|  |  |
| --- | --- |
|  | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ      ГОЛОВНОЙ УЧЕБНО-ИССЛЕДОВАТЕЛЬСКИЙ И МЕТОДИЧЕСКИЙ ЦЕНТР ПРОФЕССИОНАЛЬНОЙ РЕАБИЛИТАЦИИ ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

КАФЕДРА            СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

***К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ***

***НА ТЕМУ:***

***Речевой и слуховой тренажер***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

Студент       ИУ5Ц-84Б    **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_Папин А. В.\_\_**

(Группа) (Подпись, дата) (И.О.Фамилия)

Руководитель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_ Канев А. И.\_\_**

(Подпись, дата) (И.О.Фамилия)

*Москва, 2024 г.*

**Министерство науки и высшего образования Российской Федерации**



**Федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«Московский государственный технический университет имени**

**Н.Э. Баумана**

**(национальный исследовательский университет)»**

**(МГТУ им. Н.Э. Баумана)**

УТВЕРЖДАЮ

Заведующий кафедрой \_\_\_ИУ5\_\_\_

(Индекс)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_В.И. Терехов\_

(И.О.Фамилия)

« \_07\_ » \_\_февраля\_\_\_ 2024 г.

**ЗАДАНИЕ**

**на выполнение научно-исследовательской работы**

по теме \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Речевой и слуховой тренажер\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Студент группы \_\_\_\_ИУ5Ц-84Б\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Папин Алексей Владимирович\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Фамилия, имя, отчество)

Направленность НИР (учебная, исследовательская, практическая, производственная, др.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ИССЛЕДОВАТЕЛЬСКАЯ\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Источник тематики (кафедра, предприятие, НИР)                          КАФЕДРА                                \_\_\_\_\_\_

График выполнения НИР: 25% к \_\_\_ нед., 50% к \_\_\_ нед., 75% к \_\_ нед., 100% к \_\_\_ нед.

Техническое задание Разработать речевой и слуховой тренажер.   
Построить и исследовать метрики WER, CER, MER, WIL, IWER на основе введенного произношения и слушаемого контента. Провести тренировку речи с использованием выбранной лучшей модели и проанализировать результаты.

***Оформление научно-исследовательской работы:***

Расчетно-пояснительная записка на \_27\_ листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дата выдачи задания «07» февраля 2024 г.

**Руководитель НИР**  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_ **Канев А.И.**\_\_\_\_

(Подпись, дата) (И.О.Фамилия)

**Студент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_ Папин А.В.**\_\_\_

(Подпись, дата) (И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 4](#_Toc1)

1. [Постановка задачи 6](#_Toc2)
2. [Описание исходных данных и используемых методов 9](#_Toc3)
3. [Выполнение работы 11](#_Toc4)

[ЗАКЛЮЧЕНИЕ 26](#_Toc5)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 27](#_Toc6)

# **ВВЕДЕНИЕ**

Великий философ древности, Аристотель, изрек: «Скажи что-нибудь, я хочу тебя увидеть»[1]. Эти слова олицетворяют важность речи в формировании личного образа. Все её недостатки могут привести к трудностям в общении, которое мгновенно теряет компонент имиджа делового человека [2]. Одной из таких проблем является ротацизм, или картавость, или другие дефекты, которые встречаются у людей. С детства неправильное произношение звука может вызвать проблемы с самооценкой и затруднения в «речевых» профессиях, утверждают специалисты. В России примерно 10% населения сталкиваются с проблемами произношения звуков «р» и «рь». [3]

Несомненно, в современном образовательном процессе и реабилитации существует постоянная необходимость в инновационных подходах к развитию речевых и слуховых навыков. С учетом динамичного развития технологий, речевые и слуховые тренажеры представляют собой перспективное направление, способствующее эффективной коррекции и усилению данных навыков. Настоящая научно-исследовательская работа нацелена на изучение и анализ различных аспектов применения речевых и слуховых тренажеров в образовательной и реабилитационной практике.

Речевые и слуховые тренажеры представляют собой современные технологические решения, разработанные нейросетевыми технологиями для улучшения и коррекции речевой и слуховой активности. Эти инструменты предоставляют уникальные возможности для персонализированного обучения, основанного на потребностях конкретного человека. В сфере образования речевые тренажеры способствуют развитию навыков правильного произношения, расширению словарного запаса и повышению коммуникативных умений. В то время как слуховые тренажеры направлены на улучшение аудиального восприятия, способности к различению звуков и пониманию речи. Это особенно актуально для нашего факультета - ГУИМЦ.

Целью настоящего исследования является изучение применения технологий машинного обучения, в частности, предобученных моделей от Hugging Face, для анализа и коррекции дефектов речи и слуха. Основными задачами работы являются проведение анализа качества распознавания речи с использованием метрик WER, MER, WIL, CER, IWER, а также разработка методов и подходов к улучшению дикции и произношения и слуховых восприятий.

После изучения и выбора лучшей модели машинного обучения тренировать дикцию в течение нескольких дней. Полученные результаты провести исследование и зафиксировать динамику развития прогресса.

# **Постановка задачи**

Актуальность текущей задачи продолжает возрастать в связи возрастающей потребности в исправлении речевых дефектов и улучшении дикции. Особенно это важно для нашего факультета – ГУИМЦ. В связи с этим было принято решение о создании речевого и слухового тренажера. У каждого человека уникальный голос, который характеризуется индивидуальными особенностями: стилем произношения, тембром, интонацией, паузами, ударениями, громкостью и прочими параметрами. Поэтому было бы разумно воспользоваться нейросетевыми технологиями для создания речевого тренажера, способного учитывать все эти уникальные особенности голоса.

Однако существуют внешние факторы, которые могут затруднить обучение и использование тренажера, такие как шумы и посторонние звуки, мешающие правильному определению речевых дефектов у пользователя. Для преодоления этих проблем целесообразно использовать предварительно обученные модели, которые включают в себя различные датасеты с русским произношением и учитывающие посторонние препятствующие факторы.

В рамках выполнения поставленной задачи планируется использование метрик с целью определения точности и эффективности разработанного речевого тренажера:

1. Word Error Rate (WER);

где S — количество операций замены слов,

I — количество операций вставки слов,

D — количество операций удаления слов,

T — количество слов в эталонной фразе,

H — количество верно распознанных слов.

1. Character Error Rate (CER);

где S — количество операций замены слов,

I — количество операций вставки слов,

D — количество операций удаления слов,

C — количество правильных символов,

N — количество символов в ссылке (N = S + D + C).

1. Message Error Rate (MER);

где S — количество операций замены слов,

I — количество операций вставки слов,

D — количество операций удаления слов,

H — количество верно распознанных слов.

1. Word Information Loss (WIL);

где T0 — количество слов в гипотезе распознавания,

T — количество слов в эталонной фразе,

H — количество верно распознанных слов.

И самая подходящая метрика для русского языка с учетом особенности конструкции русского языка, например, для некоторых синтетических языков адекватным их структуре показателем является флективная ошибка распознавания слов IWER:

1. Inflectional Word Error Rate (IWER);

где — вес, который приписывает всем неверным заменам слов, которые приводят к замене лексемы слова, т.е. к грубым ошибкам распознавания,

— количество ошибок,

— меньший вес, который приписывает всем негрубым ошибкам в слова, например, где было неверно распознано окончание словоформы, но основа слова распознана правильно.

— количество негрубых ошибок,

Эти метрики позволят оценить качество распознавания речи и выявить возможные области для улучшения.

Также в рамках исследования включено, что после выбора лучшей модели необходимо провести исследование по тренировку своей дикции в течение нескольких дней. Полученные результаты после тренировок сделать исследование и зафиксировать изменения и в лучшую и в худшую сторону.

# **Описание исходных данных и используемых методов**

Для анализа и сравнения эффективности различных моделей распознавания речи на русском языке используется набор данных Common Voice, а именно нескольких наборов: «Common Voice Corpus 11», «Common Voice Corpus 12», «Common Voice Corpus 14». Рассмотрим один из датасетов этого набора – «Common Voice Corpus 11.0».

Датасет «Common Voice 11.0» представляет собой уникальные MP3-файлы и соответствующие текстовые файлы. Многие из 24 210 записанных часов в этом наборе данных также содержат демографическую метаинформацию, такую как возраст, пол и акцент, которая может помочь повысить точность систем распознавания речи. На текущий момент набор данных включает в себя 16 413 проверенных часов на 100 языках.

Рассмотрим следующий второй датасет – «Common Voice 12.0», он также представляет собой уникальные MP3-файлы и соответствующие текстовые файлы. В наборе данных содержится множество записанных часов - 26 119, из которых многие также содержат демографическую метаинформацию, такую как возраст, пол и акцент диктора. На текущий момент в наборе данных включено 17 127 проверенных часов на 104 языках.

Третий датасет – «Common Voice 14.0», который также содержит уникальные MP3-файлы и соответствующие текстовые файлы. В этом наборе данных содержится множество записанных часов - 28 117, из которых многие также содержат демографическую метаинформацию, такую как возраст, пол и акцент диктора. На текущий момент в наборе данных включено 18 651 проверенный час на 112 языках.

Каждый датасет доступен для скачивания и предварительной обработки с помощью библиотеки datasets в языке Python. Объект данных включает информацию о пути к аудиофайлу, предложение, количество положительных и отрицательных голосов, возраст и пол диктора, а также его акцент и локаль. Для работы с аудиофайлами в наборе данных используется словарь, содержащий путь к скачанному аудиофайлу, декодированный массив аудио и частота дискретизации. Предложения, включенные в датасет, являются теми, которые предлагалось произнести пользователю. Количество положительных и отрицательных голосов указывает на то, насколько хорошо аудиофайл оценен рецензентами. Для более точного анализа и сравнения результатов также учитываются возраст и пол диктора, а также его акцент и локаль.

После тренировки речи в течение нескольких дней был собран датасет, в котором включается следующее: моноканальный бинарный файл с частотой дискретизацией 32000, распознанный текст, канонический текст, дата записи, WER, CER, MER, WIL, IWER. Размер полученного датасета составляет 3020 уникальных строк.

# **Выполнение работы**

В каждой модели было проведено тщательное исследование, включающее в себя следующие этапы: загрузку модели: 'wav2vec2-large-100k-voxpopuli-ft-Common-Voice\_plus\_TTS-Dataset-russian' [4], 'wav2vec2-large-ru-golos' [5], 'wav2vec2-mbart50-ru' [6], 'wav2vec2-xls-r-1b-russian' [7] и 'vosk-model-ru-0.42' [8] и соответствующих нескольких датасетов, предобработку данных, которая включала ресемплирование звука при необходимости для изменения частоты дискретизации, а также токенизацию для получения предсказаний от модели. После этого производился расчет нескольких метрик, включая Word Error Rate (WER), Character Error Rate (CER), Message Error Rate (MER) и Word Information Loss (WIL). Для достижения основной цели и выбора наилучшей модели предполагается выполнение следующих задач: получение данных, исследование метрик, установка рейтингов, определение оптимальной модели и подготовка выводов и рекомендаций по применению лучшей модели.

Первоначально были загружены и объединены данные из файлов CSV для каждой модели с целью создания общего датафрейма для дальнейшего анализа, можно ознакомиться с таблицей 3.1. Далее был проведен тщательный исследовательский анализ, включающий сбор данных о метриках каждой модели. Полученный объединенный датафрейм был проанализирован на значения метрик WER, CER, MER и WIL для каждой модели и каждого датасета, с целью оценки эффективности моделей в различных условиях и на разных наборах данных. Для определения лучшей модели и ее применения на практике было принято решение о расчете рейтинга для каждой модели на основе суммы инвертированных значений метрик WER, CER, MER и WIL. Далее проводилась нормализация метрик для достижения 100% соотношения, а также пересчет рейтинга каждой модели в процентах относительно всех других моделей. В результате анализа данных и установления рейтинга эффективности моделей была определена оптимальная модель для распознавания речи на русском языке. Лучшей моделью оказалась 'bond005/wav2vec2-mbart50-ru', демонстрирующая следующие значения метрик: WER – 31%, CER – 12%, MER – 29% и WIL – 45%. Можно ознакомиться с таблицей 3.2, приведенной ниже.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Модель** | **Датасет** | **WER** | **CER** | **MER** | **WIL** |
| Edresson/wav2vec2-large-100k-voxpopuli-ft-Common-Voice\_plus\_TTS-Dataset-russian | common\_voice\_11 | 0,696 | 0,377 | 0,688 | 0,870 |
| common\_voice\_12 | 0,750 | 0,388 | 0,750 | 0,914 |
| common\_voice\_14 | 0,695 | 0,351 | 0,687 | 0,875 |
| bond005/wav2vec2-large-ru-golos | common\_voice\_11 | 0,385 | 0,071 | 0,385 | 0,606 |
| common\_voice\_12 | 0,478 | 0,110 | 0,478 | 0,715 |
| common\_voice\_14 | 0,433 | 0,088 | 0,433 | 0,679 |
| bond005/wav2vec2-mbart50-ru | common\_voice\_11 | 0,311 | 0,118 | 0,290 | 0,448 |
| common\_voice\_12 | 0,302 | 0,124 | 0,286 | 0,454 |
| common\_voice\_14 | 0,383 | 0,155 | 0,356 | 0,550 |
| jonatasgrosman/wav2vec2-xls-r-1b-russian | common\_voice\_11 | 0,367 | 0,094 | 0,367 | 0,589 |
| common\_voice\_12 | 0,393 | 0,100 | 0,384 | 0,607 |
| common\_voice\_14 | 0,390 | 0,093 | 0,381 | 0,603 |
| vosk-model-ru-0.42 | common\_voice\_11 | 0,380 | 0,158 | 0,380 | 0,600 |
| common\_voice\_12 | 0,369 | 0,135 | 0,365 | 0,577 |
| common\_voice\_14 | 0,341 | 0,130 | 0,341 | 0,555 |

Таблица 3.1 – Таблица результатов оценки эффективности моделей распознавания речи на русском языке по каждому датасету

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Модель** | **WER** | **CER** | **MER** | **WIL** | **Рейтинг** |
| bond005/wav2vec2-mbart50-ru | 31% | 12% | 29% | 45% | 100% |
| vosk-model-ru-0.42 | 37% | 13% | 36% | 58% | 82% |
| jonatasgrosman/wav2vec2-xls-r-1b-russian | 39% | 9% | 38% | 60% | 80% |
| bond005/wav2vec2-large-ru-golos | 43% | 9% | 43% | 68% | 69% |
| Edresson/wav2vec2-large-100k-voxpopuli-ft-Common-Voice\_plus\_TTS-Dataset-russian | 70% | 38% | 69% | 88% | 0% |

Таблица 3.2 – Таблица рейтинга эффективности моделей распознавания речи на русском языке

После выбора лучшей модели «bond005/wav2vec2-mbart50-ru» было решено потренировать свою речь в течение нескольких дней, начиная с 22 апреля по 17 мая с небольшими перерывами. Для наилучшего анализа и эффективного исследования было решено разбить на несколько разных тем общей сложности 6 тем и 179 уникальных слов:

* Повседневные разговоры:

Привет, здравствуй, дела, новости, планы, увлечения, погода, спорт, еда, фильмы, книги, отпуск, друзья, музыка, работа, хобби, здоровье, компьютер, животные, путешествия, учеба, семья, техника, встреча, отдых, финансы, покупки, событие, время, праздник, до свидания.

* Технические:

Робот, инженер, автоматизация, электроника, кластеризация, механизм, технология, схема, компонент, программное обеспечение, микроконтроллер, конструкция, инновация, кодирование, энергия, проектирование, технический рисунок, сенсор, трансформатор, оптика, датчик, реактор, прототип, печатная плата, алгоритм, двигатель, вычислитель, автоматика, цифровой.

* Бытовые:

Пылесос, стиральная машина, микроволновка, кофеварка, утюг, тостер, холодильник, телевизор, пылесборник, посудомоечная машина, блендер, электрочайник, вентилятор, увлажнитель, кондиционер, лампочка, весы, фен, мультиварка, электронагреватель, фильтр, чайник, усилитель сигнала, ручной миксер, видеодомофон, газовая плита, мусорный бак, кофемолка, противень, робот-пылесос.

* Научные:

Эксперимент, исследование, гипотеза, теория, модель, данные, публикация, анализ, результат, обнаружение, методика, лаборатория, эксперт, протокол, статистика, измерение, оборудование, исследователь, доказательство, заключение, вывод, инженерия, наблюдение, тестирование, реакция, контроль, моделирование, опыт, теорема, спектр.

* Космос:

Планета, звезда, галактика, космический корабль, орбита, астрономия, космонавт, созвездие, луна, космическая станция, телескоп, солнце, марс, сатурн, плутон, космическая прогулка, гравитация, космический мусор, астрофизика, черная дыра, путешествие, луч, аппарат, воздух, туризм, экспедиция, ракета, среда, костюм, полет.

* Медицина:

Здоровье, болезнь, лекарство, врач, хирургия, диагноз, терапия, вакцина, лечение, пациент, медицинский прибор, биология, анатомия, психология, иммунитет, операция, экстренная помощь, кардиология, неврология, онкология, реабилитация, история болезни, персонал, практика, скальпель, окулист, больница, анестез, осмотр.

Разнообразие слова по каждой теме раскрывает исследование шире и позволяет заметить пробелы, которых необходимо устранять. Рассмотрим динамику метрики по каждой теме, выбрав случайное слово. Для справки, чем меньше значение метрики, тем лучше.

* Повседневные разговоры - слово "Фильмы"

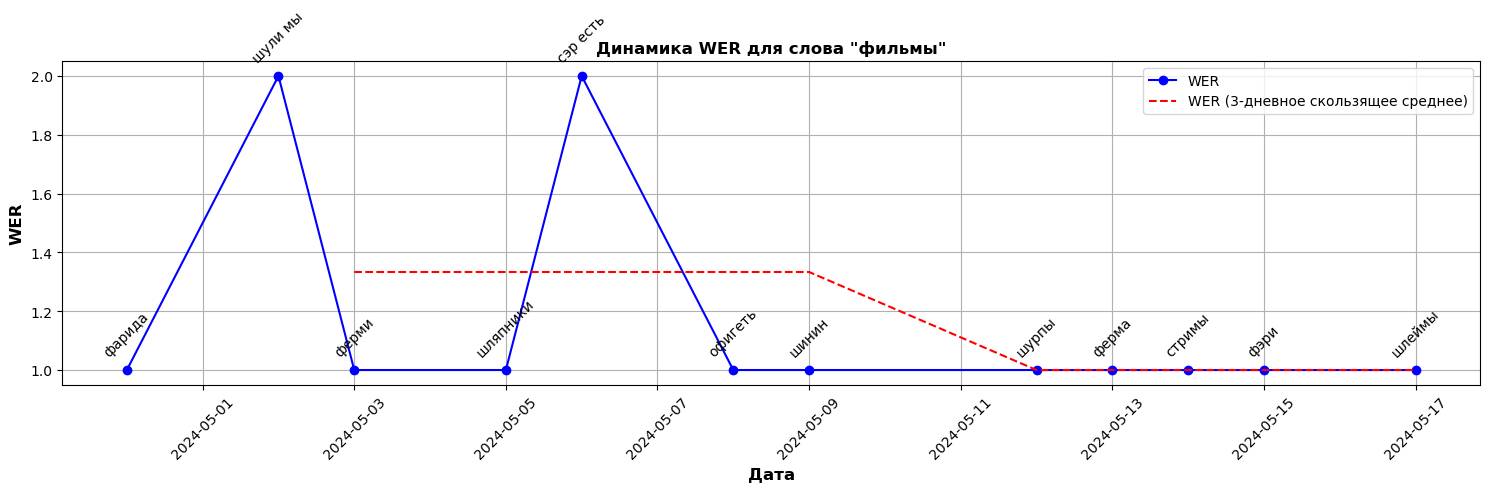


Рисунок 1 - Динамика WER для слова «Фильмы»

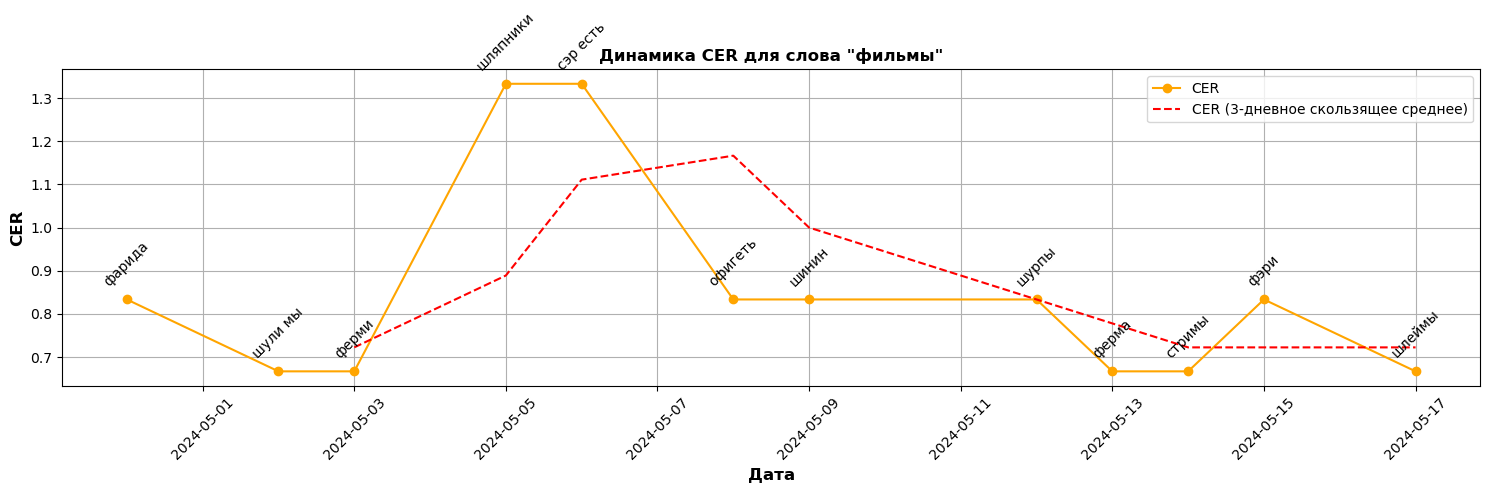


Рисунок 2 - Динамика CER для слова «Фильмы»

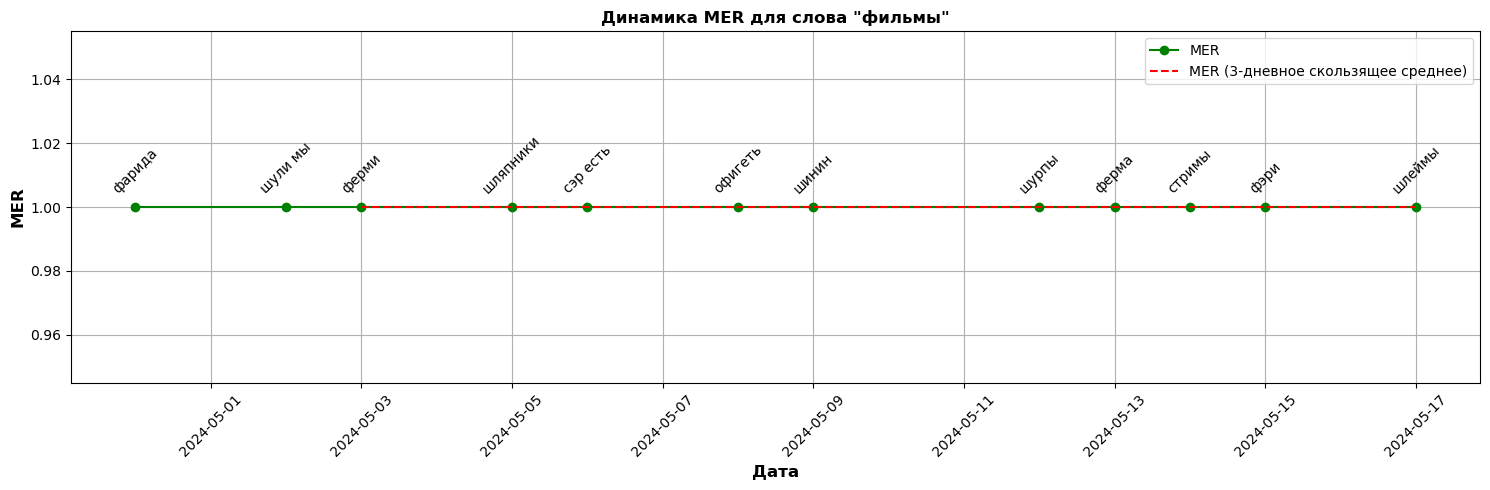


Рисунок 3 - Динамика MER для слова «Фильмы»

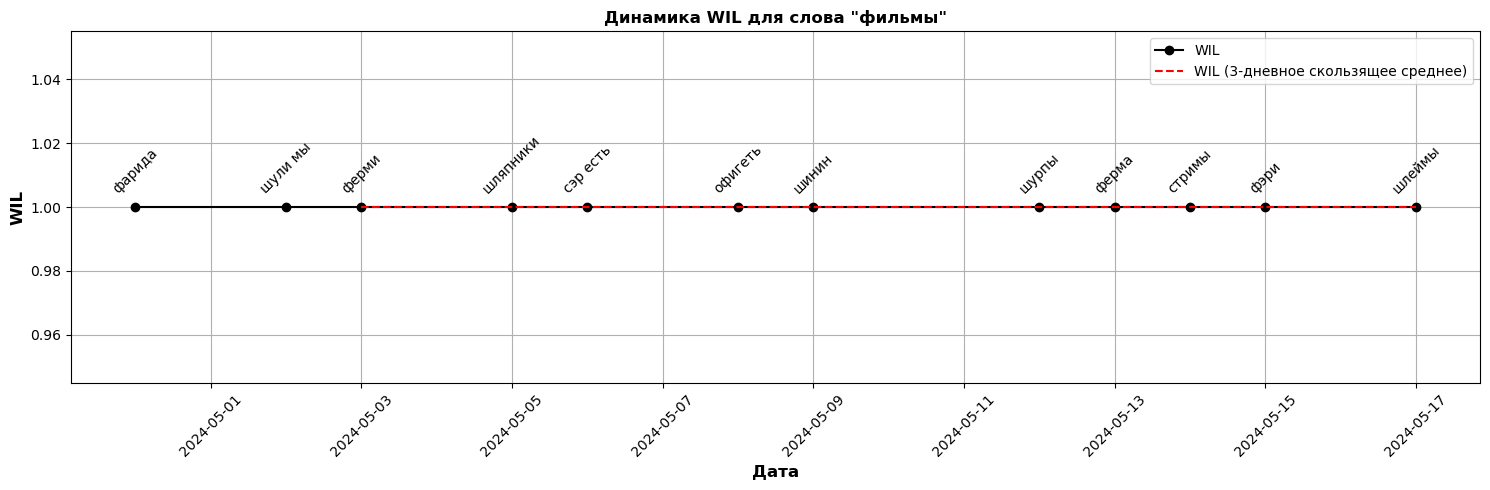


Рисунок 4 - Динамика WIL для слова «Фильмы»

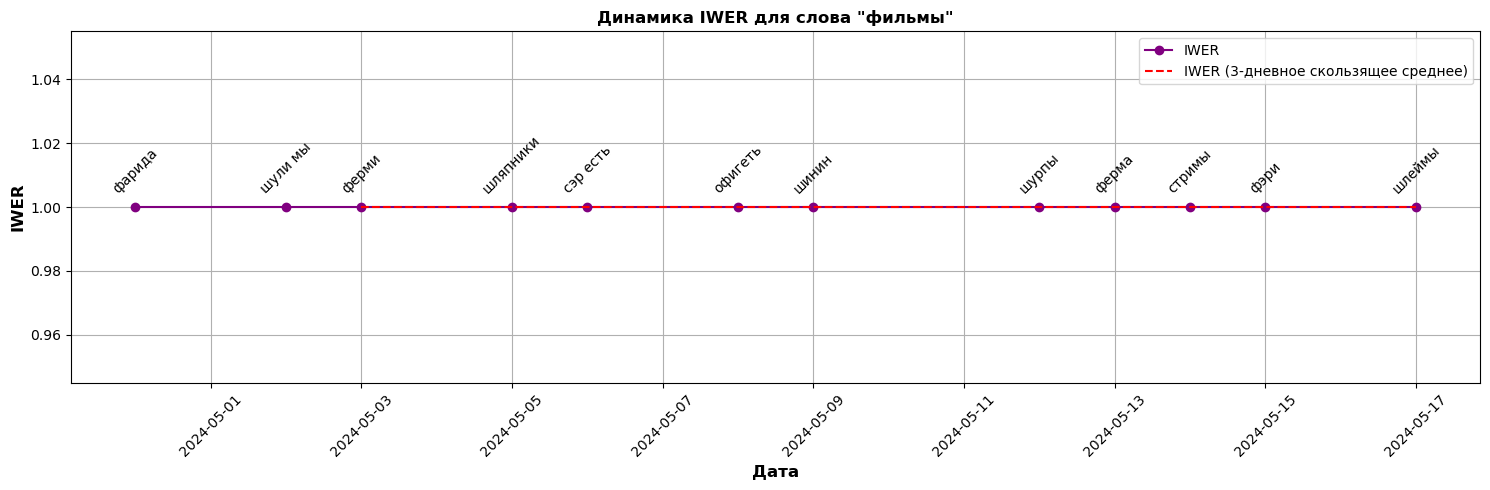


Рисунок 5 - Динамика IWER для слова «Фильмы»

Как и видим, что в последнее время мне удалось добиться оценку, удовлетворяющую метрики WER – 1 и CER – 0,7. С остальными метриками не происходит каких-либо значительных изменений.

* Технические - слово "Робот"

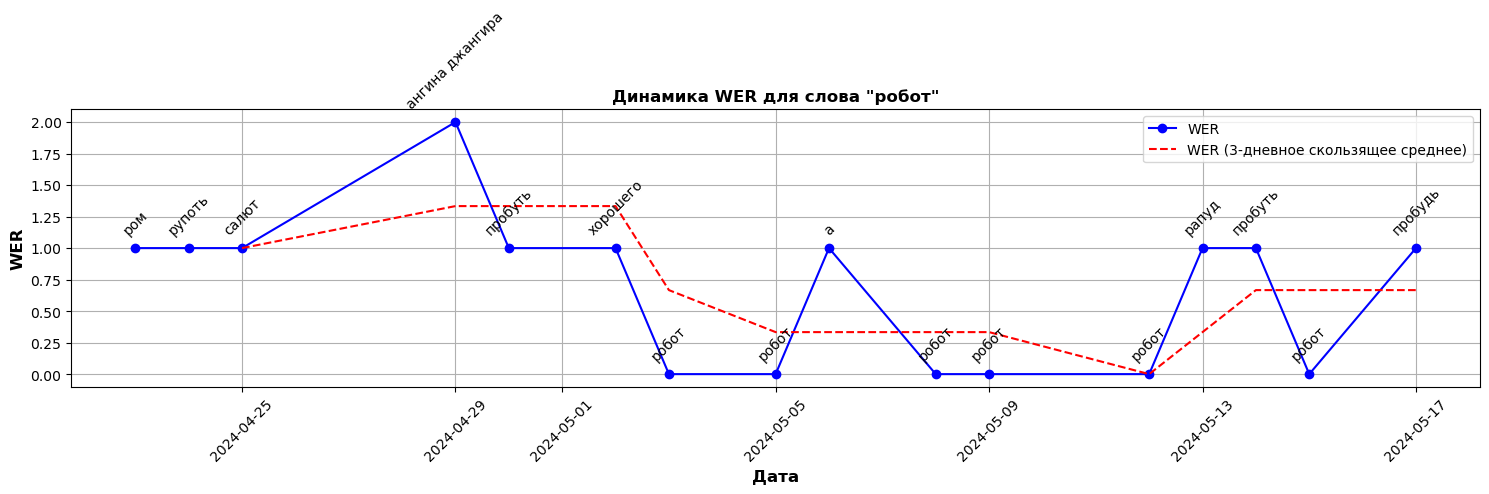


Рисунок 6 - Динамика WER для слова «Робот»

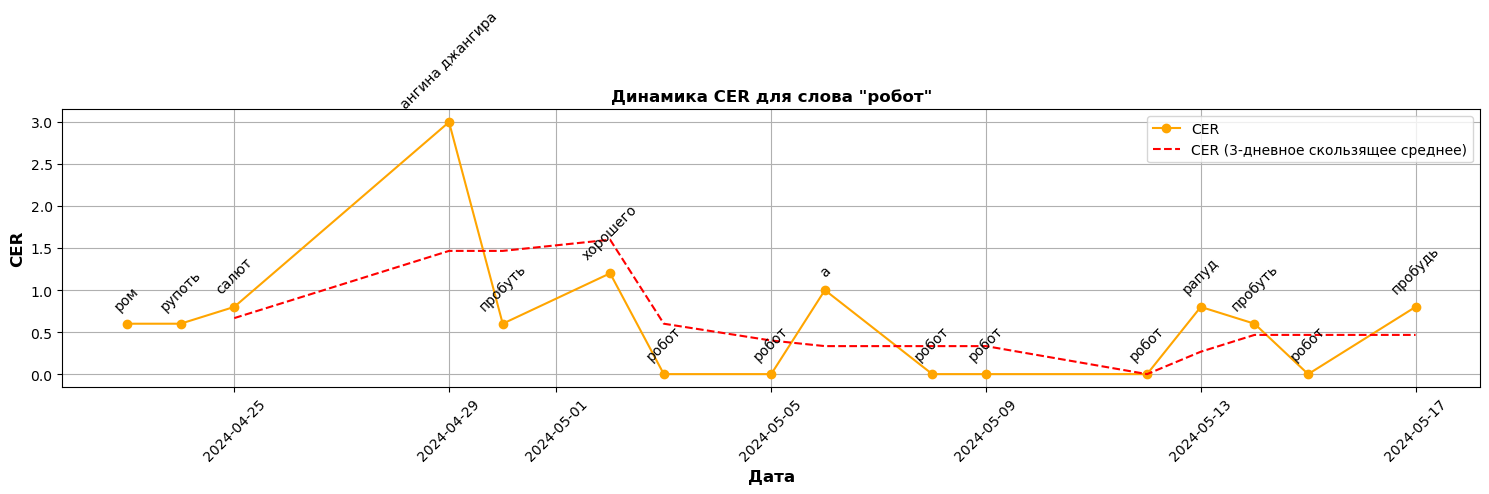


Рисунок 7 - Динамика СER для слова «Робот»

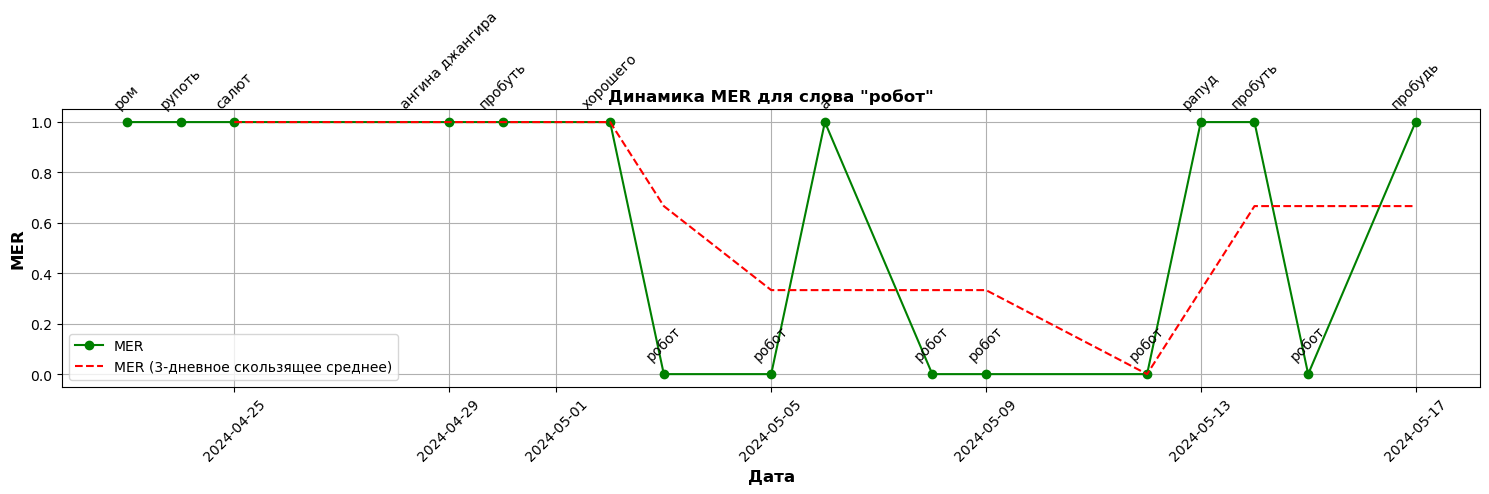


Рисунок 8 - Динамика MER для слова «Робот»

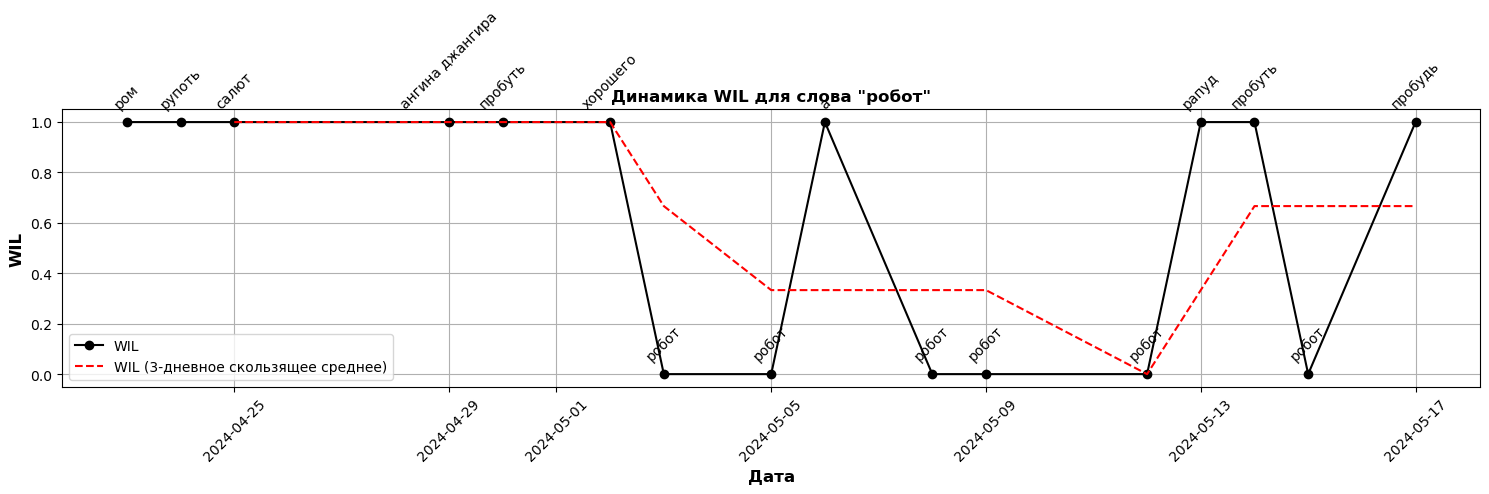


Рисунок 9 - Динамика WIL для слова «Робот»

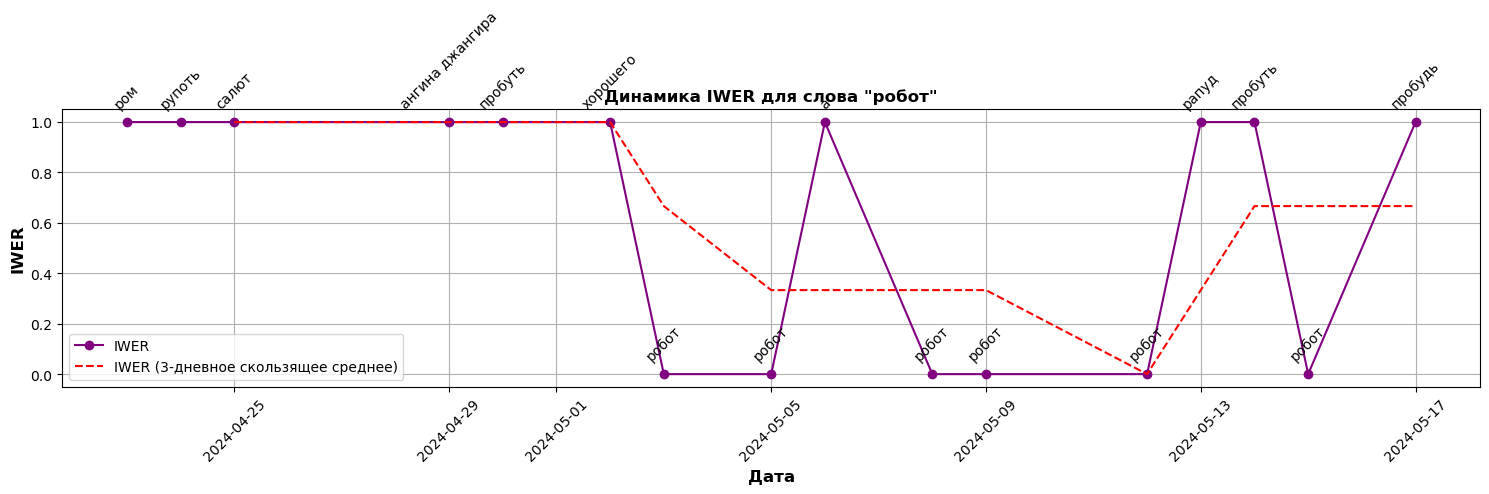


Рисунок 10 - Динамика IWER для слова «Робот»

Здесь также наблюдается позитивный рост улучшения дикции, если понаблюдать 3-дневное скользящее среднее, то метрики WER и CER держатся низкую планку в долгое время. С остальными метриками происходит скачкообразующее изменение, но не наблюдается критичное значительное изменение.

* Бытовые - слово "Микроволновка"

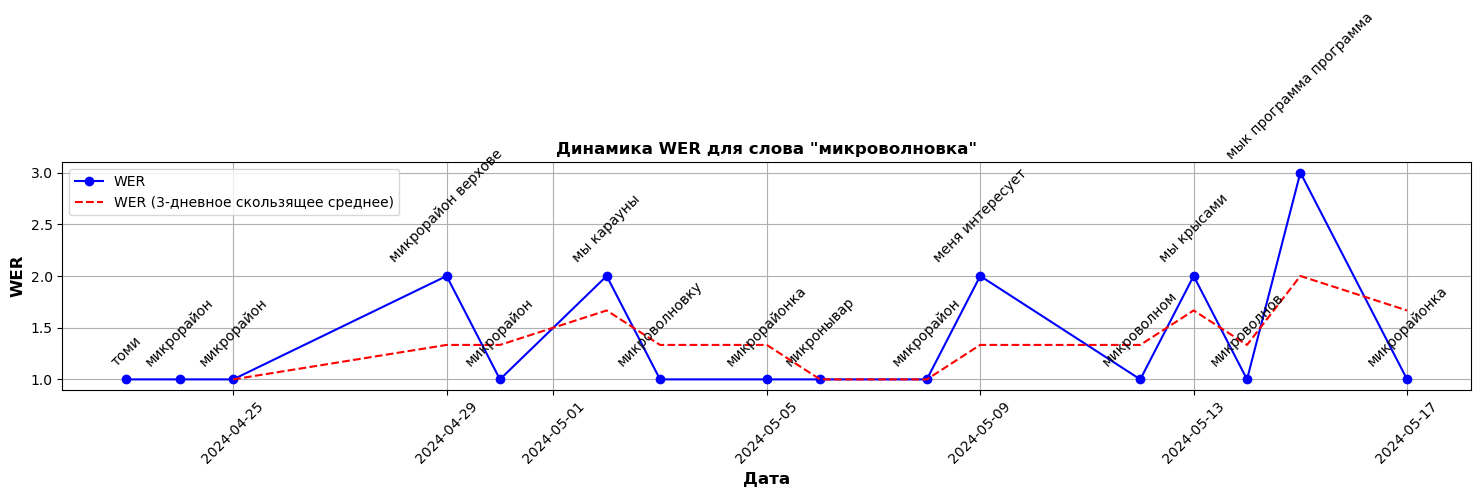


Рисунок 11 - Динамика WER для слова «Микроволновка»

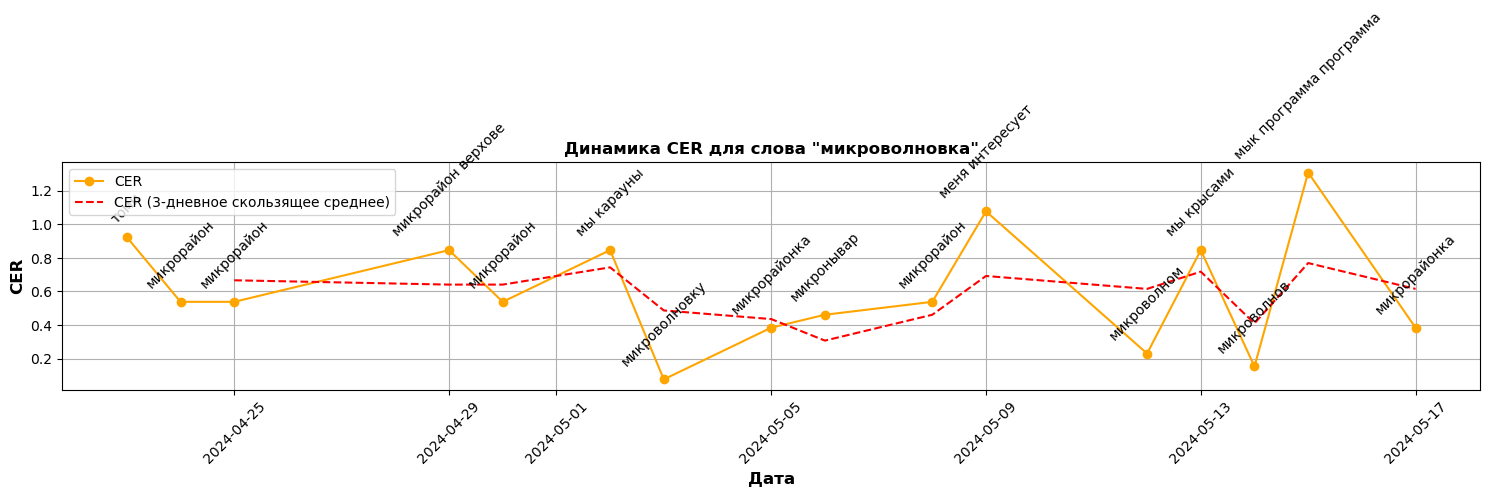


Рисунок 12 - Динамика CER для слова «Микроволновка»

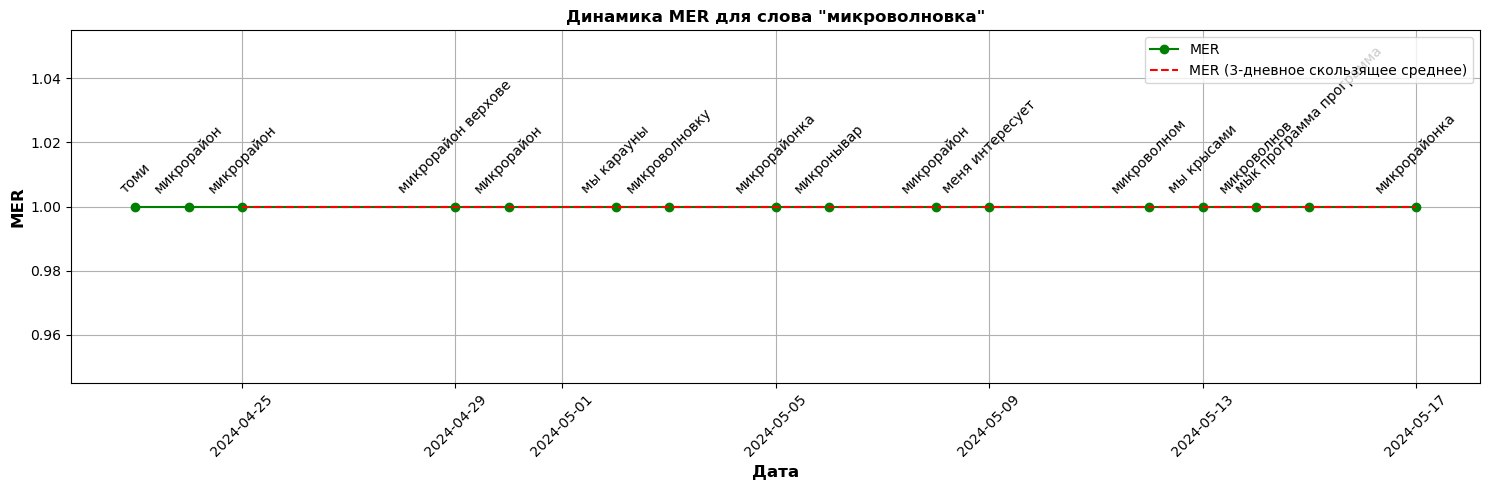


Рисунок 13 - Динамика MER для слова «Микроволновка»



Рисунок 14 - Динамика WIL для слова «Микроволновка»

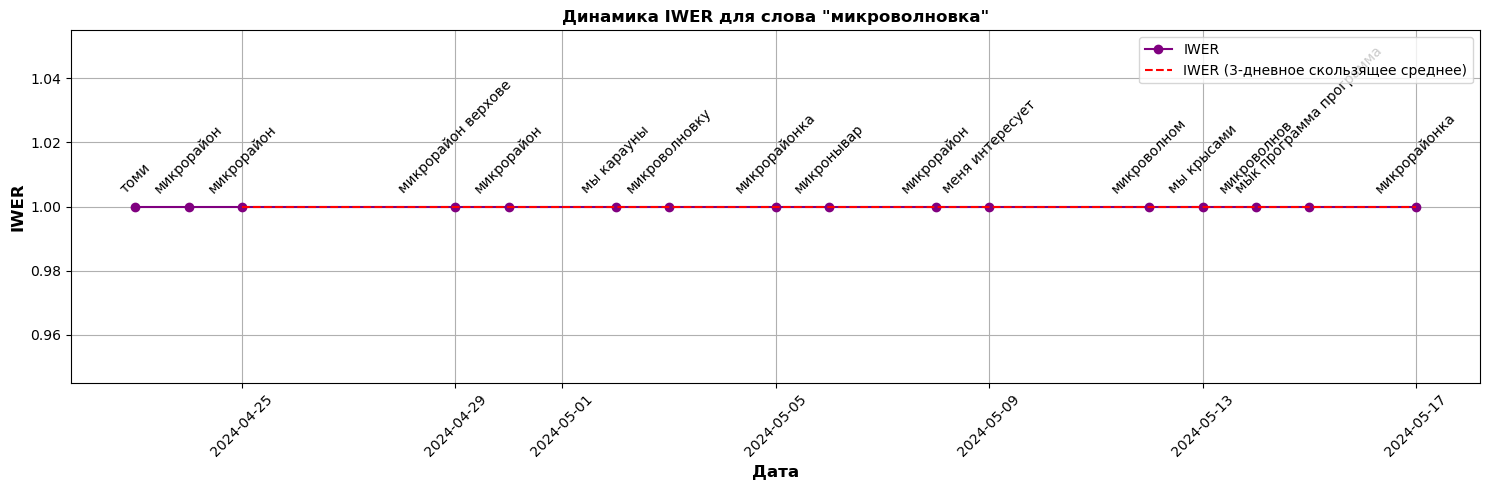


Рисунок 15 - Динамика IWER для слова «Микроволновка»

Здесь наблюдается стабильный рост, если понаблюдать 3-дневное скользящее среднее, то метрики WER и CER держатся одинаковую планку в долгое время. С остальными метриками не происходит каких-либо изменение.

* Научные - слово "Инженерия"

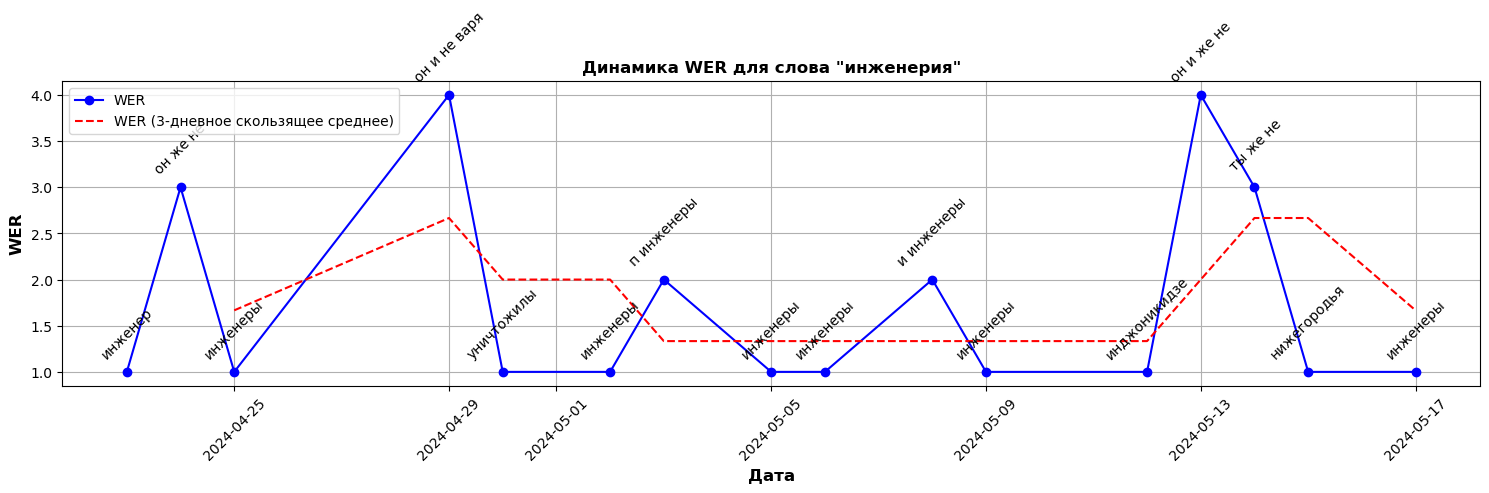


Рисунок 16 - Динамика WER для слова «Инженерия»

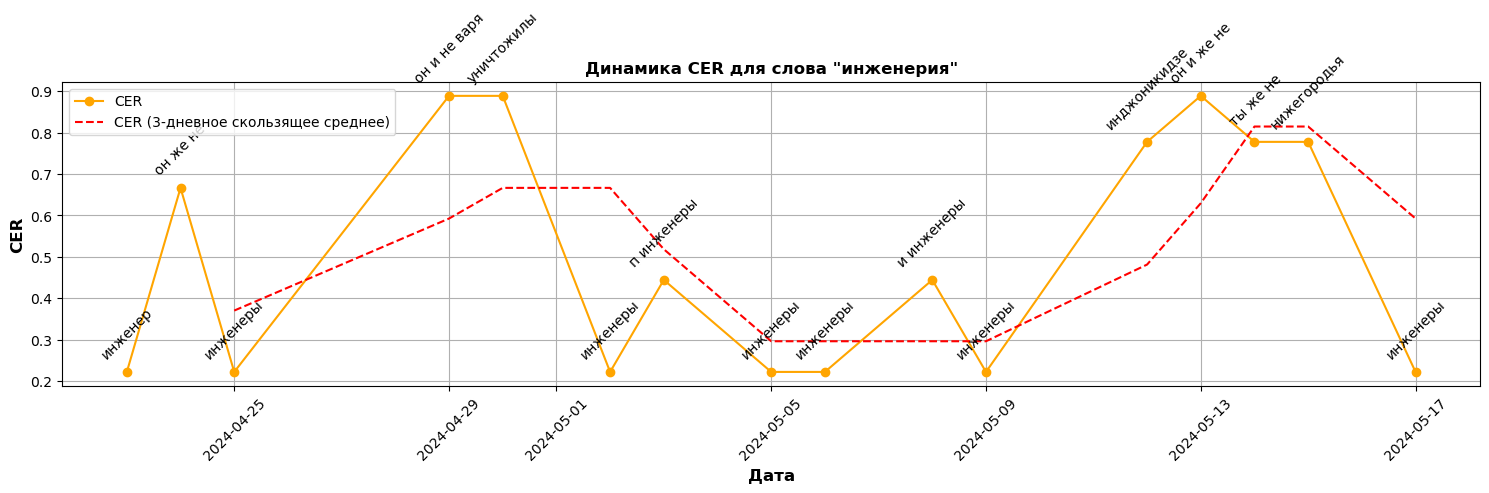


Рисунок 17 - Динамика CER для слова «Инженерия»

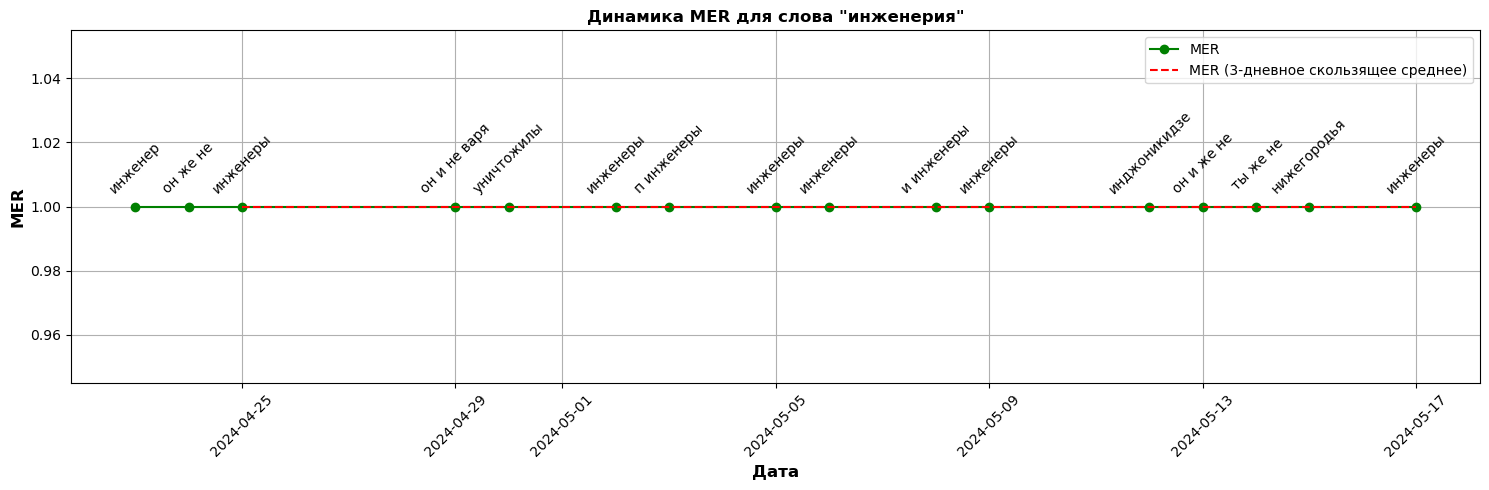


Рисунок 18 - Динамика MER для слова «Инженерия»

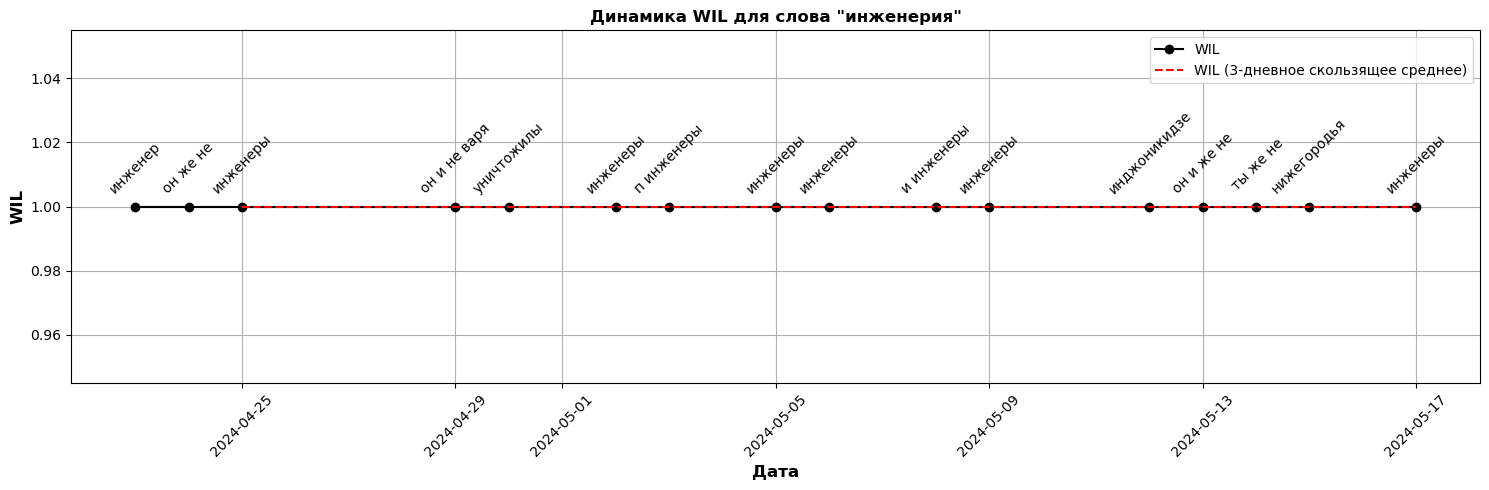


Рисунок 19 - Динамика WIL для слова «Инженерия»

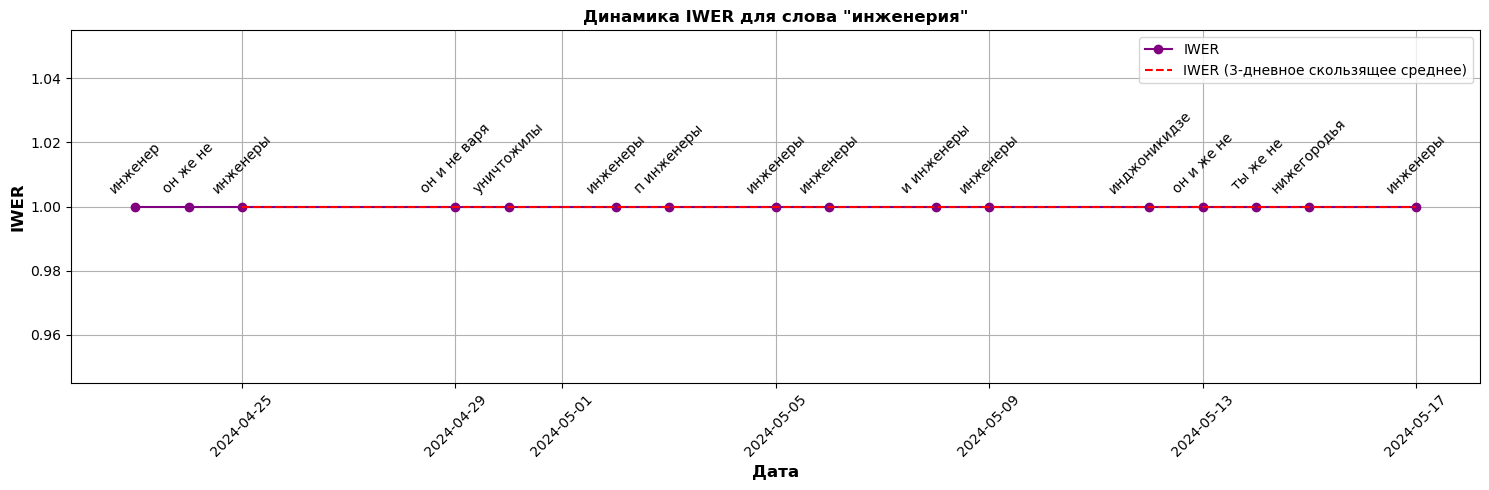


Рисунок 20 - Динамика IWER для слова «Инженерия»

Здесь также наблюдается позитивный рост улучшения дикции, если понаблюдать 3-дневное скользящее среднее, то метрика WER держится низкую планку в долгое время. А метрика CER в последнее время что-то подскочила, но вскоре вниз пошел. С остальными метриками не происходит каких-либо изменение.

* Космос - слово "Астрофизика"

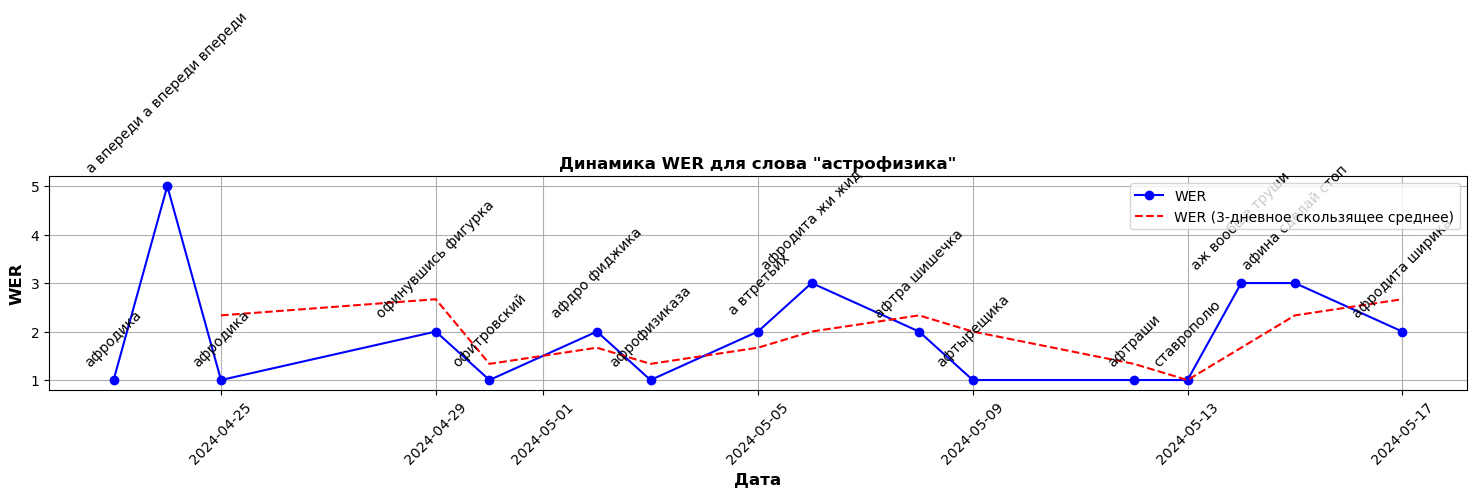


Рисунок 21 - Динамика WER для слова «Астрофизика»

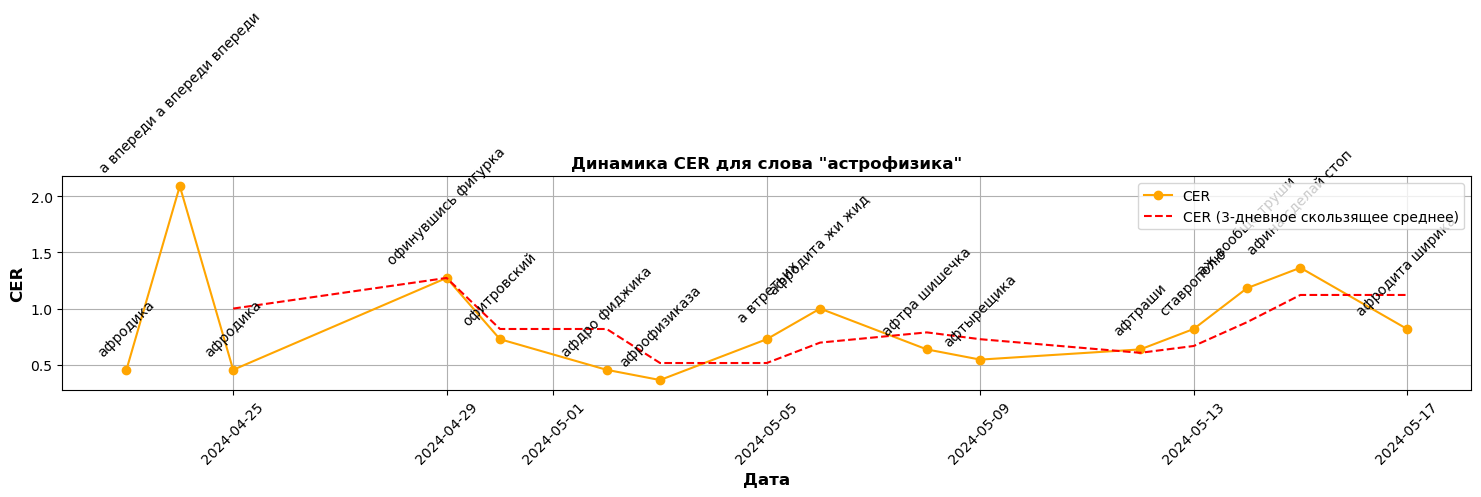


Рисунок 22 - Динамика CER для слова «Астрофизика»

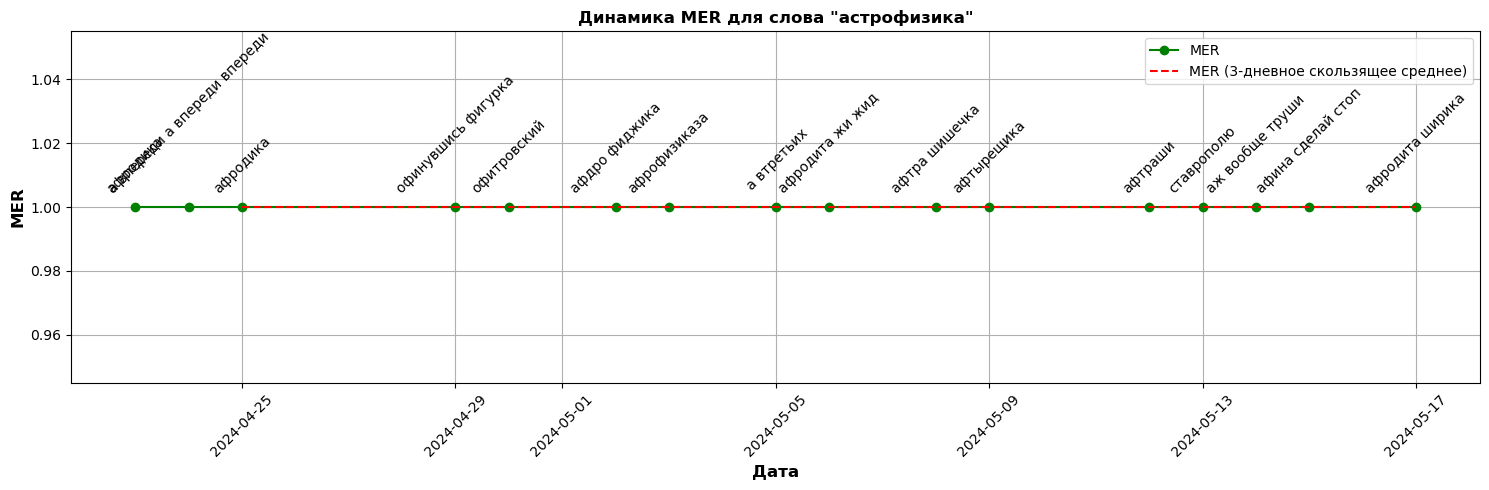


Рисунок 23 - Динамика WIL для слова «Астрофизика»



Рисунок 24 - Динамика WIL для слова «Астрофизика»

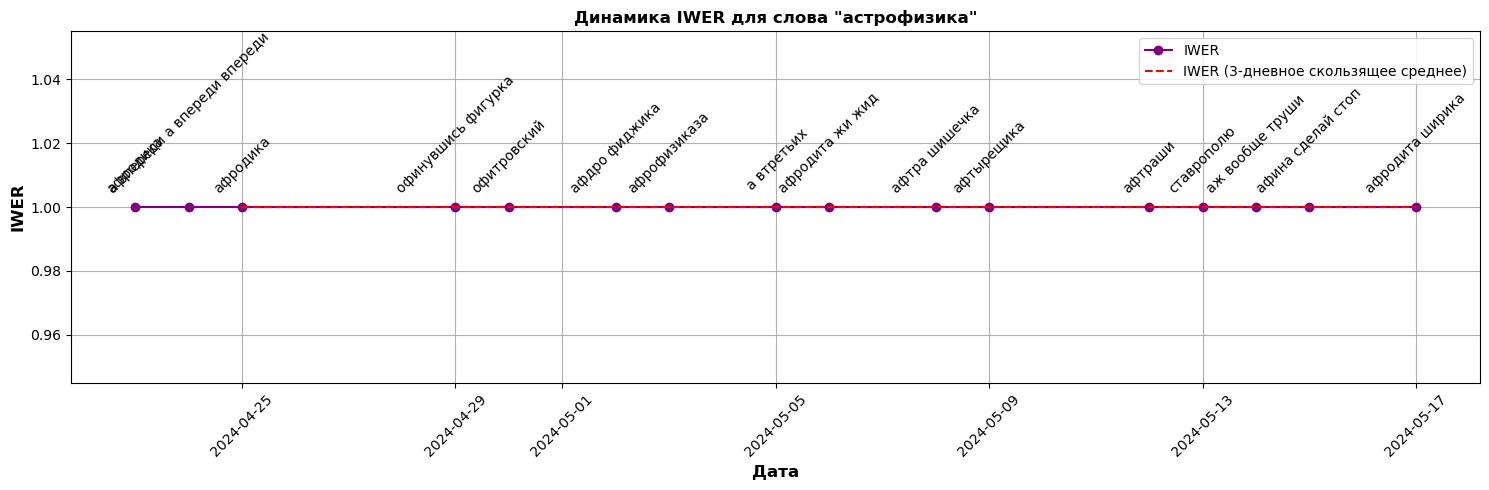


Рисунок 25 - Динамика IWER для слова «Астрофизика»

Здесь наблюдается стабильный рост, если понаблюдать 3-дневное скользящее среднее, то метрики WER и CER держатся одинаковую планку в долгое время. С остальными метриками не происходит каких-либо изменение.

* Медицина - слово "Биология"

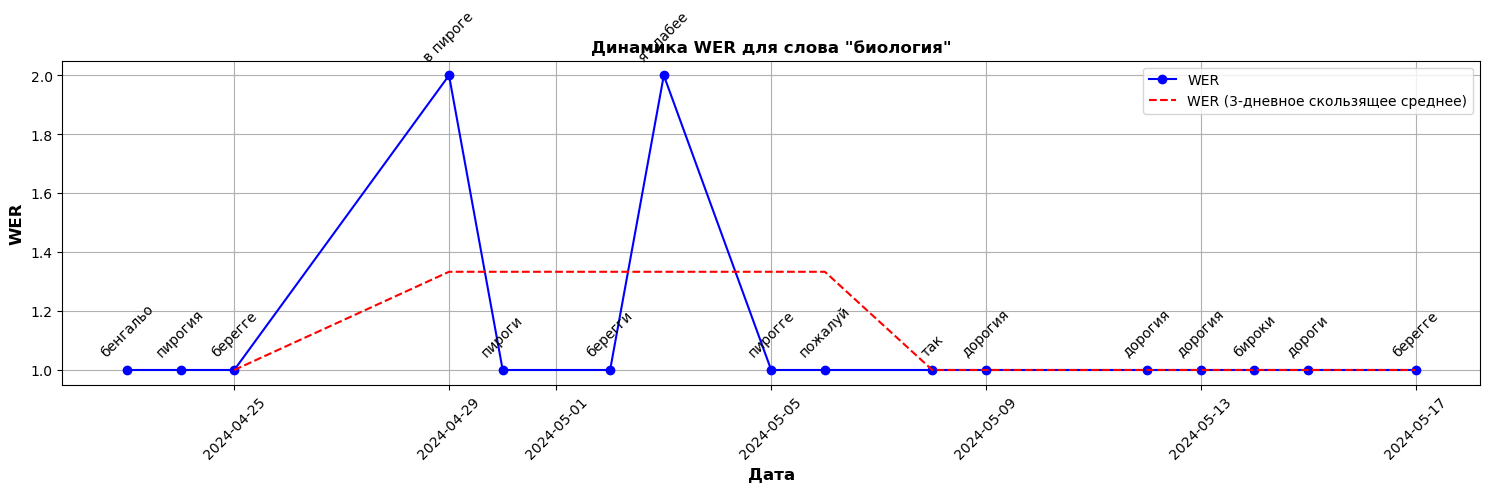


Рисунок 26 - Динамика WER для слова «Биология»

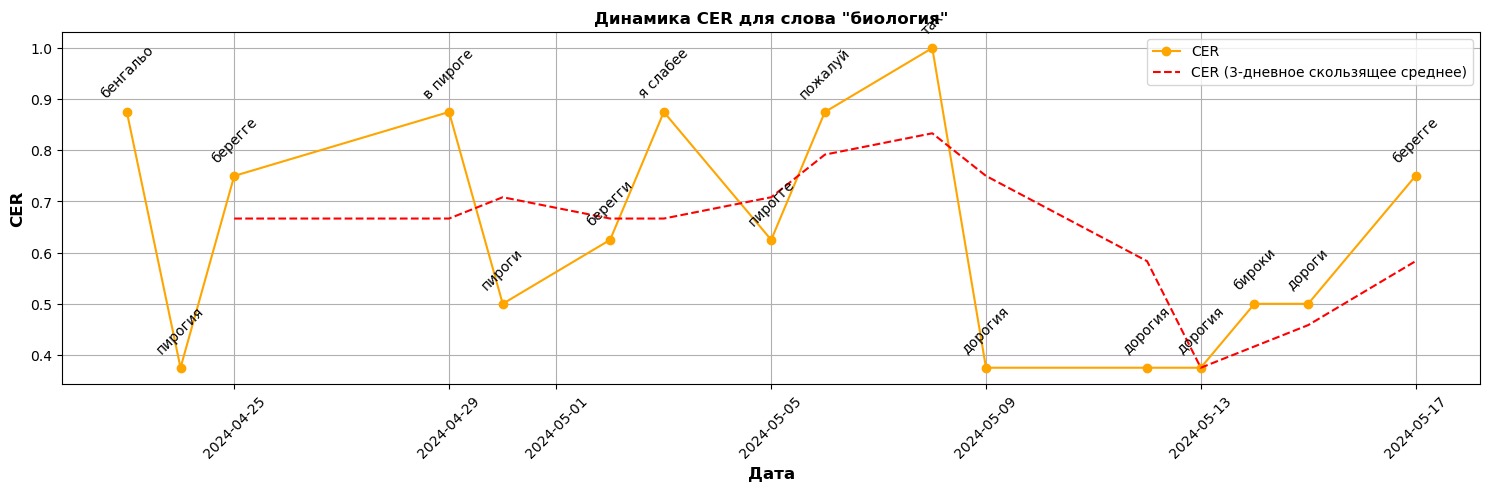


Рисунок 27 - Динамика CER для слова «Биология»

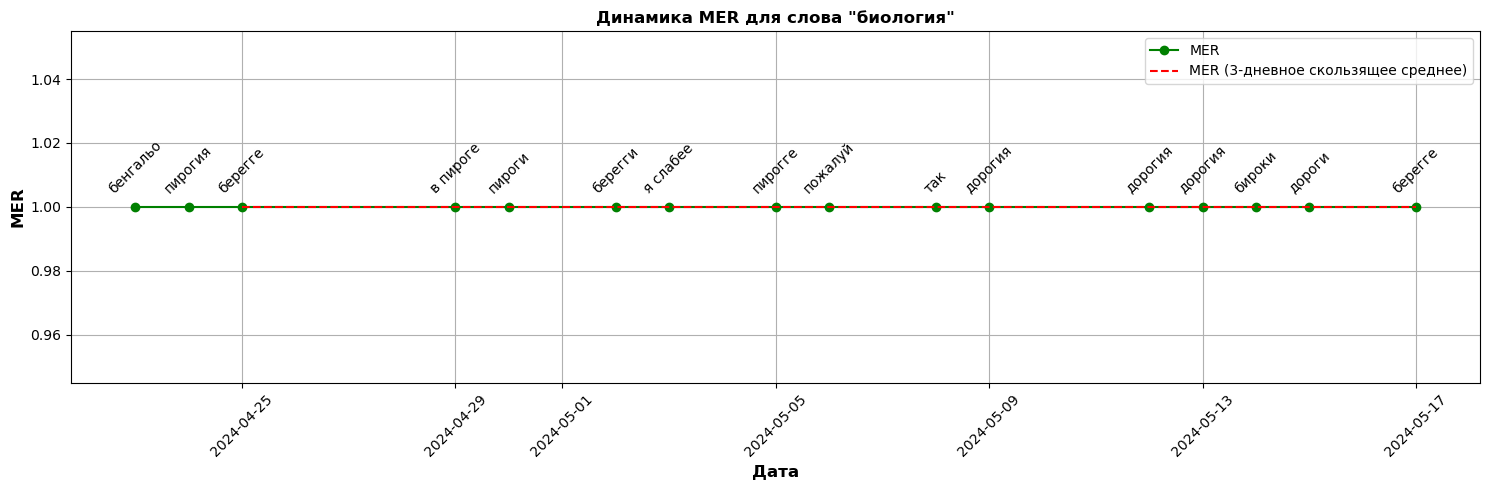


Рисунок 28 - Динамика MER для слова «Биология»

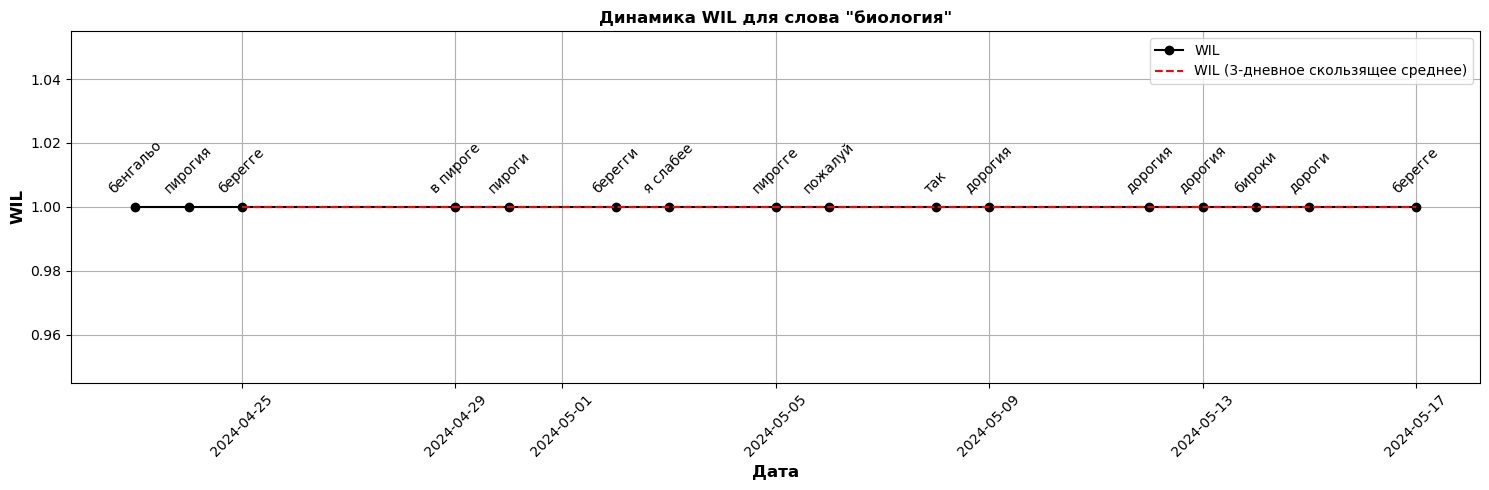


Рисунок 29 - Динамика WIL для слова «Биология»

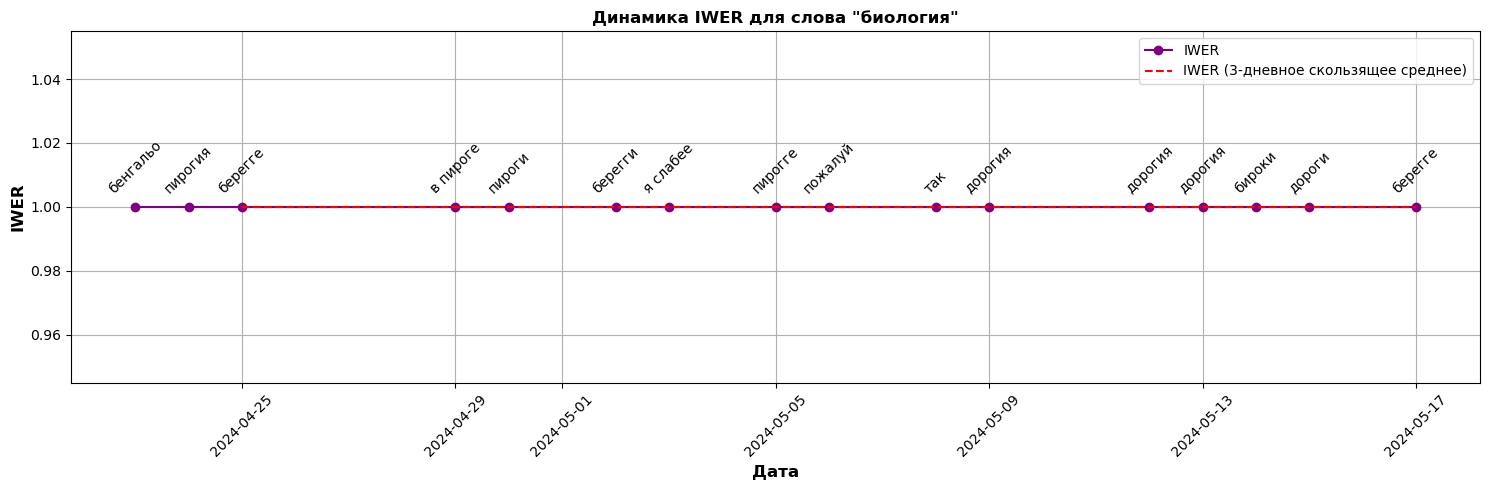


Рисунок 30 - Динамика IWER для слова «Биология»

Если понаблюдать 3-дневное скользящее среднее, то здесь наблюдается стабильный рост у метрики WER, а у метрики CER стремится к минимальному значению. С остальными метриками не происходит каких-либо изменение.

Рассмотрим общую картинку – изменение метрики по каждой теме.

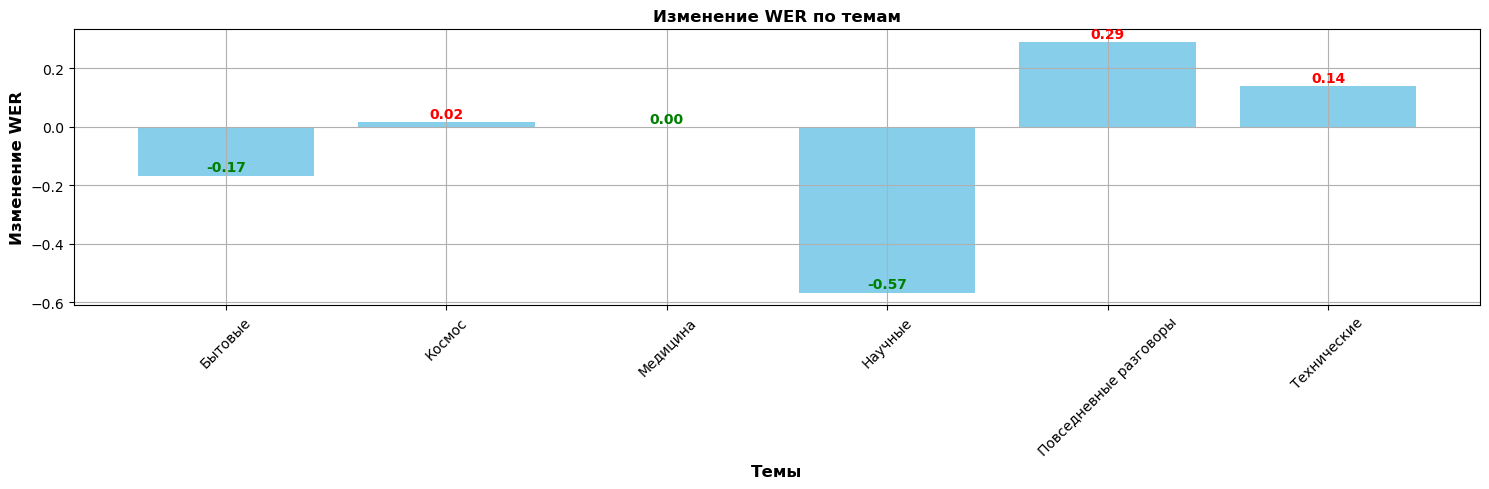


Рисунок 31 - Изменение WER по темам

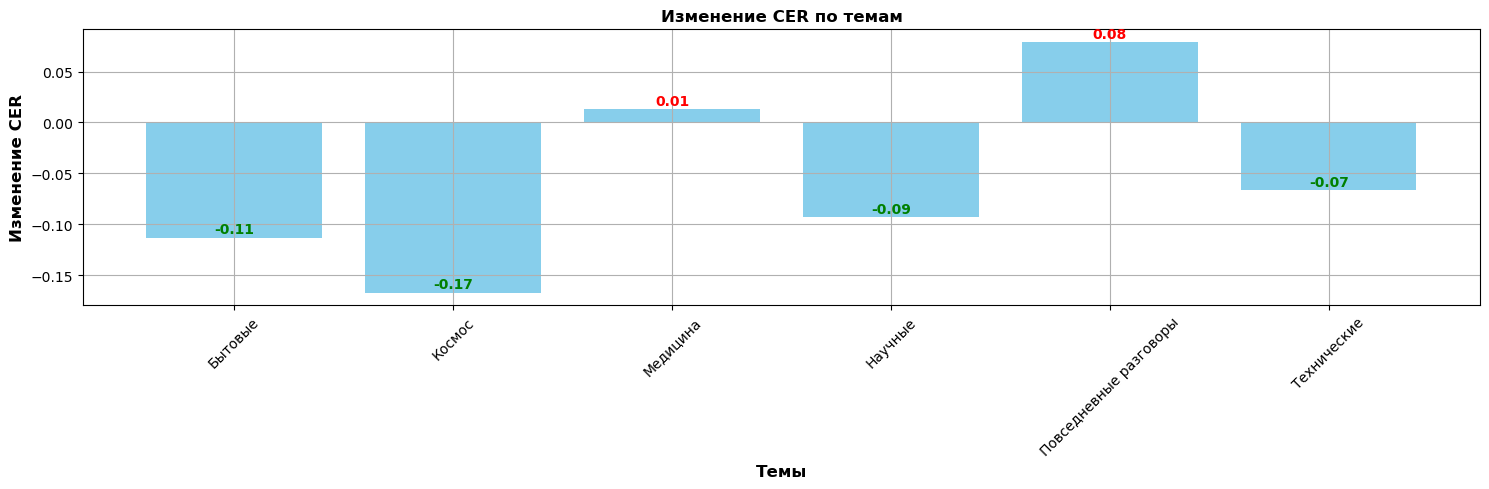


Рисунок 32 - Изменение CER по темам

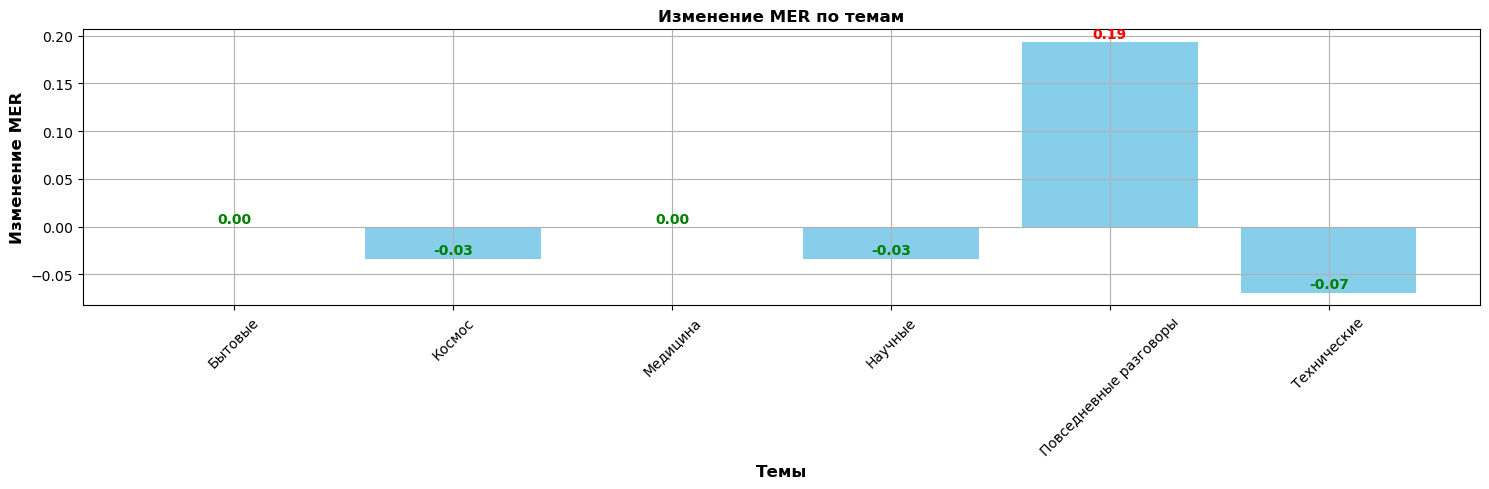


Рисунок 33 - Изменение MER по темам

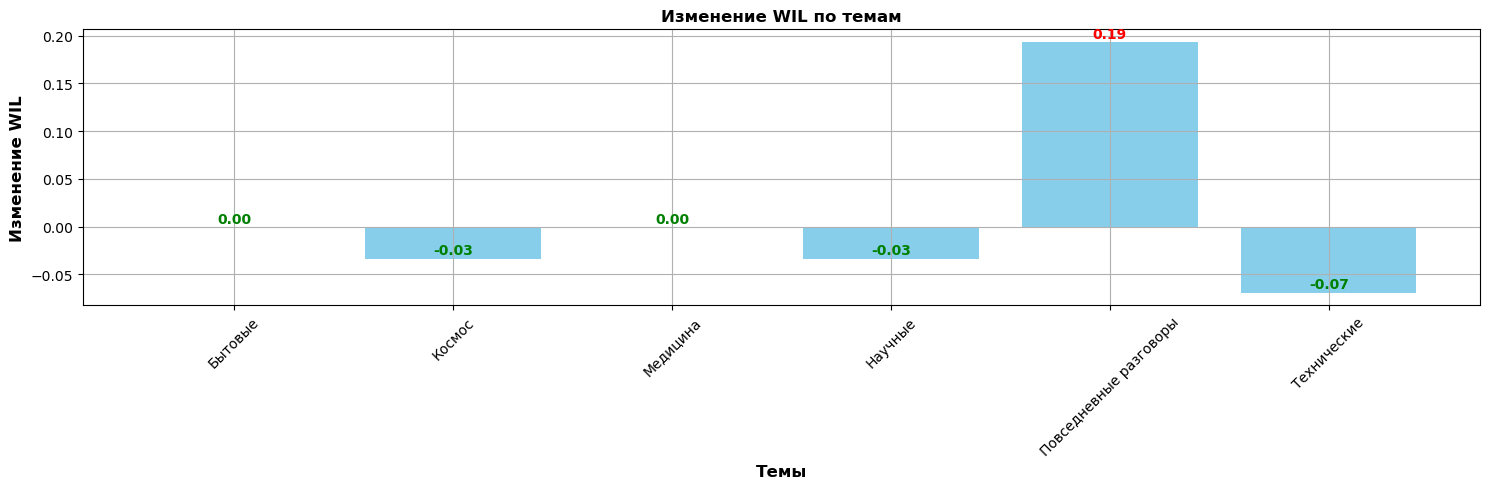


Рисунок 34 - Изменение WIL по темам

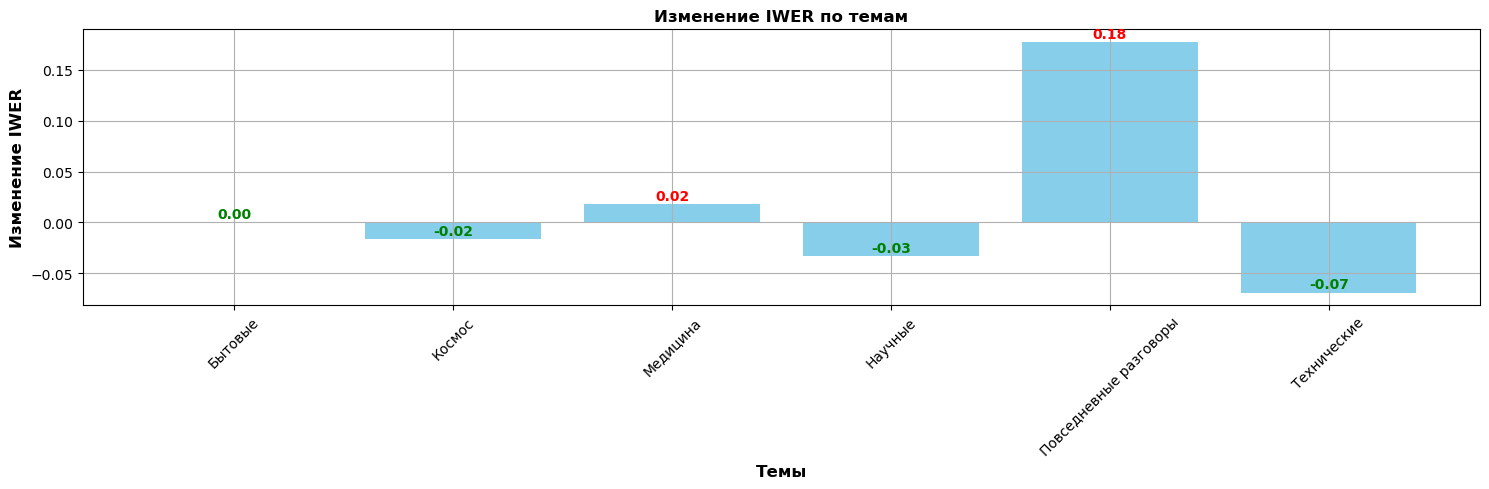


Рисунок 1 - Изменение IWER по темам

На основе вышеприведённого графика можно заметить значительный прогресс в моём произношении по научным и техническим темам. Иными словами, можно сказать, что со сверстниками я в основном общаюсь на русском жестовом языке, а с преподавателями — устной речью. Темы, поднимаемые для поддержания диалога с преподавателями, обычно находятся в рамках нашего учебного процесса, то есть касаются технических и научных терминов. Что и доказывают вышеперечисленные графики успеваемости моего произношения.

# **ЗАКЛЮЧЕНИЕ**

В результате выполнения данной научно-исследовательской работы были выполнены следующие задачи:

1. Была проведена тщательная аналитика каждой модели распознавания речи, включающая загрузку моделей: 'wav2vec2-large-100k-voxpopuli-ft-Common-Voice\_plus\_TTS-Dataset-russian', 'wav2vec2-large-ru-golos', 'wav2vec2-mbart50-ru', 'wav2vec2-xls-r-1b-russian' и 'vosk-model-ru-0.42', а также соответствующих датасетов. Это включало в себя предварительную обработку данных, включая ресемплирование аудиофайлов при необходимости, и токенизацию для получения предсказаний от моделей.
2. Был осуществлен расчет нескольких метрик, таких как Word Error Rate (WER), Character Error Rate (CER), Message Error Rate (MER) и Word Information Loss (WIL), для оценки производительности каждой модели.
3. Для выбора наилучшей модели были выполнены следующие этапы: получение данных, анализ метрик, установка рейтингов, определение оптимальной модели и подготовка выводов и рекомендаций по применению лучшей модели.
4. В ходе анализа данных была определена наиболее эффективная модель для распознавания речи на русском языке. Лучшей моделью оказалась 'bond005/wav2vec2-mbart50-ru', продемонстрировавшая высокие значения метрик: WER – 31%, CER – 12%, MER – 29% и WIL – 45%.
5. Проведена тренировка речи с использованием выбранной лучшей модели в течение нескольких дней.
6. В ходе анализа было выявлено, что темы, связанные с областью технических и научных направлений, способствовали улучшению моего произношения.
7. На основе тренировки и анализа были зафиксированы результаты, демонстрирующие прогресс в развитии дикции и произношения.

# **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Сократ цитаты и афоризмы: [Электронный ресурс]. // <https://quotar.org/>. URL: https://quotar.org/quotes/authors/sokrat (Дата обращения: 17.03.2024)
2. Вербальный имидж — как речь влияет на успех: [Электронный ресурс]. // <https://www.artimage.ru>. URL: <https://www.artimage.ru/info/articles/verbalnyy_imidzh/> (Дата обращения: 17.03.2024)
3. Дефекты речи: можно ли выступать на публике, если вы картавите или шепелявите?: [Электронный ресурс]. // https://www.7ya.ru/. [URL: https://www.7ya.ru/article/Defekty-rechi-mozhno-li-vystupat-na-publike-esli-vy-kartavite-ili-shepelyavite/](https://www.7ya.ru/article/Defekty-rechi-mozhno-li-vystupat-na-publike-esli-vy-kartavite-ili-shepelyavite/) (Дата обращения: 28.05.2023)
4. Edresson/wav2vec2-large-100k-voxpopuli-ft-Common-Voice\_plus\_TTS-Dataset-russian: [Электронный ресурс]. // <https://huggingface.co/>. URL: <https://huggingface.co/Edresson/wav2vec2-large-100k-voxpopuli-ft-Common-Voice_plus_TTS-Dataset-russian> (Дата обращения: 17.03.2024)
5. Bond005/wav2vec2-large-ru-golos: [Электронный ресурс]. // <https://huggingface.co/>. URL: <https://huggingface.co/bond005/wav2vec2-large-ru-golos> (Дата обращения: 17.03.2024)
6. Bond005/wav2vec2-mbart50-ru: [Электронный ресурс]. // <https://huggingface.co/>. URL: <https://huggingface.co/bond005/wav2vec2-mbart50-ru> (Дата обращения: 17.03.2024)
7. Jonatasgrosman/wav2vec2-xls-r-1b-russian: [Электронный ресурс]. // https://huggingface.co/. URL: <https://huggingface.co/jonatasgrosman/wav2vec2-xls-r-1b-russian> (Дата обращения: 17.03.2024)
8. Vosk-model-ru-0.42: [Электронный ресурс]. // <https://alphacephei.com/ru/>. URL: <https://alphacephei.com/vosk/models> (Дата обращения: 17.03.2024)