

Содержание

ВВЕДЕНИЕ	3
1 Аналитическая часть	4
2 Конструкторская часть	5
2.1 Дисперсия	5
2.2 Математическое ожидание	5
2.3 Ковариация	5
2.4 Корреляция Пирсона	5
2.5 Корреляция Спирмена	5
3 Технологическая часть	7
4 Исследовательская часть	10
4.1 Анализ статистических характеристик и отбор признаков	10
4.2 Анализ матриц сходства	11
4.3 Анализ топ-слов по группам и метрикам	12
4.4 Заключение и направления дальнейших исследований	12
4.5 Вывод	13
ЗАКЛЮЧЕНИЕ	14
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	15
Приложение А	16

ВВЕДЕНИЕ

Целью данной лабораторной работы является исследование характеристик набора текстовых данных для определения значимости различных параметров векторного пространства. В рамках исследования были применены методы статистического анализа для оценки дисперсии, математического ожидания, ковариации и корреляций (Пирсона и Спирмена) параметров, определяющих векторы документов. Анализ проводился как для всего набора документов, так и для каждой выделенной тематической группы. Основной задачей является выявление параметров пространства, которые могут быть удалены без существенной потери информативности, что может способствовать оптимизации обработки и анализа текстовых данных.

1 Аналитическая часть

В рамках данного исследования проводится анализ характеристик набора текстовых данных с целью определения значимости различных параметров векторного пространства. Основное внимание уделяется следующим статистическим показателям:

- Дисперсия
- Математическое ожидание
- Ковариация
- Корреляция Пирсона
- Корреляция Спирмена

Эти показатели рассчитываются как для всего набора документов, так и для каждой тематической группы отдельно. Такой подход позволяет выявить общие закономерности в данных и специфические особенности каждой темы.

Дисперсия отражает степень разброса значений параметра относительно его среднего значения. Высокая дисперсия может указывать на информативность параметра, так как он способен различать разные документы.

Математическое ожидание показывает среднюю частоту встречаемости слова в документах. Слова с высоким математическим ожиданием могут быть общими для многих документов и не обязательно несут специфическую информацию о теме.

Ковариация и корреляции (Пирсона и Спирмена) позволяют оценить взаимосвязь между параметрами. Сильно коррелирующие параметры могут содержать избыточную информацию, и один из них потенциально может быть удален.

Анализ этих показателей помогает выявить наиболее значимые слова для каждой темы и определить, какие параметры вносят наибольший вклад в различие документов разных тематик.

Вывод

В аналитической части были рассмотрены основные статистические показатели, используемые для анализа параметров векторного пространства текстовых документов. Обоснована важность расчета этих показателей как для всего набора данных, так и для отдельных тематических групп. Определены критерии, по которым можно оценивать значимость параметров и потенциальную возможность их удаления из векторного пространства.

2 Конструкторская часть

В данном разделе представлены математические формулы для расчета статистических показателей, используемых в анализе.

2.1 Дисперсия

Дисперсия характеризует разброс значений случайной величины относительно её математического ожидания:

$$\text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (2.1)$$

где X - случайная величина, $\mu = \mathbb{E}[X]$ - математическое ожидание, N - количество наблюдений, x_i - i-е наблюдение.

2.2 Математическое ожидание

Математическое ожидание дискретной случайной величины:

$$\mathbb{E}[X] = \mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.2)$$

2.3 Ковариация

Ковариация между двумя случайными величинами X и Y :

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_X)(y_i - \mu_Y) \quad (2.3)$$

где μ_X и μ_Y - математические ожидания X и Y соответственно.

2.4 Корреляция Пирсона

Коэффициент корреляции Пирсона:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^N (x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\sum_{i=1}^N (x_i - \mu_X)^2} \sqrt{\sum_{i=1}^N (y_i - \mu_Y)^2}} \quad (2.4)$$

где σ_X и σ_Y - стандартные отклонения X и Y .

2.5 Корреляция Спирмена

Коэффициент ранговой корреляции Спирмена:

$$\rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2.5)$$

где d_i - разность между рангами соответствующих значений X и Y , n - количество наблюдений.

Вывод

В конструкторской части были представлены математические формулы для расчета основных статистических показателей: дисперсии, математического ожидания, ковариации, корреляции Пирсона и Спирмена. Эти формулы являются основой для анализа характеристик набора данных и определения значимости параметров векторного пространства.

3 Технологическая часть

Для реализации анализа характеристик набора текстовых данных был использован язык программирования Python версии 3.12. Разработка велась в среде Jupyter Notebook, что позволило интерактивно выполнять код и анализировать промежуточные результаты.

Основные используемые библиотеки:

- `pandas` - для эффективной работы с табличными данными и расчета статистических показателей
- `numpy` - для выполнения математических операций с массивами
- `matplotlib` и `seaborn` - для визуализации результатов анализа
- `scipy` - для расчета расстояний между векторами

Ключевые этапы реализации:

- 1) Загрузка и предварительная обработка текстовых данных
- 2) Расчет статистических показателей (дисперсия, математическое ожидание, ковариация, корреляции)
- 3) Анализ топ-слов по различным метрикам
- 4) Сравнение результатов между тематическими группами
- 5) Отбор признаков на основе дисперсии и корреляции
- 6) Расчет матриц сходства с использованием различных метрик

Основные реализованные функции:

- 1) `calculate_stats()` - рассчитывает статистические показатели для всего набора данных и для каждой группы.
- 2) `top_words()` - выбирает топ-N слов по заданному критерию.
- 3) `compare_metrics_for_group()` - сравнивает топ-слова по различным метрикам для конкретной группы.
- 4) `compare_groups_for_metric()` - сравнивает топ-слова между группами для конкретной метрики.
- 5) `find_variance_threshold()` - находит пороговое значение дисперсии для отбора признаков.
- 6) `find_correlation_threshold()` - находит пороговое значение корреляции для отбора признаков.
- 7) `select_features()` - выполняет отбор признаков на основе пороговых значений дисперсии и корреляции.
- 8) `cosine_similarity()` - вычисляет косинусное сходство между векторами.
- 9) `ohai_similarity()` - вычисляет сходство Охай между векторами.
- 10) `euclidean_distance()` - вычисляет евклидово расстояние между векторами.
- 11) `calculate_similarity_matrix()` - создает матрицу сходства для всех текстов.
- 12) `plot_heatmap()` - визуализирует матрицу сходства в виде тепловой карты.

Ниже представлен фрагмент кода для расчета статистических показателей:

```

def calculate_stats(X, y):
    full_stats = {
        'variance': X.var(),
        'mean': X.mean(),
        'covariance': X.cov(),
        'pearson_corr': X.corr(method='pearson'),
        'spearman_corr': X.corr(method='spearman')
    }

    group_stats = {}
    for group in y.unique():
        group_data = X[y == group]
        group_stats[group] = {
            'variance': group_data.var(),
            'mean': group_data.mean(),
            'covariance': group_data.cov(),
            'pearson_corr': group_data.corr(method='pearson'),
            'spearman_corr': group_data.corr(method='spearman')
        }

    return full_stats, group_stats

```

Ниже представлен фрагмент кода для отбора признаков:

```

def select_features(X, variance_threshold, correlation_threshold):
    # Удаление признаков с низкой дисперсией
    high_variance_features = X.columns[X.var() >= variance_threshold]
    X_high_var = X[high_variance_features]

    # Удаление сильно коррелирующих признаков
    corr_matrix = X_high_var.corr().abs()
    upper_tri = corr_matrix.where(np.triu(np.ones(corr_matrix.shape), k
                                          =1).astype(bool))
    to_drop = [column for column in upper_tri.columns if any(upper_tri[
        column] >= correlation_threshold)]

    return X_high_var.drop(columns=to_drop)

target_features = 30 # Желаемое количество признаков
variance_threshold = find_variance_threshold(X, target_features)
correlation_threshold = find_correlation_threshold(X, target_features)

```

```
X_selected = select_features(X, variance_threshold,
                             correlation_threshold)
```

Для визуализации результатов была реализована функция построения тепловой карты:

```
def plot_heatmap(similarity_matrix: np.ndarray, title: str, vmin=0,
                  vmax=1):
    plt.figure(figsize=(12, 10))
    sns.heatmap(similarity_matrix, cmap='YlGnBu',
                xticklabels=True, yticklabels=True,
                cbar_kws={'label': 'Сходство/Расстояние'}, vmin=vmin, vmax=vmax)
    plt.title(title, fontsize=16)
    plt.xlabel('Номер текста', fontsize=12)
    plt.ylabel('Номер текста', fontsize=12)
    plt.tight_layout()
    plt.savefig(f"{title}.png")
    plt.show()
```

Вывод

В ходе технологической части была реализована программа на языке Python для анализа характеристик набора текстовых данных. Были рассчитаны различные статистические показатели, проведён авто-анализ топ-слов по метрикам и выполнено сравнение результатов между тематическими группами. Реализован алгоритм отбора признаков на основе дисперсии и корреляции, что позволило уменьшить размерность данных. Также были реализованы функции для расчета сходства между текстами с использованием различных метрик и визуализации результатов. Полученные результаты позволят оценить значимость различных параметров векторного пространства и определить потенциальных кандидатов на удаление.

4 Исследовательская часть

Целью исследования было проведение анализа характеристик набора текстовых данных с использованием различных метрик сходства и оценка эффективности применения методов отбора признаков для снижения размерности данных.

Исследование проводилось на следующем оборудовании:

- Устройство: Marqich-LAPTOP
- Процессор: 13th Gen Intel(R) Core(TM) i7-13700H 2.40 GHz
- Оперативная память: 16,0 ГБ
- Операционная система: Windows 11, 64-разрядная, процессор x64

Исследование проводилось в несколько этапов:

- 1) Предварительная обработка текстовых данных и их векторизация.
- 2) Анализ статистических характеристик полученных векторов.
- 3) Отбор признаков на основе дисперсии и корреляции.
- 4) Расчет матриц сходства с использованием различных метрик (косинусное сходство, метрика Охай, евклидово расстояние).
- 5) Визуализация результатов в виде тепловых карт.
- 6) Сравнительный анализ результатов до и после отбора признаков.

4.1 Анализ статистических характеристик и отбор признаков

В результате векторизации текстовых данных было получено 3630 признаков. Для снижения размерности данных был применен метод отбора признаков на основе дисперсии и корреляции. В результате количество признаков было уменьшено до 19.

Параметр	Значение
Исходное количество признаков	3630
Количество признаков после отбора	19
Пороговое значение дисперсии	4.063673469387755
Пороговое значение корреляции	0.6679196697561163

Таблица 4.1 — Результаты отбора признаков

Оставшиеся после отбора признаки: анамнез, состояние, уровень, спортивный, комментатор, бензин, физический, должный, анализ, закон, синтез, заболевание, химия, ставка, органический, пациент, метод, вещество, лечебный.

Анализ оставшихся признаков показывает, что они относятся к различным тематическим областям. Это свидетельствует о том, что отобранные признаки хорошо отражают разнообразие тем в исходном наборе данных.

4.2 Анализ матриц сходства

Для оценки близости текстовых документов были использованы три метрики: косинусное сходство, метрика Охай и евклидово расстояние. Результаты были визуализированы в виде тепловых карт до и после отбора признаков.

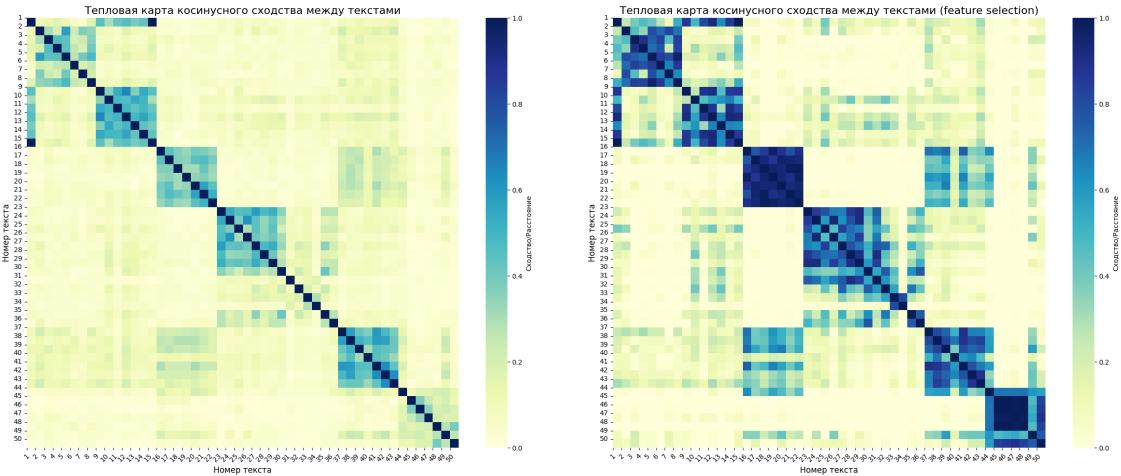


Рисунок 4.1 — Тепловые карты косинусного сходства: до отбора признаков (слева) и после (справа)

Анализ тепловых карт косинусного сходства (рис. 4.1) показывает, что после отбора признаков на карте появилось больше шума. Однако, некоторые классы (3, 6, 7) стали более четко выраженным, что может свидетельствовать о лучшем разделении этих тематических групп. В то же время, классы 1-4 и 5 начали сливатся, что может указывать на потерю некоторой дифференцирующей информации при отборе признаков.

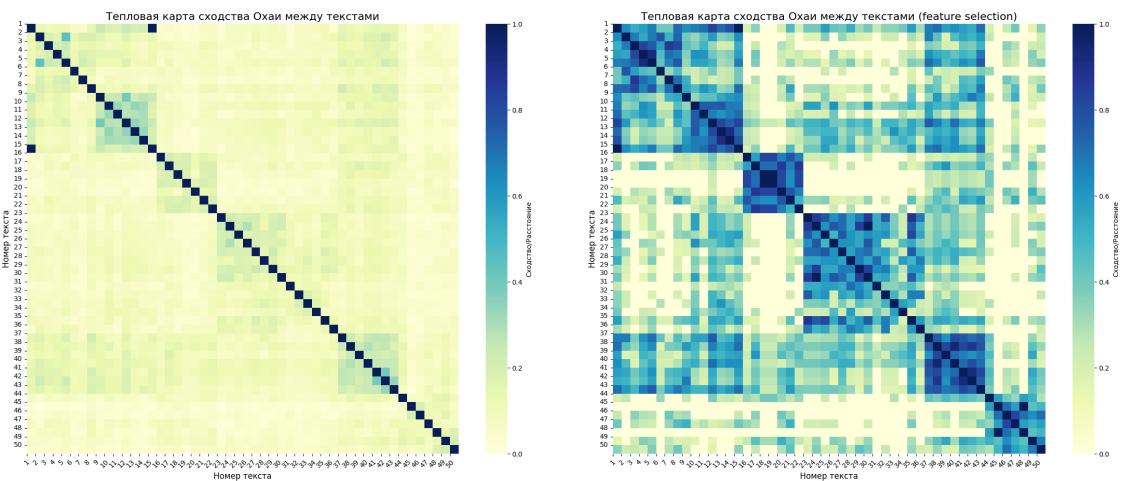


Рисунок 4.2 — Тепловые карты сходства Охай: до отбора признаков (слева) и после (справа)

Метрика Охай (рис. 4.2) после отбора признаков показала значительное увеличение шума на тепловой карте. Несмотря на это, общая структура сходства между документами сохраняется и в целом согласуется с результатами косинусной меры. Это может свидетельствовать о том, что метрика Охай более чувствительна к изменению количества признаков.

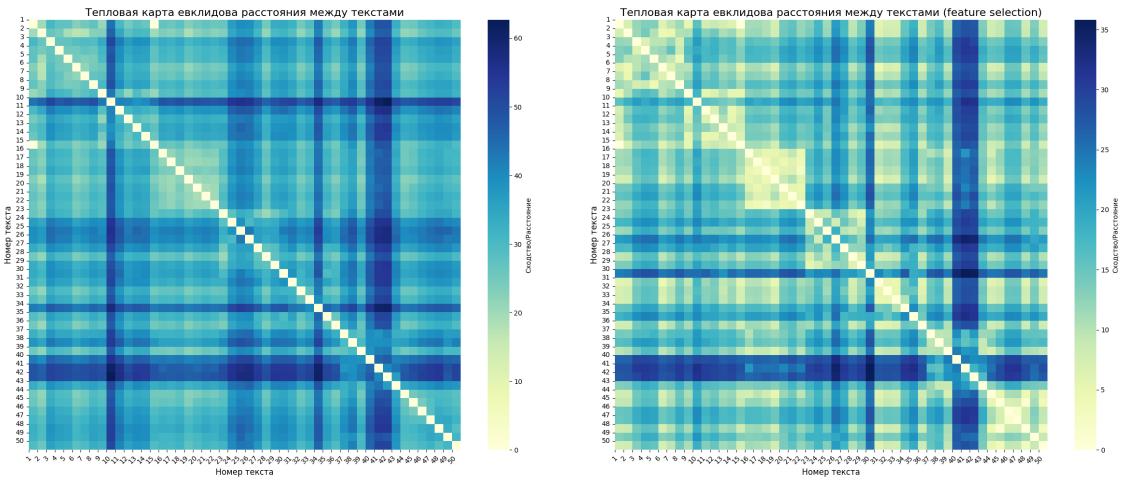


Рисунок 4.3 — Тепловые карты евклидова расстояния: до отбора признаков (слева) и после (справа)

Евклидово расстояние (рис. 4.3) после отбора признаков показало улучшение в визуализации тематических групп. Темы текстов стали более четко различимыми, что может свидетельствовать о том, что отобранные признаки хорошо отражают ключевые различия между документами. Однако, по общей эффективности в выявлении структуры данных евклидово расстояние все еще уступает косинусной мере.

4.3 Анализ топ-слов по группам и метрикам

Анализ топ-слов по группам и метрикам, представленный в Приложении А, показал интересные результаты. Пересечений множеств слов по группам внутри метрик практически не наблюдается, что свидетельствует о высокой степени различия между словами в разных тематических группах. Это подтверждает эффективность выбранного метода векторизации и отбора признаков в сохранении уникальных характеристик каждой группы текстов.

Важно отметить, что даже в случаях, когда у двух групп встречаются общие слова, их частота употребления существенно отличается. Это указывает на то, что не только наличие определенных слов, но и их относительная важность в контексте группы играет ключевую роль в характеристике текстов.

Отсутствие в топ-списках служебных слов (союзов, предлогов и т.д.) подтверждает эффективность предварительной обработки текста и фокусирование на содержательных словах, несущих основную смысловую нагрузку.

4.4 Заключение и направления дальнейших исследований

Проведенное исследование показало эффективность применения методов отбора признаков для анализа текстовых данных. Снижение размерности с 3630 до 19 признаков позволило сохранить ключевую информацию о тематической структуре текстов, что подтверждается анализом тепловых карт различных метрик сходства.

Для дальнейшего исследования предлагается:

Провести анализ влияния количества отобранных признаков на качество классификации текстов. Это позволит определить оптимальное соотношение между размерностью данных и сохранением информативности.

Исследовать эффективность других методов снижения размерности, таких как латентно-семантический анализ (LSA) или t-SNE, и сравнить их результаты с полученными в данном исследовании.

Применить методы кластерного анализа к полученным векторным представлениям текстов для автоматического выявления тематических групп и сравнения с исходной классификацией.

Провести более детальный анализ слов, оставшихся после отбора признаков, для выявления их роли в определении тематики текстов и потенциального использования в задачах автоматической классификации.

4.5 Вывод

В ходе исследования был проведен комплексный анализ характеристик набора текстовых данных с использованием методов отбора признаков и различных метрик сходства. Применение метода отбора признаков позволило значительно снизить размерность данных с 3630 до 19 признаков, сохранив при этом ключевую информацию о тематической структуре текстов. Анализ тепловых карт различных метрик сходства (косинусное сходство, метрика Охай, евклидово расстояние) до и после отбора признаков показал, что, несмотря на некоторое увеличение шума, основная структура сходства между документами сохраняется. Особенно эффективным оказалось применение косинусной меры сходства, которая даже после значительного сокращения числа признаков продолжает четко выделять основные тематические группы текстов. Анализ топ-слов по группам и метрикам подтвердил высокую степень различия между словами в разных тематических группах, что свидетельствует об эффективности выбранного подхода к векторизации и отбору признаков. Полученные результаты открывают перспективы для дальнейших исследований в области автоматической классификации текстов и оптимизации методов снижения размерности данных.

ЗАКЛЮЧЕНИЕ

В рамках данной лабораторной работы было проведено комплексное исследование характеристик набора текстовых данных с целью определения значимости различных параметров векторного пространства и оптимизации процессов обработки и анализа текстовой информации.

Основные результаты и выводы исследования:

- 1) **Статистический анализ:** Применение методов оценки дисперсии, математического ожидания, ковариации и корреляций позволило выявить ключевые характеристики векторного представления документов как для всего набора, так и для отдельных тематических групп.
- 2) **Снижение размерности:** Использование метода отбора признаков на основе дисперсии и корреляции позволило сократить размерность данных с 3630 до 19 признаков без критической потери информативности.
- 3) **Анализ метрик сходства:** Исследование косинусного сходства, метрики Охай и евклидова расстояния выявило их различную чувствительность к изменению количества признаков. Косинусная мера показала наибольшую устойчивость.
- 4) **Тематическая дифференциация:** Анализ топ-слов подтвердил высокую степень различия между словами в разных тематических группах, что свидетельствует об эффективности выбранного подхода к векторизации и отбору признаков.
- 5) **Оптимизация обработки данных:** Выявленная возможность существенного сокращения размерности данных открывает перспективы для оптимизации процессов обработки больших объемов текстовой информации.
- 6) **Практическая применимость:** Результаты имеют широкое применение в областях информационного поиска, автоматической классификации текстов и тематического моделирования.

Направления дальнейших исследований включают:

- Изучение влияния различных методов предобработки текста на качество векторного представления.
- Исследование эффективности других методов снижения размерности (t-SNE, UMAP).
- Разработку адаптивных методов отбора признаков, учитывающих специфику конкретных тематических областей.
- Интеграцию полученных результатов с методами глубокого обучения.

В заключение стоит отметить, что проведенное исследование не только достигло поставленной цели по определению значимости различных параметров векторного пространства, но и предоставило ценные инсайты о структуре и характеристиках текстовых данных. Полученные результаты демонстрируют возможность существенной оптимизации процессов обработки текстовой информации, что имеет важное значение в контексте современных задач анализа больших данных и искусственного интеллекта.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. scikit-learn: PCA [Электронный ресурс] // Scikit-learn.org. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html> (дата обращения: 07.10.2024).
2. NumPy Documentation [Электронный ресурс] // NumPy.org. URL: <https://numpy.org/doc/stable/index.html> (дата обращения: 07.10.2024).
3. Pandas Documentation [Электронный ресурс] // Pandas.pydata.org. URL: <https://pandas.pydata.org/docs/> (дата обращения: 07.10.2024).

Приложение А

Сравнение слов различными метриками в пределах одной группы

Сравнение топ-50 слов между метриками для группы 'Анамнез':

Общие слова для всех метрик: {'приступ', 'врач', 'должный', 'деталь'}

Топ слова для метрики variance: {'событие', 'наследственность', 'семейный', 'объективный', 'болезнь', 'влияние', 'родственник', 'ребёнок', 'свидетель', 'выяснить', 'тяжёлый', 'случай', 'должный', 'деталь', 'эпилепсия', 'история', 'опрос', 'медицинский', 'вопрос', 'кровь', 'беременность', 'анамнез', 'приступ', 'пациент', 'наличие', 'психиатрический', 'риск', 'эпизод', 'платформа', 'цель', 'далее', 'больной', 'сведение', 'программный', 'эпилептический', 'вена', 'необходимо', 'ранее', 'заболевание', 'анализ', 'врач', 'жалоба', 'дать', 'симптом', 'сбор', 'проведение', 'качество', 'оценка', 'иметься', 'зabor'}

Топ слова для метрики mean: {'событие', 'наследственность', 'семейный', 'болезнь', 'лицо', 'развитие', 'родственник', 'ребёнок', 'выяснить', 'информация', 'свой', 'случай', 'должный', 'позволять', 'деталь', 'эпилепсия', 'диагностика', 'данные', 'опрос', 'медицинский', 'вопрос', 'важно', 'кровь', 'беременность', 'уровень', 'анамнез', 'приступ', 'пациент', 'наличие', 'метод', 'являться', 'риск', 'осмотр', 'помогать', 'особенность', 'цель', 'далее', 'больной', 'сведение', 'реакция', 'получить', 'необходимо', 'заболевание', 'анализ', 'врач', 'дать', 'сбор', 'качество', 'оценка', 'иметься'}

Топ слова для метрики covariance: {'событие', 'влияние', 'родственник', 'проблема', 'ребёнок', 'свидетель', 'информация', 'задавать', 'уточнить', 'должный', 'деталь', 'эпилепсия', 'позволять', 'большой', 'опрос', 'сторона', 'вопрос', 'важно', 'следовать', 'доктор', 'степень', 'анамнез', 'рассказ', 'приступ', 'пациент', 'прежде', 'целесообразно', 'эпизод', 'сообщить', 'помогать', ' позвонить', 'далее', 'цель', 'возможность', 'особенность', 'эпилептический', 'получить', 'трактовка', 'необходимо', 'высказаться', 'заболевание', 'информированность', 'врач', 'привести', 'дать', 'относиться', 'порой', 'сбор', 'разный', 'помощь'}

Топ слова для метрики pearson_corr: {'собеседование', 'намного', 'встреча', 'привязать', 'понятный', 'рассказчик', 'деталь', 'должный', 'отойти', 'уста', 'максимально', 'взаимодополняющий', 'предмет', 'полностью', 'шаблон', 'конец', 'восстановление', 'повторный', 'спешить', 'беспечность', 'незначительный', 'конечность', 'прийти', 'скорый', 'опросить', 'кабинет', 'приступ', 'культура', 'уточнение', 'эффективность', 'далее', 'рефлексия', 'возможность', 'воспользоваться', 'упущение', 'сосредоточить', 'эмоциональный', 'эпилептический', 'случайно', 'появление', 'частично', 'произойти', 'врач', 'интересовать', 'объективность', 'коллега', 'порой', 'относиться', 'разный', 'информационный'}

Топ слова для метрики spearman_corr: {'собеседование', 'картина', 'путь', 'идеал', 'вспомнить', 'припадок', 'рассказчик', 'деталь', 'должный', 'происходить', 'закладываться', 'при-

ступать', 'шаблон', 'полностью', 'взаимодополняющий', 'спешить', ' позвать', 'развернуть', 'синдром', 'доктор', 'приступ', 'культура', 'уточнение', 'дверь', 'ставить', 'эффективность', 'тонус', 'отклониться', 'парциальный', 'попросить', 'возможность', 'сосед', 'рефлексия', 'чужой', 'предусмотреть', 'воспользоваться', 'эпилептический', 'трактовка', 'продумать', 'случайно', 'комплексный', 'врач', 'объективность', 'коллега', 'порой', 'относиться', 'старший', 'размышление', 'уставновленный', 'информационный'}

Сравнение топ-50 слов между метриками для группы 'Биохимия крови':

Общие слова для всех метрик: {'общий', 'содержание', 'печень', 'холестерин', 'уровень', 'белок'}

Топ слова для метрики variance: {'различный', 'билирубин', 'плазма', 'организм', 'проблема', 'фермент', 'образование', 'происходить', 'общий', 'указывать', 'печень', 'патология', 'холестерин', 'кровь', 'биохимия', 'уровень', 'количество', 'глобулин', 'результат', 'биохимический', 'повышение', 'орган', 'ммольть', 'снижение', 'фракция', 'являться', 'липид', 'альбумин', 'например', 'повышенный', 'норма', 'белок', 'выявить', 'наблюдаться', 'концентрация', 'заболевание', 'анализ', 'клетка', 'содержание', 'колебаться', 'исследование', 'ткань', 'говорить', 'показатель', 'вещество', 'процесс', 'работа', 'нарушение', 'травма', 'зabor'}

Топ слова для метрики mean: {'различный', 'билирубин', 'плазма', 'состояние', 'свидетельствовать', 'организм', 'фермент', 'проблема', 'сахарный', 'образование', 'позволять', 'общий', 'изменение', 'диагностика', 'признак', 'печень', 'патология', 'кровь', 'холестерин', 'биохимия', 'обмен', 'уровень', 'гепатит', 'количество', 'результат', 'глобулин', 'биохимический', 'повышение', 'орган', 'ммольть', 'снижение', 'почка', 'являться', 'фракция', 'высокий', 'альбумин', 'мочевина', 'норма', 'белок', 'заболевание', 'анализ', 'содержание', 'исследование', 'распад', 'глюкоза', 'показатель', 'вещество', 'процесс', 'работа', 'нарушение'}

Топ слова для метрики covariance: {'поэтому', 'билирубин', 'характеризовать', 'плазма', 'болезнь', 'организм', 'проблема', 'фермент', 'непрямой', 'причина', 'аста', 'образование', 'прямой', 'происходить', 'общий', 'рвота', 'признак', 'следствие', 'метаболический', 'лпвп', 'печень', 'продукт', 'недостаточный', 'лпнп', 'холестерин', 'уровень', 'количество', 'глобулин', 'повышение', 'пища', 'снижение', 'фракция', 'являться', 'почка', 'альбумин', 'мочевина', ' повышенный', 'белок', 'наблюдаться', 'приём', 'клетка', 'содержание', 'распад', 'либо', 'глюкоза', 'эритроцит', 'вещество', 'процесс', 'работа', 'нарушение'}

Топ слова для метрики pearson_corr: {'поэтому', 'снижаться', 'характеризовать', 'плазма', 'отношение', 'аста', 'прогностический', 'нерасторимый', 'происходить', 'общий', 'половой', 'инсулин', 'рвота', 'животное', 'признак', 'сразу', 'лпвп', 'печень', 'недостаточный', 'зло-качественный', 'способность', 'холестерин', 'совокупность', 'мускулатура', 'уровень', 'физический', 'исключение', 'пища', 'сродство', 'снижение', 'помимо', 'являться', 'участвовать', 'тонус', 'диуретик', 'мышца', 'обезвоживание', 'клеточный', 'повышенный', 'служащий', 'белок', 'физиологически', 'наблюдаться', 'содержание', 'клетка', 'переизбыток', 'глюкоза', 'потребление', 'активность', 'утилизация'}

Топ слова для метрики spearman_corr: {'возможный', 'дегидратация', 'нагрузка', 'ку-

рение', 'прочее', 'аста', 'сдаваться', 'вследствие', 'обеспечивать', 'значительно', 'оптимальный', 'общий', 'сразу', 'животное', 'лпвп', 'онкотический', 'печень', 'продукт', 'постоянно', 'относительно', 'лпнп', 'холестерин', 'гормон', 'судорога', 'уровень', 'физический', 'наоборот', 'гепатит', 'дозировка', 'циркуляция', 'пища', 'сродство', 'рост', 'мочевыделительный', 'бляшка', 'любой', 'диуретик', 'становиться', 'серёзный', 'окраска', 'белок', 'вместе', 'содержание', 'переизбыток', 'определенный', 'сознание', 'глюкоза', 'активность', 'глюкуронов', 'направить'}

Сравнение топ-50 слов между метриками для группы 'Комментаторы':

Общие слова для всех метрик: {'риторический', 'средство'}

Топ слова для метрики variance: {'событие', 'зритель', 'плавание', 'риторический', 'игрок', 'свой', 'дистанция', 'должный', 'происходить', 'спортсмен', 'репортаж', 'вести', 'быстрый', 'большой', 'интерес', 'среди', 'спорт', 'комментатор', 'сеть', 'журналист', 'активно', 'следовать', 'подготовить', 'социальный', 'количество', 'панкрат', 'соревнование', 'являться', 'комментировать', 'взаимодействие', 'собственный', 'хороший', 'комментарий', 'денис', 'речь', 'возможность', 'спортивный', 'например', 'личный', 'жанр', 'современный', 'личность', 'использование', 'чемпион', 'метафора', 'определенный', 'средство', 'спонтанный', 'языковой', 'работа'}

Топ слова для метрики mean: {'событие', 'зритель', 'знание', 'матч', 'плавание', 'аудитория', 'риторический', 'игрок', 'свой', 'дистанция', 'должный', 'происходить', 'спортсмен', 'репортаж', 'вести', 'комментирование', 'большой', 'спорт', 'комментатор', 'журналист', 'следовать', 'подготовить', 'количество', 'соревнование', 'разный', 'разбираться', 'являться', 'комментировать', 'собственный', 'стиль', 'комментарий', 'денис', 'речь', 'возможность', 'спортивный', 'например', 'жанр', 'современный', 'язык', 'личность', 'статья', 'говорить', 'определенный', 'метафора', 'средство', 'использовать', 'понимать', 'спонтанный', 'языковой', 'работа'}

Топ слова для метрики covariance: {'олимпийский', 'различный', ' отметить', 'зритель', 'плавание', 'россия', 'риторический', 'европа', 'дистанция', 'эмоциональность', 'водный', 'данные', 'комментирование', 'комментаторпрофессионал', 'среди', 'спорт', 'тонкость', 'важно', 'комментатор', 'рассказ', 'количество', 'панкрат', 'чемпионат', 'соревнование', 'тропа', 'ведущий', 'являться', 'комментировать', 'фигура', 'индивидуальный', 'собственный', 'проводиться', 'начаться', 'восклицание', 'денис', 'характеристика', 'больший', 'самый', 'современный', 'олимпиада', 'половина', 'личность', 'формировать', 'медаль', 'многие', 'чемпион', 'метафора', 'средство', 'языковой', 'комментаторский'}

Топ слова для метрики pearson_corr: {'олимпийский', ' отметить', 'несоблюдаемый', 'отождествлять', 'алексей', 'профессионализм', 'ретранслятор', 'заимствование', 'малышев', 'принцип', 'риторический', 'прослушивание', 'начать', 'мастерски', 'честь', 'граница', 'изображение', 'изменение', 'регулярный', 'данные', 'требоваться', 'эмоция', 'активный', 'защита', 'установка', 'существенно', 'подкреплять', 'внести', 'профессиональный', 'рассказ', 'чемпионат', 'соревнование', 'комментаторрепортёр', 'предъявляться', 'комментаторзнатка', 'стиль', 'восклицание', 'диалогизация', 'распространить', 'самый', 'комментаторироника', 'чемпионка', 'формировать', 'средство', 'эстетический', 'ирония', 'процесс', 'комментаторский', 'попов', 'демонстрировать'}

Топ слова для метрики spearman_corr: {'олимпийский', ' отметить', 'отождествлять', 'алексей', 'ретранслятор', 'взрастать', 'малышев', 'принцип', 'риторический', 'воздействовать', 'мастерски', 'сионим', 'граница', 'образование', 'изображение', 'градация', 'честь', 'изменение', 'ведение', 'зависеть', 'требоваться', 'эмоция', 'активный', 'защита', 'учёт', 'установка', 'тонкость', 'требование', 'мастерство', 'ольга', 'внести', 'профессиональный', 'рассказ', 'сохранять', 'чемпионат', 'восприятие', 'расширить', 'комментаторактёр', 'стиль', 'восхищание', 'диализация', 'атлетика', 'распространить', 'самый', 'комментаторионика', 'средство', 'эстетический', 'процесс', 'комментаторский', 'демонстрировать'}

Сравнение топ-50 слов между метриками для группы 'Неорганическая химия':

Общие слова для всех метрик: {'изучить', 'основа'}

Топ слова для метрики variance: {'различный', 'понятие', 'теоретический', 'французский', 'развитие', 'открытие', 'химик', 'неорганический', 'масса', 'образование', 'научный', 'период', 'изучить', 'история', 'свойство', 'кислота', 'соль', 'физик', 'теория', 'окислительновосстановительный', 'основа', 'синтез', 'открыть', 'закон', 'физический', 'химический', 'химия', 'метод', 'раствор', 'периодический', 'являться', 'физикохимический', 'учёный', 'система', 'металл', 'элемент', 'английский', 'реакция', 'например', 'атомный', 'получить', 'минеральный', 'анализ', 'новый', 'исследование', 'стать', 'комплексообразование', 'соединение', 'вещество', 'связь'}

Топ слова для метрики mean: {'различный', 'понятие', 'теоретический', 'развитие', 'материал', 'открытие', 'химик', 'неорганический', 'масса', 'образование', 'научный', 'период', 'изучить', 'история', 'свойство', 'атом', 'соль', 'теория', 'основа', 'синтез', 'открыть', 'закон', 'физический', 'химический', 'структура', 'химия', 'метод', 'периодический', 'раствор', 'являться', 'природа', 'учёный', 'система', 'металл', 'элемент', 'реакция', 'например', 'атомный', 'получить', 'минеральный', 'современный', 'анализ', 'новый', 'исследование', 'стать', 'использование', 'соединение', 'вещество', 'процесс', 'связь'}

Топ слова для метрики covariance: {'разработка', 'направление', 'платиновый', 'понятие', 'влияние', 'развитие', 'французский', 'открытие', 'химик', 'предложить', 'масса', 'установить', 'научный', 'конец', 'изучить', 'история', 'свойство', 'соль', 'теория', 'изучение', 'разработать', 'основа', 'строение', 'открыть', 'закон', 'солевой', 'химический', 'супрамолекулярный', 'практический', 'раствор', 'метод', 'физикохимический', 'комплекс', 'система', 'металл', 'элемент', 'друг', 'электропроводность', 'атомный', 'получить', 'лиганд', 'электронный', 'анализ', 'новый', 'исследование', 'дать', 'координационный', 'комплексообразование', 'соединение', 'работа'}

Топ слова для метрики pearson_corr: {'направление', 'платиновый', 'электрохимический', 'оборудование', 'предложить', 'химикатехнолог', 'ндчичиров', 'установить', 'научный', 'конец', 'лгберг', 'сельский', 'изучить', 'совершенствование', 'многокомпонентный', 'справочник', 'соль', 'ииия', 'изучение', 'методика', 'афгерасим', 'разработать', 'основа', 'строение', 'гпдезидерьев', 'практический', 'супрамолекулярный', 'изготовление', 'раствор', 'создать', 'расмотреть', 'лмамиров', 'система', 'фармацевтический', 'возможность', 'электропроводность', 'неизвестный', 'высококачественный', 'лантаноид', 'растворимость', 'дать', 'комплексообразо-

вание', 'ансамбль', 'гидролитический', 'гетеровалентный', 'рутений', 'состав', 'работа', 'стекло', 'нанесение'}

Топ слова для метрики spearman_corr: {'измерение', 'направление', 'тгахмет', 'платиновый', 'сдгромаков', 'электрохимический', 'осмийиридийплатина', 'сложный', 'сольватация', 'металлалиганд', 'жидкокристаллический', 'предложить', 'активность', 'кимочалов', 'ндчицииров', 'установить', 'конец', 'лгберг', 'гомый', 'изучить', 'сельский', 'многокомпонентный', 'изучение', 'основа', 'устойчивость', 'строительство', 'ивовчинник', 'авзахар', 'действие', 'полимерный', 'практический', 'создать', 'протонный', 'электролиз', 'вфторопов', 'неизвестный', 'вестись', 'овмихайл', 'армустафина', 'вывести', 'дать', 'экспериментальный', 'ансамбль', 'гидролитический', 'диссоциация', 'гетероядерный', 'генератор', 'оценка', 'работа', 'стекло'}

Сравнение топ-50 слов между метриками для группы 'Органическая химия':

Общие слова для всех метрик: {'степень'}

Топ слова для метрики variance: {'нефть', 'горение', 'радикал', 'этilen', 'развитие', 'детонационный', 'материал', 'углеродный', 'природный', 'детонация', 'углерод', 'период', 'смесь', 'конец', 'спирт', 'воздух', 'бензин', 'атом', 'цепь', 'продукт', 'пропан', 'гидропероксид', 'топливо', 'алкан', 'получение', 'сгорание', 'синтез', 'название', 'углеводород', 'степень', 'область', 'органический', 'молекула', 'химический', 'химия', 'правило', 'наука', 'двигатель', 'цилиндр', 'изооктан', 'большой', 'давление', 'получить', 'сжатие', 'число', 'образоваться', 'попутный', 'состав', 'вещество', 'октановый'}

Топ слова для метрики mean: {'нефть', 'горение', 'радикал', 'этilen', 'качество', 'развитие', 'материал', 'производство', 'углеродный', 'природный', 'детонация', 'углерод', 'конец', 'смесь', 'период', 'спирт', 'бензин', 'атом', 'большой', 'продукт', 'цепь', 'топливо', 'получение', 'алкан', 'гидропероксид', 'сгорание', 'синтез', 'углеводород', 'название', 'степень', 'молекула', 'органический', 'химический', 'область', 'химия', 'метод', 'являться', 'правило', 'наука', 'двигатель', 'путём', 'получить', 'сжатие', 'число', 'соединение', 'образоваться', 'состав', 'вещество', 'октановый', 'связь'}

Топ слова для метрики covariance: {'горение', 'бензиновый', 'детонационный', 'температура', 'камера', 'поршень', 'детонация', 'резкий', 'сжиматься', 'смесь', 'воздух', 'нгептан', 'бензин', 'опасность', 'большой', 'цепь', 'принять', 'спокойный', 'определять', 'выше', 'гидропероксид', 'топливо', 'скорость', 'объём', 'углеводород', 'степень', 'уменьшаться', 'фронт', 'возникать', 'повышение', 'сгорать', 'двигатель', 'повышаться', 'характеристика', 'цилиндр', 'изооктан', 'большой', 'давление', 'маленький', 'автомобильный', 'легко', 'мощность', 'содержание', 'сжатие', 'число', 'образоваться', 'увеличить', 'шкала', 'процесс', 'октановый'}

Топ слова для метрики pearson_corr: {'подшипник', 'горение', 'бензиновый', 'электрический', 'процент', 'температура', 'вспомнить', 'детонация', 'сжиматься', 'неустойчивый', 'сильный', 'обладать', 'нгептан', 'сравнивать', 'равный', 'работать', 'достигаться', 'гидропероксид', 'скорость', 'расход', 'устойчивость', 'неравномерный', 'степень', 'уменьшаться', 'взрыво-подобный', 'предел', 'шестичленный', 'сгорать', 'экономичность', 'дымный', 'характеристика', 'цилиндр', 'возможность', 'рабочий', 'изооктан', 'случайно', 'мощность', 'содержание', 'кольцо', 'стекло'}

'сжатие', 'пара', 'говорить', 'стук', 'увеличение', 'тесно', 'коленчатый', 'показатель', 'задолго', 'алкен', 'октановый'}

Топ слова для метрики spearman_corr: {'непредельный', 'дело', 'двигаться', 'электрический', 'детонационный', 'температура', 'доля', 'испытуемый', 'достигать', 'развиваться', 'обладать', 'неустойчивый', 'сжиматься', 'сорт', 'окисление', 'работать', 'достигаться', 'триметилпентан', 'расход', 'устойчивость', 'неравномерный', 'степень', 'взрывоподобный', 'уменьшаться', 'предел', 'шестичленный', 'дымный', 'мера', 'экономичность', 'характеристика', 'возможность', 'рабочий', 'кислородсодержащий', 'изооктан', 'ухудшаться', 'сжимать', 'кольцо', 'пара', 'говорить', 'стук', 'карбюраторный', 'увеличение', 'тесно', 'коленчатый', 'эфир', 'детонировать', 'распространение', 'алкен', 'задолго', 'сжать'}

Сравнение топ-50 слов между метриками для группы 'Спортивная медицина':

Общие слова для всех метрик: {'восстановление', 'специальный', 'тренер', 'средство', 'контроль'}

Топ слова для метрики variance: {'состояние', 'нагрузка', 'лицо', 'специальный', 'развитие', 'организм', 'гимнастика', 'работоспособность', 'обследование', 'контроль', 'здравье', 'должный', 'функциональный', 'спортсмен', 'питание', 'диагностика', 'восстановление', 'спорт', 'раздел', 'рациональный', 'медицина', 'физический', 'профессиональный', 'тело', 'пациент', 'культура', 'орган', 'метод', 'занятие', 'лечебный', 'тренер', 'воспитание', 'система', 'задача', 'тренировочный', 'мышца', 'цель', 'спортивный', 'курс', 'заниматься', 'комплексный', 'врач', 'наблюдение', 'лечение', 'средство', 'проведение', 'работка', 'травма', 'упражнение', 'врачебный'}

Топ слова для метрики mean: {'различный', 'состояние', 'нагрузка', 'специальный', 'организм', 'развитие', 'обследование', 'контроль', 'здравье', 'свой', 'должный', 'функциональный', 'спортсмен', 'питание', 'восстановление', 'спорт', 'медицинский', 'раздел', 'медицина', 'физический', 'тренировка', 'профессиональный', 'определение', 'уровень', 'тело', 'результат', 'культура', 'соревнование', 'метод', 'занятие', 'режим', 'лечебный', 'тренер', 'являться', 'система', 'задача', 'цель', 'спортивный', 'заниматься', 'заболевание', 'врач', 'наблюдение', 'лечение', 'средство', 'использовать', 'процесс', 'работка', 'травма', 'упражнение', 'врачебный'}

Топ слова для метрики covariance: {'различный', 'специальный', 'развитие', 'организм', 'россия', 'работоспособность', 'гимнастика', 'контроль', 'форма', 'упражнение', 'массовый', 'функциональный', 'период', 'общий', 'диагностика', 'контингент', 'восстановление', 'спорт', 'изучение', 'раздел', 'рациональный', 'физический', 'медицина', 'следовать', 'культура', 'повышение', 'метод', 'занятие', 'лечебный', 'тренер', 'являться', 'воспитание', 'система', 'задача', 'спортивный', 'курс', 'заниматься', 'содержание', 'заболевание', 'комплексный', 'использование', 'лечение', 'средство', 'использовать', 'разный', 'процесс', 'работка', 'клинический', 'связь', 'врачебный'}

Топ слова для метрики pearson_corr: {'самостоятельный', 'зависимость', 'сравнение', 'составной', 'специальный', 'россия', 'стимулятор', 'гимнастика', 'контроль', 'форма', 'целое', 'период', 'общий', 'обострение', 'основный', 'восстановление', 'активный', 'продукт', 'работать', 'врачебный'}

'существенный', 'касаться', 'специфика', 'поддержание', 'часть', 'контролировать', 'следовать', 'человеческий', 'связать', 'действие', 'режим', 'тренер', 'являться', 'повреждение', 'наука', 'хороший', 'советский', 'повышенный', 'значение', 'посвятить', 'заболевание', 'использование', 'иной', 'средство', 'использовать', 'относиться', 'жизненный', 'трудоспособность', 'физкультура', 'клинический', 'настоящий'}

Топ слова для метрики spearman_corr: {'самостоятельный', 'зависимость', 'сравнение', 'составной', 'специальный', 'россия', 'стимулятор', 'гимнастика', 'психологический', 'контроль', 'форма', 'целое', 'период', 'общий', 'обострение', 'основный', 'восстановление', 'активный', 'продукт', 'работать', 'существенный', 'касаться', 'поддержание', 'специфика', 'часть', 'контролировать', 'человеческий', 'связать', 'действие', 'тренер', 'являться', 'повреждение', 'наука', 'хороший', 'советский', 'повышенный', 'значение', 'посвятить', 'заболевание', 'использование', 'иной', 'средство', 'использовать', 'относиться', 'трудоспособность', 'жизненный', 'качество', 'физкультура', 'клинический', 'настоящий'}

Сравнение топ-50 слов между метриками для группы 'Ставки на спорт':

Общие слова для всех метрик: {'контора', 'свой', 'игрок'}

Топ слова для метрики variance: {'событие', 'стратегия', 'контора', 'доход', 'азартный', 'лошадь', 'математический', 'свой', 'должный', 'изменение', 'грайнд', 'некоторый', 'счёт', 'маржа', 'спорт', 'лошадиный', 'букмекерский', 'цупис', 'прибыль', 'закон', 'ставка', 'профессиональный', 'карта', 'ситуация', 'беттинг', 'забег', 'система', 'спорный', 'налог', 'пари', 'спортивный', 'регистрация', 'ожидание', 'скачка', 'букмекер', 'азарт', 'федерация', 'отчисление', 'поддержка', 'оскар', 'коэффициент', 'вероятность', 'игра', 'рубль', 'делать', 'федеральный', 'формула', 'сумма', 'игрок', 'организация'}

Топ слова для метрики mean: {'событие', 'стратегия', 'размер', 'контора', 'выигрыш', 'азартный', 'лошадь', 'математический', 'деятельность', 'свой', 'должный', 'некоторый', 'грайнд', 'счёт', 'маржа', 'спорт', 'букмекерский', 'цупис', 'прибыль', 'нужный', 'закон', 'ставка', 'профессиональный', 'карта', 'ситуация', 'расчёт', 'беттинг', 'забег', 'система', 'спорный', 'налог', 'пари', 'спортивный', 'российский', 'поддержка', 'букмекер', 'скачка', 'федерация', 'регистрация', 'азарт', 'оскар', 'коэффициент', 'вероятность', 'игра', 'рубль', 'делать', 'федеральный', 'сумма', 'игрок', 'организация'}

Топ слова для метрики covariance: {'ошибка', 'возможный', 'блокировка', 'контора', 'технический', 'проблема', 'причина', 'деятельность', 'обращение', 'свой', 'случай', 'должный', 'некоторый', 'ресурс', 'легальный', 'спорцбоокий', 'большинство', 'служба', 'переписка', 'букмекерский', 'закон', 'профессиональный', 'лимит', 'действие', 'клиент', 'ставить', 'ситуация', 'любой', 'правило', 'ревиэ', 'спорный', 'налог', 'помогать', 'аккаунт', 'российский', 'спортивный', 'регистрация', 'букмекер', 'поддержка', 'федерация', 'польза', 'жалоба', 'рубль', 'репутация', 'федеральный', 'ассоциация', 'игрок', 'неправильный', 'решить', 'организация'}

Топ слова для метрики pearson_corr: {'анды', 'ошибка', 'поэтому', 'возможный', 'блокировка', 'оставлять', 'паспортный', 'контора', 'бизнес', 'сомнительный', 'обращение', 'свой', 'данный', 'ресурс', 'решение', 'полный', 'позволять', 'уменьшить', 'легальный', 'спорцбоокий', 'спорт', 'игрок', 'неправильный', 'решить', 'организация'}

'плюсовой', 'большинство', 'лимит', 'действие', 'пригодиться', 'ставить', 'ситуация', 'выплачивать', 'беттинг', 'специализированный', 'любой', 'целенаправленный', 'правильно', 'хороший', 'помогать', 'спорный', 'английский', 'аккаунт', 'сделать', 'грешить', 'условие', 'язык', 'продолжать', 'выигрышный', 'жалоба', 'стабильный', 'игорный', 'игрок', 'редакция', 'решить'}

Топ слова для метрики spearman_corr: {'оставлять', 'поэтому', 'блокировка', 'паспортный', 'контора', 'отказ', 'технический', 'проблема', 'обратиться', 'понижение', 'бизнес', 'сомнительный', 'ассоциатион', 'обращение', 'наказывать', 'свой', 'данный', 'полный', 'оставить', 'легальный', 'авторитетный', 'спорцбоокий', 'указывать', 'правота', 'договорный', 'расставаться', 'наверняка', 'действие', 'смириться', 'ставить', 'неточность', 'выплачивать', 'беттинг', 'специализированный', 'любой', 'целенаправленный', 'критика', 'помогать', 'английский', 'российский', 'удаться', 'грешить', 'выигрышный', 'стабильный', 'репутация', 'игорный', 'будущее', 'игрок', 'редакция', 'организация'}

Сравнение слов с "наилучшими" метриками а каждой группе

Сравнение топ-50 слов по метрике variance между группами:

Общие слова для всех групп: —

Топ слова для группы Анамнез: {'событие', 'наследственность', 'семейный', 'объективный', 'болезнь', 'влияние', 'родственник', 'ребёнок', 'свидетель', 'выяснить', 'тяжёлый', 'случай', 'должный', 'деталь', 'эпилепсия', 'история', 'опрос', 'медицинский', 'вопрос', 'кровь', 'беременность', 'анамнез', 'приступ', 'пациент', 'наличие', 'психиатрический', 'риск', 'эпизод', 'платформа', 'цель', 'далее', 'больной', 'сведение', 'программный', 'эпилептический', 'вена', 'необходимо', 'ранее', 'заболевание', 'анализ', 'врач', 'жалоба', 'дать', 'симптом', 'сбор', 'проведение', 'качество', 'оценка', 'иметься', 'зabor'}

Топ слова для группы Биохимия крови: {'различный', 'билирубин', 'плазма', 'организм', 'проблема', 'фермент', 'образование', 'происходить', 'общий', 'указывать', 'печень', 'патология', 'холестерин', 'кровь', 'биохимия', 'уровень', 'количества', 'глобулин', 'результат', 'биохимический', 'повышение', 'орган', 'ммоль', 'снижение', 'фракция', 'являться', 'липид', 'альбумин', 'например', 'повышенный', 'норма', 'белок', 'выявить', 'наблюдаться', 'концентрация', 'заболевание', 'анализ', 'клетка', 'содержание', 'колебаться', 'исследование', 'ткань', 'говорить', 'показатель', 'вещество', 'процесс', 'работа', 'нарушение', 'травма', 'зabor'}

Топ слова для группы Комментаторы: {'событие', 'зритель', 'плавание', 'риторический', 'игрок', 'свой', 'дистанция', 'должный', 'происходить', 'спортсмен', 'репортаж', 'вести', 'быстрый', 'большой', 'интерес', 'среди', 'спорт', 'комментатор', 'сеть', 'журналист', 'активно', 'следовать', 'подготовить', 'социальный', 'количества', 'панкрат', 'соревнование', 'являться', 'комментировать', 'взаимодействие', 'собственный', 'хороший', 'комментарий', 'денис', 'речь', 'возможность', 'спортивный', 'например', 'личный', 'жанр', 'современный', 'личность', 'использование', 'чемпион', 'метафора', 'определенный', 'средство', 'спонтанный', 'языковой', 'работа'}

Топ слова для группы Неорганическая химия: {'различный', 'понятие', 'теоретический',

'французский', 'развитие', 'открытие', 'химик', 'неорганический', 'масса', 'образование', 'научный', 'период', 'изучить', 'история', 'свойство', 'кислота', 'соль', 'физик', 'теория', 'окислительновосстановительный', 'основа', 'синтез', 'открыть', 'закон', 'физический', 'химический', 'химия', 'метод', 'раствор', 'периодический', 'являться', 'физикохимический', 'учёный', 'система', 'металл', 'элемент', 'английский', 'реакция', 'например', 'атомный', 'получить', 'минеральный', 'анализ', 'новый', 'исследование', 'стать', 'комплексообразование', 'соединение', 'вещество', 'связь'}

Топ слова для группы Органическая химия: {'нефть', 'горение', 'радикал', 'этilen', 'развитие', 'детонационный', 'материал', 'углеродный', 'природный', 'детонация', 'углерод', 'период', 'смесь', 'конец', 'спирт', 'воздух', 'бензин', 'атом', 'цепь', 'продукт', 'пропан', 'гидропероксид', 'топливо', 'алкан', 'получение', 'сгорание', 'синтез', 'название', 'углеводород', 'степень', 'область', 'органический', 'молекула', 'химический', 'химия', 'правило', 'наука', 'двигатель', 'цилиндр', 'изооктан', 'большой', 'давление', 'получить', 'сжатие', 'число', 'образоваться', 'попутный', 'состав', 'вещество', 'октановый'}

Топ слова для группы Спортивная медицина: {'состояние', 'нагрузка', 'лицо', 'специальный', 'развитие', 'организм', 'гимнастика', 'работоспособность', 'обследование', 'контроль', 'здравье', 'должный', 'функциональный', 'спортсмен', 'питание', 'диагностика', 'восстановление', 'спорт', 'раздел', 'рациональный', 'медицина', 'физический', 'профессиональный', 'тело', 'пациент', 'культура', 'орган', 'метод', ' занятие', 'лечебный', 'тренер', 'воспитание', 'система', 'задача', 'тренировочный', 'мышца', 'цель', 'спортивный', 'курс', 'заниматься', 'комплексный', 'врач', 'наблюдение', 'лечение', 'средство', 'проведение', 'работка', 'травма', 'упражнение', 'врачебный'}

Топ слова для группы Ставки на спорт: {'событие', 'стратегия', 'контора', 'доход', 'азартный', 'лошадь', 'математический', 'свой', 'должный', 'изменение', 'грайнд', 'некоторый', 'счёт', 'маржа', 'спорт', 'лошадиный', 'букмекерский', 'цупис', 'прибыль', 'закон', 'ставка', 'профессиональный', 'карта', 'ситуация', 'беттинг', 'забег', 'система', 'спорный', 'налог', 'пари', 'спортивный', 'регистрация', 'ожидание', 'скачка', 'букмекер', 'азарт', 'федерация', 'отчисление', 'поддержка', 'оскар', 'коэффициент', 'вероятность', 'игра', 'рубль', 'делать', 'федеральный', 'формула', 'сумма', 'игрок', 'организация'}

Сравнение топ-50 слов по метрике mean между группами:

Общие слова для всех групп: —

Топ слова для группы Анамнез: {'событие', 'наследственность', 'семейный', 'болезнь', 'лицо', 'развитие', 'родственник', 'ребёнок', 'выяснить', 'информация', 'свой', 'случай', 'должный', 'позволять', 'деталь', 'эпилепсия', 'диагностика', 'данные', 'опрос', 'медицинский', 'вопрос', 'важно', 'кровь', 'беременность', 'уровень', 'анамнез', 'приступ', 'пациент', 'наличие', 'метод', 'являться', 'риск', 'осмотр', 'помогать', 'особенность', 'цель', 'далее', 'больной', 'сведение', 'реакция', 'получить', 'необходимо', 'заболевание', 'анализ', 'врач', 'дать', 'сбор', 'качество', 'оценка', 'иметься'}

Топ слова для группы Биохимия крови: {'различный', 'билирубин', 'плазма', 'состоя-

ние', 'свидетельствовать', 'организм', 'фермент', 'проблема', 'сахарный', 'образование', 'позволять', 'общий', 'изменение', 'диагностика', 'признак', 'печень', 'патология', 'кровь', 'холестерин', 'биохимия', 'обмен', 'уровень', 'гепатит', 'количество', 'результат', 'глобулин', 'биохимический', 'повышение', 'орган', 'ммольть', 'снижение', 'почка', 'являться', 'фракция', 'высокий', 'альбумин', 'мочевина', 'норма', 'белок', 'заболевание', 'анализ', 'содержание', 'исследование', 'распад', 'глюкоза', 'показатель', 'вещество', 'процесс', 'работа', 'нарушение'}

Топ слова для группы Комментаторы: { 'событие', 'зритель', 'знание', 'матч', 'плавание', 'аудитория', 'риторический', 'игрок', 'свой', 'дистанция', 'должный', 'происходить', 'спортсмен', 'репортаж', 'вести', 'комментирование', 'большой', 'спорт', 'комментатор', 'журналист', 'следовать', 'подготовить', 'количество', 'соревнование', 'разный', 'разбираться', 'являться', 'комментировать', 'собственный', 'стиль', 'комментарий', 'денис', 'речь', 'возможность', 'спортивный', 'например', 'жанр', 'современный', 'язык', 'личность', 'статья', 'говорить', 'определенный', 'метафора', 'средство', 'использовать', 'понимать', 'спонтанный', 'языковой', 'работа' }

Топ слова для группы Неорганическая химия: { 'различный', 'понятие', 'теоретический', 'развитие', 'материал', 'открытие', 'химик', 'неорганический', 'масса', 'образование', 'научный', 'период', 'изучить', 'история', 'свойство', 'атом', 'соль', 'теория', 'основа', 'синтез', 'открыть', 'закон', 'физический', 'химический', 'структура', 'химия', 'метод', 'периодический', 'раствор', 'являться', 'природа', 'учёный', 'система', 'металл', 'элемент', 'реакция', 'например', 'атомный', 'получить', 'минеральный', 'современный', 'анализ', 'новый', 'исследование', 'стать', 'использование', 'соединение', 'вещество', 'процесс', 'связь' }

Топ слова для группы Органическая химия: { 'нефть', 'горение', 'радикал', 'этilen', 'качество', 'развитие', 'материал', 'производство', 'углеродный', 'природный', 'детонация', 'углерод', 'конец', 'смесь', 'период', 'спирт', 'бензин', 'атом', 'большой', 'продукт', 'цепь', 'топливо', 'получение', 'алкан', 'гидропероксид', 'сгорание', 'синтез', 'углеводород', 'название', 'степень', 'молекула', 'органический', 'химический', 'область', 'химия', 'метод', 'являться', 'правило', 'наука', 'двигатель', 'путём', 'получить', 'сжатие', 'число', 'соединение', 'образоваться', 'состав', 'вещество', 'октановый', 'связь' }

Топ слова для группы Спортивная медицина: { 'различный', 'состояние', 'нагрузка', 'специальный', 'организм', 'развитие', 'обследование', 'контроль', 'здравье', 'свой', 'должный', 'функциональный', 'спортсмен', 'питание', 'восстановление', 'спорт', 'медицинский', 'раздел', 'медицина', 'физический', 'тренировка', 'профессиональный', 'определение', 'уровень', 'тело', 'результат', 'культура', 'соревнование', 'метод', ' занятие', 'режим', 'лечебный', 'тренер', 'являться', 'система', 'задача', 'цель', 'спортивный', 'заниматься', 'заболевание', 'врач', 'наблюдение', 'лечение', 'средство', 'использовать', 'процесс', 'работа', 'травма', 'упражнение', 'врачебный' }

Топ слова для группы Ставки на спорт: { 'событие', 'стратегия', 'размер', 'контора', 'выигрыш', 'азартный', 'лошадь', 'математический', 'деятельность', 'свой', 'должный', 'некоторый', 'грайнд', 'счёт', 'маржа', 'спорт', 'букмекерский', 'цупис', 'прибыль', 'нужный', 'закон', 'ставка', 'профессиональный', 'карта', 'ситуация', 'расчёт', 'беттинг', 'забег', 'система', 'спорный', 'на-

лог', 'пари', 'спортивный', 'российский', 'поддержка', 'букмекер', 'скачка', 'федерация', 'регистрация', 'азарт', 'оскар', 'коэффициент', 'вероятность', 'игра', 'рубль', 'делать', 'федеральный', 'сумма', 'игрок', 'организация'}

Сравнение топ-50 слов по метрике covariance между группами:

Общие слова для всех групп: —

Топ слова для группы Анамнез: { 'событие', 'влияние', 'родственник', 'проблема', 'ребёнок', 'свидетель', 'информация', 'задавать', 'уточнить', 'должный', 'деталь', 'эпилепсия', 'позволять', 'большой', 'опрос', 'сторона', 'вопрос', 'важно', 'следовать', 'доктор', 'степень', 'анамнез', 'рассказ', 'приступ', 'пациент', 'прежде', 'целесообразно', 'эпизод', 'сообщить', 'помогать', 'позвонить', 'далее', 'цель', 'возможность', 'особенность', 'эпилептический', 'получить', 'трактовка', 'необходимо', 'высказаться', 'заболевание', 'информированность', 'врач', 'привести', 'дать', 'относиться', 'порой', 'сбор', 'разный', 'помощь' }

Топ слова для группы Биохимия крови: { 'поэтому', 'билирубин', 'характеризовать', 'плазма', 'болезнь', 'организм', 'проблема', 'фермент', 'непрямой', 'причина', 'аста', 'образование', 'прямой', 'происходить', 'общий', 'рвота', 'признак', 'следствие', 'метаболический', 'лпвп', 'печень', 'продукт', 'недостаточный', 'лппп', 'холестерин', 'уровень', 'количество', 'глобулин', 'повышение', 'пища', 'снижение', 'фракция', 'являться', 'почка', 'альбумин', 'мочевина', ' повышенный', 'белок', 'наблюдаться', 'приём', 'клетка', 'содержание', 'распад', 'либо', 'глюкоза', 'эритроцит', 'вещество', 'процесс', 'работа', 'нарушение' }

Топ слова для группы Комментаторы: { 'олимпийский', 'различный', 'отметить', 'зритель', 'плавание', 'россия', 'риторический', 'европа', 'дистанция', 'эмоциональность', 'водный', 'данные', 'комментирование', 'комментаторпрофессионал', 'среди', 'спорт', 'тонкость', 'важно', 'комментатор', 'рассказ', 'количество', 'панкрат', 'чемпионат', 'соревнование', 'тропа', 'ведущий', 'являться', 'комментировать', 'фигура', 'индивидуальный', 'собственный', 'проводиться', 'начаться', 'восклицание', 'денис', 'характеристика', 'больший', 'самый', 'современный', 'олимпиада', 'половина', 'личность', 'формировать', 'медаль', 'многие', 'чемпион', 'метафора', 'средство', 'языковой', 'комментаторский' }

Топ слова для группы Неорганическая химия: { 'разработка', 'направление', 'платиновый', 'понятие', 'влияние', 'развитие', 'французский', 'открытие', 'химик', 'предложить', 'масса', 'установить', 'научный', 'конец', 'изучить', 'история', 'свойство', 'соль', 'теория', 'изучение', 'разработать', 'основа', 'строительство', 'открыть', 'закон', 'солевой', 'химический', 'супрамолекулярный', 'практический', 'раствор', 'метод', 'физикохимический', 'комплекс', 'система', 'металл', 'элемент', 'друг', 'электропроводность', 'атомный', 'получить', 'лиганд', 'электронный', 'анализ', 'новый', 'исследование', 'дать', 'координационный', 'комплексообразование', 'соединение', 'работа' }

Топ слова для группы Органическая химия: { 'горение', 'бензиновый', 'детонационный', 'температура', 'камера', 'поршень', 'детонация', 'резкий', 'сжиматься', 'смесь', 'воздух', 'нгептан', 'бензин', 'опасность', 'большой', 'цепь', 'принять', 'спокойный', 'определять', 'выше', 'гидропероксид', 'топливо', 'скорость', 'объём', 'углеводород', 'степень', 'уменьшаться', 'фронт', 'реакция' }

'возникать', 'повышение', 'сгорать', 'двигатель', 'повышаться', 'характеристика', 'цилиндр', 'изооктан', 'больший', 'давление', 'маленький', 'автомобильный', 'легко', 'мощность', 'содержание', 'сжатие', 'число', 'образоваться', 'увеличить', 'шкала', 'процесс', 'октановый'}

Топ слова для группы Спортивная медицина: {'различный', 'специальный', 'развитие', 'организм', 'россия', 'работоспособность', 'гимнастика', 'контроль', 'форма', 'упражнение', 'массовый', 'функциональный', 'период', 'общий', 'диагностика', 'контингент', 'восстановление', 'спорт', 'изучение', 'раздел', 'рациональный', 'физический', 'медицина', 'следовать', 'культура', 'повышение', 'метод', 'занятие', 'лечебный', 'тренер', 'являться', 'воспитание', 'система', 'задача', 'спортивный', 'курс', 'заниматься', 'содержание', 'заболевание', 'комплексный', 'использование', 'лечение', 'средство', 'использовать', 'разный', 'процесс', 'работа', 'клинический', 'связь', 'врачебный'}

Топ слова для группы Ставки на спорт: {'ошибка', 'возможный', 'блокировка', 'контора', 'технический', 'проблема', 'причина', 'деятельность', 'обращение', 'свой', 'случай', 'должный', 'некоторый', 'ресурс', 'легальный', 'спорцбоокий', 'большинство', 'служба', 'переписка', 'букмекерский', 'закон', 'профессиональный', 'лимит', 'действие', 'клиент', 'ставить', 'ситуация', 'любой', 'правило', 'ревиэ', 'спорный', 'налог', 'помогать', 'аккаунт', 'российский', 'спортивный', 'регистрация', 'букмекер', 'поддержка', 'федерация', 'польза', 'жалоба', 'рубль', 'репутация', 'федеральный', 'ассоциация', 'игрок', 'неправильный', 'решить', 'организация'}

Сравнение топ-50 слов по метрике pearson_corr между группами:

Общие слова для всех групп: —

Топ слова для группы Анамнез: {'собеседование', 'намного', 'встреча', 'привязать', 'понятный', 'рассказчик', 'деталь', 'должный', 'отойти', 'уста', 'максимально', 'взаимодополняющий', 'предмет', 'полностью', 'шаблон', 'конец', 'восстановление', 'повторный', 'спешить', 'беспечность', 'незначительный', 'конечность', 'прийти', 'скорый', 'опросить', 'кабинет', 'приступ', 'культура', 'уточнение', 'эффективность', 'далее', 'рефлексия', 'возможность', 'воспользоваться', 'упущение', 'сосредоточить', 'эмоциональный', 'эпилептический', 'случайно', 'появление', 'частично', 'произойти', 'врач', 'интересовать', 'объективность', 'коллега', 'порой', 'относиться', 'разный', 'информационный'}

Топ слова для группы Биохимия крови: {'поэтому', 'снижаться', 'характеризовать', 'плазма', 'отношение', 'аста', 'прогностический', 'нерасторимый', 'происходить', 'общий', 'половой', 'инсулин', 'рвота', 'животное', 'признак', 'сразу', 'лпвп', 'печень', 'недостаточный', 'злокачественный', 'способность', 'холестерин', 'совокупность', 'мускулатура', 'уровень', 'физический', 'исключение', 'пища', 'сродство', 'снижение', 'помимо', 'являться', 'участвовать', 'тонус', 'диуретик', 'мышца', 'обезвоживание', 'клеточный', ' повышенный', 'служащий', 'белок', 'физиологически', 'наблюдаться', 'содержание', 'клетка', 'переизбыток', 'глюкоза', 'потребление', 'активность', 'утилизация'}

Топ слова для группы Комментаторы: {'олимпийский', ' отметить', 'несоблюдаемый', 'отождествлять', 'алексей', 'профессионализм', 'ретранслятор', 'заимствование', 'мальшев', 'принцип', 'риторический', 'прослушивание', 'начать', 'мастерски', 'честь', 'граница', 'изоб-

ражение', 'изменение', 'регулярный', 'данные', 'требоваться', 'эмоция', 'активный', 'защита', 'установка', 'существенно', 'подкреплять', 'внести', 'профессиональный', 'рассказ', 'чемпионат', 'соревнование', 'комментаторрепортёр', 'предъявляться', 'комментаторзнатка', 'стиль', 'восхищание', 'диалогизация', 'распространить', 'самый', 'комментаториороника', 'чемпионка', 'формировать', 'средство', 'эстетический', 'ирония', 'процесс', 'комментаторский', 'попов', 'демонстрировать'}

Топ слова для группы Неорганическая химия: { 'направление', 'платиновый', 'электрохимический', 'оборудование', 'предложить', 'химикатехнолог', 'ндчичиров', 'установить', 'научный', 'конец', 'лгберг', 'сельский', 'изучить', 'совершенствование', 'многокомпонентный', 'справочник', 'соль', 'ииия', 'изучение', 'методика', 'афгерасим', 'разработать', 'основа', 'строение', 'тпдезидерьев', 'практический', 'супрамолекулярный', 'изготовление', 'раствор', 'создать', 'расмотреть', 'лмамиров', 'система', 'фармацевтический', 'возможность', 'электропроводность', 'неизвестный', 'высококачественный', 'лантаноид', 'растворимость', 'дать', 'комплексообразование', 'ансамбль', 'гидролитический', 'гетеровалентный', 'рутений', 'состав', 'работа', 'стекло', 'нанесение' }

Топ слова для группы Органическая химия: { 'подшипник', 'горение', 'бензиновый', 'электрический', 'процент', 'температура', 'вспомнить', 'детонация', 'сжиматься', 'неустойчивый', 'сильный', 'обладать', 'нгептан', 'сравнивать', 'равный', 'работать', 'достигаться', 'гидропероксид', 'скорость', 'расход', 'устойчивость', 'неравномерный', 'степень', 'уменьшаться', 'взрывоподобный', 'предел', 'шестичленный', 'сгорать', 'экономичность', 'дымный', 'характеристика', 'цилиндр', 'возможность', 'рабочий', 'изооктан', 'случайно', 'мощность', 'содержание', 'кольцо', 'сжатие', 'пара', 'говорить', 'стук', 'увеличение', 'тесно', 'коленчатый', 'показатель', 'задолго', 'алкен', 'октановый' }

Топ слова для группы Спортивная медицина: { 'самостоятельный', 'зависимость', 'сравнение', 'составной', 'специальный', 'россия', 'стимулятор', 'гимнастика', 'контроль', 'форма', 'целое', 'период', 'общий', 'обострение', 'основный', 'восстановление', 'активный', 'продукт', 'работать', 'существенный', 'касаться', 'специфика', 'поддержание', 'часть', 'контролировать', 'следовать', 'человеческий', 'связать', 'действие', 'режим', 'тренер', 'являться', 'повреждение', 'наука', 'хороший', 'советский', ' повышенный', 'значение', 'посвятить', 'заболевание', 'использование', 'иной', 'средство', 'использовать', 'относиться', 'жизненный', 'трудоспособность', 'физкультура', 'клинический', 'настоящий' }

Топ слова для группы Ставки на спорт: { 'анды', 'ошибка', 'поэтому', 'возможный', 'блокировка', 'оставлять', 'паспортный', 'контора', 'бизнес', 'сомнительный', 'обращение', 'свой', 'данный', 'ресурс', 'решение', 'полный', 'позволять', 'уменьшить', 'легальный', 'спорцбоокий', 'плюсовой', 'большинство', 'лимит', 'действие', 'пригодиться', 'ставить', 'ситуация', 'выплачивать', 'беттинг', 'специализированный', 'любой', 'целенаправленный', 'правильно', 'хороший', 'помогать', 'спорный', 'английский', 'аккаунт', 'сделать', 'грешить', 'условие', 'язык', 'продолжать', 'выигрышный', 'жалоба', 'стабильный', 'игорный', 'игрок', 'редакция', 'решить' }

Сравнение топ-50 слов по метрике spearman_corr между группами:

Общие слова для всех групп: —

Топ слова для группы Анамнез: { 'собеседование', 'картина', 'путь', 'идеал', 'вспомнить', 'припадок', 'рассказчик', 'деталь', 'должный', 'происходить', 'закладываться', 'приступать', 'шаблон', 'полностью', 'взаимодополняющий', 'спешить', ' позвать', 'развернуть', 'синдром', 'доктор', 'приступ', 'культура', 'уточнение', 'дверь', 'ставить', 'эффективность', 'тонус', 'отклониться', 'парциальный', 'попросить', 'возможность', 'сосед', 'рефлексия', 'чужой', 'предусмотреть', 'воспользоваться', 'эпилептический', 'трактовка', 'продумать', 'случайно', 'комплексный', 'врач', 'объективность', 'коллега', 'порой', 'относиться', 'старший', 'размышление', 'уставновленный', 'информационный' }

Топ слова для группы Биохимия крови: { 'возможный', 'дегидратация', 'нагрузка', 'куриение', 'прочее', 'аста', 'сдаваться', 'вследствие', 'обеспечивать', 'значительно', 'оптимальный', 'общий', 'сразу', 'животное', 'лпвп', 'онкотический', 'печень', 'продукт', 'постоянно', 'относительно', 'лпнп', 'холестерин', 'гормон', 'судорога', 'уровень', 'физический', 'наоборот', 'гепатит', 'дозировка', 'циркуляция', 'пища', 'сродство', 'рост', 'мочевыделительный', 'бляшка', 'любой', 'диуретик', 'становиться', 'серёзный', 'окраска', 'белок', 'вместе', 'содержание', 'переизбыток', 'определенный', 'сознание', 'глюкоза', 'активность', 'глюкуронов', 'направить' }

Топ слова для группы Комментаторы: { 'олимпийский', ' отметить', 'отождествлять', 'алексей', 'ретранслятор', 'возрастать', 'малышев', 'принцип', 'риторический', 'воздействовать', 'мастерски', 'синоним', 'граница', 'образование', 'изображение', 'градация', 'честь', 'изменение', 'ведение', 'зависеть', 'требоваться', 'эмоция', 'активный', 'защита', 'учёт', 'установка', 'тонкость', 'требование', 'мастерство', 'ольга', 'внести', 'профессиональный', 'рассказ', 'сохранять', 'чемпионат', 'восприятие', 'расширить', 'комментаторактёр', 'стиль', 'восхищение', 'диализация', 'атлетика', 'распространить', 'самый', 'комментаторироника', 'средство', 'эстетический', 'процесс', 'комментаторский', 'демонстрировать' }

Топ слова для группы Неорганическая химия: { 'измерение', 'направление', 'тгахмет', 'платиновый', 'сдгромаков', 'электрохимический', 'осмийиридийплатина', 'сложный', 'сольватация', 'металлалиганд', 'жидкокристаллический', 'предложить', 'активность', 'кимочалов', 'ндчичиров', 'установить', 'конец', 'лгберг', 'гомый', 'изучить', 'сельский', 'многокомпонентный', 'изучение', 'основа', 'устойчивость', 'строение', 'ивовчинник', 'авзахар', 'действие', 'полимерный', 'практический', 'создать', 'протонный', 'электролиз', 'вфторопов', 'неизвестный', 'вестись', 'овмихайл', 'армустафина', 'вывести', 'дать', 'экспериментальный', 'ансамбль', 'гидролитический', 'диссоциация', 'гетероядерный', 'генератор', 'оценка', 'работа', 'стекло' }

Топ слова для группы Органическая химия: { 'непредельный', 'дело', 'двигаться', 'электрический', 'детонационный', 'температура', 'доля', 'испытуемый', 'достигать', 'развиваться', 'обладать', 'неустойчивый', 'сжиматься', 'сорт', 'окисление', 'работать', 'достигаться', 'триметилпентан', 'расход', 'устойчивость', 'неравномерный', 'степень', 'взрывоподобный', 'уменьшаться', 'предел', 'шестичленный', 'дымный', 'мера', 'экономичность', 'характеристика', 'возможность', 'рабочий', 'кислородсодержащий', 'изооктан', 'ухудшаться', 'сжимать', 'кольцо', 'пара', 'говорить', 'стук', 'карбюраторный', 'увеличение', 'тесно', 'коленчатый', 'эфир', 'дето-

нировать', 'распространение', 'алкен', 'задолго', 'сжать'}

Топ слова для группы Спортивная медицина: {'самостоятельный', 'зависимость', 'сравнение', 'составной', 'специальный', 'россия', 'стимулятор', 'гимнастика', 'психологический', 'контроль', 'форма', 'целое', 'период', 'общий', 'обострение', 'основный', 'восстановление', 'активный', 'продукт', 'работать', 'существенный', 'касаться', 'поддержание', 'специфика', 'часть', 'контролировать', 'человеческий', 'связать', 'действие', 'тренер', 'являться', 'повреждение', 'написать', 'хороший', 'советский', 'повышенный', 'значение', 'посвятить', 'заболевание', 'использование', 'иной', 'средство', 'использовать', 'относиться', 'трудоспособность', 'жизненный', 'качество', 'физкультура', 'клинический', 'настоящий'}

Топ слова для группы Ставки на спорт: {'оставлять', 'поэтому', 'блокировка', 'паспортный', 'контора', 'отказ', 'технический', 'проблема', 'обратиться', 'понижение', 'бизнес', 'сомнительный', 'ассоциативный', 'обращение', 'наказывать', 'свой', 'данный', 'полный', 'оставить', 'легальный', 'авторитетный', 'спорцбоокий', 'указывать', 'правота', 'договорный', 'расставаться', 'наверняка', 'действие', 'смириться', 'ставить', 'неточность', 'выплачивать', 'беттинг', 'специализированный', 'любой', 'целенаправленный', 'критика', 'помогать', 'английский', 'российский', 'удаться', 'грешить', 'выигрышный', 'стабильный', 'репутация', 'игорный', 'будущее', 'игрок', 'редакция', 'организация'}