема текста по количеству общих и уникальных слов. Самыми большими объемами текста обладают академические базы (тексты статей из открытой библиотеки «Киберленинка», в том числе по количеству уникальных слов.

Показатель энтропии необходимо анализировать в сравнении. В данном случае анализируется нормализованная энтропия, то есть вариативность/разнообразие текста с учетом именно уникальных слов, так как объемы текстовых массивов отличаются друг от друга и сравнивать энтропию без поправки на уникальные слова некорректно. График демонстрирует наибольшие показатели в «новостном» блоке, а наименьшие в «поисковом», что говорит о меньшей в первом случае и наибольшей во втором сфокусированности тематики текста, узости описываемого объекта и сопряженных с ним тем. Если проводить сравнение, по ключевым словам, то слово «производство» дает наиболее разнообразные тексты и соответственно семантические связи в применении в русском языке.

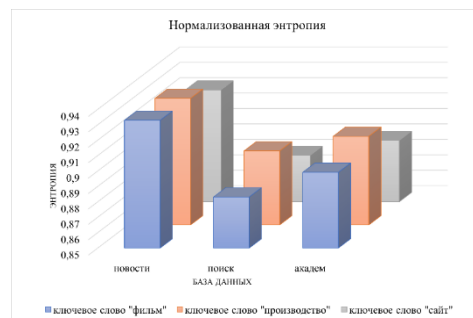


рис. 7. Сравнение нормализованной энтропии текстовых массивов

Заключение

В этимологических словарях русского языка слово «производство» нет, но можно предположить смысл его двух частей. Первая – это префикс про- вперед, дальше, действующий в интересах, замещающий. Вторая часть - изводство – водить – вести: у слова праславянские истоки, несущие значение – двигать по поверхности, держать, действовать и в этом смысле – прилагать усилия.

Древнеанглийское – film. В английском языке, фильмом первоначально называли тонкую кожицу, пленку. После создания кинематографа слово стало обозначать – кинолентку. Слово появилось в русском языке после 1917 года.

Слово «сайт» образованно от англ. Website, где web – это паутина, сеть, а site – место. Термин появился в 1990-х годов с появлением первых веб-страниц. В русском языке слово «сайт» появилось в конце 1990-х с распространением интернета в России. В 19 веке слово имело иное значение (местоположение), а также использовалось в контексте Японии и города Сайто.

В современном языке слово «сайт» функционирует в трёх основных контекстах: бытовом (сделать сайт), информационно-справочном (найти на сайте) и профессионально-техническом (оптимизировать сайт), в свою очередь слово «фильм» функционирует в трёх ключевых контекстах: поисковом (что посмотреть?), новостно-событийном (что происходит?) и аналитико-оценочном (как устроено? что нравится?), первый контекст — самый часто встречающийся, второй — связан с актуальными событиями, третий — наиболее специализированный.

Слово «производство» является многогранным термином, который применяется в

трёх основных контекстах: теоретическом (что это такое и каким бывает?), фактологическо-событийном (что и где производится сейчас? какой объём? какие процедуры идут?), прикладном профессиональном (как эффективно организовать или усовершенствовать конкретный производственный процесс?). Наиболее частым и структурированным является первый, образовательный контекст.

На основе анализа трёх ключевых слов — «сайт», «фильм» и «производство» — можно сделать общий вывод о закономерностях употребления подобных многозначных терминов. Наиболее частотным является базовый контекст, так как он обслуживает самые массовые потребности (поиск, первичное ознакомление, простое действие). Это основа цифровой и медийной коммуникации.

Источники

1. Курбонов Д. К., Каримов У. Р. Углы развитие технологии и социальных сетей как фактор изменения русского языка // orienss. 2024. №7. URL: <https://cyberleninka.ru/article/n/razvitie-tehnologii-i-sotsialnyh-setey-kak-faktor-izmeneniya-russkogo-yazyka> (дата обращения: 10.11.2025).
2. Мамедова А. А. Современный русский язык // Science and Education. 2024. №12. URL: <https://cyberleninka.ru/article/n/sovremennyy-russkiy-yazyk-2> (дата обращения: 12.11.2025).
3. Оразкулыев Б. Дискуссии о нормах русского языка: влияние глобализации и технологий на грамматические правила // ceteris paribus. 2024. №12. URL: <https://cyberleninka.ru/article/n/diskussii-o-normah-russkogo-yazyka-vliyanie-globalizatsii-i-tehnologiy-na-grammaticheskie-pravila> (дата обращения: 12.11.2025).

Интернет источники

- I. Интернет-портал «Экономика -Аналитика-Этимология». Статья «Производство». URL: <https://ecanet.ru/word/%D0%9F%D1%80%D0%BE%D0%B8%D0%B7%D0%B2%D0%BE%D0%B4%D1%81%D1%82%D0%B2%D0%BE> (дата обращения: 02.12.2025).
- II. Интернет-портал онлайн-коллекции словарей. Статья «Фильм». URL: <https://gufo.me/dict/semenov/%D1%84%D0%B8%D0%BB%D1%8C%D0%BC> (дата обращения: 05.12.2025).
- III. Официальный сайт Большой Российской Энциклопедии. Статья «производство». URL: <https://bigenc.ru/c/proizvodstvo-d9d5dc> (дата обращения: 02.12.2025).
- IV. Официальный сайт интернет-агентства Nimbi/ Cnfmz «Что такое сайт: Полное руководство по истории, возможностям и особенностям» URL: <https://nimbi.ru/blog/sait/> (дата обращения: 05.12.2025).

Приложение А

Чат-бот с материалами исследования @TrendUKON_bot

