



Д.Е. Прокудин, О.В. Кононова, Г.С. Левит

**Применение методов и технологий поиска, экспликации и анализа контекстного знания в науковедческих исследованиях (на примере научного наследия Г.Ф. Гаузе)**

***Рекомендуемая форма библиографической ссылки***

Прокудин Д.Е., Кононова О.В., Левит Г.С. Применение методов и технологий поиска, экспликации и анализа контекстного знания в науковедческих исследованиях (на примере научного наследия Г.Ф. Гаузе) // Научный сервис в сети Интернет: труды XXIII Всероссийской научной конференции (20-23 сентября 2021 г., онлайн). — М.: ИПМ им. М.В.Келдыша, 2021. — С. 322-338.

<https://doi.org/10.20948/abrau-2021-25>

<https://keldysh.ru/abrau/2021/theses/25.pdf>

***Видеозапись выступления***

# **Применение методов и технологий поиска, экспликации и анализа контекстного знания в науковедческих исследованиях (на примере научного наследия Г.Ф. Гаузе)**

**Д.Е. Прокудин<sup>1,2,4</sup>, О.В. Кононова<sup>2,4</sup>, Г.С. Левит<sup>3</sup>**

<sup>1</sup> *Санкт-Петербургский государственный университет*

<sup>2</sup> *Университет ИТМО*

<sup>3</sup> *Йенский университет имени Фридриха Шиллера*

<sup>4</sup> *Центр исследований цифрового общества, Россия*

**Аннотация.** Целью настоящего исследования является изучение применения методов поиска, экспликации и анализа текстовых данных научных публикаций с использованием информационно-коммуникационных технологий для применения в науковедческих исследованиях. Исследование проведено на русскоязычных научных публикациях, отражающих научное наследие Г.Ф. Гаузе. Предлагаемое исследование основано на результатах проведённых тематических исследований по оценке возможностей применения цифровых информационных ресурсов в науковедческих исследованиях, извлечения метаданных из цифровых электронных ресурсов с применением методов последующей количественной их обработки. В исследовании были изучены методы экспликации и анализа текстовых данных, извлечённых из цифровых научных ресурсов (на примере НЭБ, Elibrary). Для анализа использовалась информационная система Sketch Engine, предоставляющая инструменты обработки текстов на естественном языке (NLP). На основе анализа полученных результатов сделаны выводы о возможности использования изученных методов не только в науковедческих исследованиях, но и в широком круге научных исследований различной тематики.

**Ключевые слова:** научная информация, контекстное знание, поиск, экспликация, анализ данных, синтетический метод, Digital Humanities, Sketch Engine, дигитализация научного наследия, Г.Ф. Гаузе

# Methods and Technologies of Search, Explication and Analysis of Contextual Knowledge for Scientific Heritage Studies: A Case Study of Georgy F. Gause

D.E. Prokudin<sup>1,2</sup>, O.V. Kononova<sup>2,4</sup>, G.S. Levit<sup>3</sup>

<sup>1</sup> *St. Petersburg State University, Russia*

<sup>2</sup> *ITMO University, Russia*

<sup>3</sup> *Jena University, Germany*

<sup>4</sup> *Center digital society research, Russia*

**Abstract.** The objective of this research is to study methods of search, explication and analysis of text data of scientific publications with information and communication technologies for use in scientific research. The research is based on Russian-language scientific publications reflecting the scientific heritage of G. F. Gause. The proposed study is based on the results of case studies conducted to assess the possibilities of using digital information resources in scientific research, extracting metadata from digital electronic resources using methods of their subsequent quantitative processing. The study examined the methods of explication and analysis of text data extracted from digital scientific resources (for example, Elibrary). For the analysis, the information system Sketch Engine was used, which provides natural language text processing (NLP) tools. Based on the analysis of the obtained results, conclusions are drawn about the possibility of using the studied methods not only in scientific research, but also in a wide range of scientific research on various topics.

**Keywords:** scientific information, contextual knowledge, search, explication, data analysis, synthetic method, Digital Humanities, Sketch Engine, digitalization of scientific heritage, G.F. Gause

## 1. Введение

Данное исследование является очередным этапом комплексного исследования, ставящего перед собой цель изучения эффективного использования инструментария Digital Humanities в науковедческих исследованиях. В нём используются результаты проведённых ранее тематических исследований по оценке возможностей применения цифровых информационных ресурсов в науковедческих исследованиях [14, 20], извлечения метаданных из цифровых электронных ресурсов и применения методов последующей количественной их обработки [21], а также изучения методов визуализации геоданных, извлечённых из метаданных научных публикаций для применения в науковедческих исследованиях [15]. Актуальность исследования определяется возрастающим значением использования информационно-коммуникационных технологий в научных исследованиях, а также тем, что

научная информация в настоящее время в подавляющем большинстве доступна в электронной форме. А количественное увеличение не позволяет эффективно использовать её в научных исследованиях традиционными методами. В качестве фактологической базы исследования было выбрано влияние идей выдающегося советского биолога Г.Ф. Гаузе на развитие науки [3]. Это обусловлено рядом причин. Прежде всего он был выдающимся советским учёным сталинской и пост-сталинской эпохи. При этом не был подвержен репрессиям и его карьера развивалась непрерывно. Научный авторитет Гаузе простирался далеко за пределы СССР. Помимо этого, его работы опубликованы на английском языке в ведущих западных издательствах и журналах. Его основной научный труд «Борьба за существование» был впервые опубликован на английском языке в США [18]. В связи с этими фактами биографии Гаузе правомочен вопрос о влиянии учёного, жившего в тоталитарном изолированном обществе на развитие науки как внутри этого общества, так и за его пределами. Ещё одним важным обстоятельством в пользу выбора научного наследия Гаузе является то, что он известен сразу в нескольких научных областях. Наиболее известен во всём мире его вклад в эволюционную теорию и экологию. Гаузе был первым, кто экспериментально исследовал борьбу за существование. Выведенный им закон, известный как «принцип конкурентного исключения», доказывает, что никакие два вида, конкурирующие за ограниченный ресурс, не могут сосуществовать в рамках единой экологической ниши. В СССР Гаузе был выдающимся исследователем в области медицины и участвовал в разработке антибиотиков. В годы Великой Отечественной Войны он разработал первый советский антибиотик Gramacidin S (то есть Soviet – советский). Также Гаузе были проведены исследования асимметрии протоплазмы, которую он рассматривал в аспекте биологической эволюции.

С точки зрения науковедения смысл проводимого исследования заключается в сравнении влияния Гаузе в СССР, современной России и за рубежом с целью нахождения закономерностей, которые позднее могли бы быть объединены традиционными средствами истории науки. А исследуемые нами методы Digital Humanities позволяют как производить эффективный поиск и отбор научных публикаций, так и осуществлять их анализ с использованием соответствующего инструментария.

Целью данного исследования является изучение эффективности применения методов поиска, экспликации и анализа текстовых данных научных публикаций с использованием информационно-коммуникационных технологий в исследованиях науковедческих исследования на примере научного наследия Г.Ф. Гаузе.

## **2. Методы поиска, экспликации и анализа контекстного знания с использованием информационно-коммуникационных технологий в научных исследованиях**

В научных исследованиях применяются различные методы обработки текстовых данных, например, контент-анализ [10], различные виды тематического моделирования [9], обработка и анализ библиометрической информации [17, 22] и др. Также одними из наиболее доступных и часто применяемых являются методы обработки текстов на естественном языке, которые традиционно востребованы в лингвистических исследованиях [1, 12]. Применение этих методов основано на использовании соответствующего программного обеспечения и информационных систем. В целом, методы анализа текстов могут применяться в исследованиях по различным предметным областям, например, в исторических исследованиях, да и в изучении гуманитарного знания в общем. При этом использование данных методов предполагает поиск и выделение контекстов, экспликацию из них контекстного знания с последующим его анализом. На комплексное применение этих методов направлен разработанный подход, названный «синтетическим методом» [5, 6]. Данный подход основывается на применении информационно-коммуникационных технологий на всех этапах проведения исследования по анализу контекстного знания – от выбора цифровых электронных ресурсов и поиска научных текстов до экспликации контекстного знания и его анализа. Синтетический метод использовался в науковедческих исследованиях формирования терминологической базы и развития тематик развивающихся междисциплинарных направлений научных исследований [6, 7, 8], показав при этом высокую эффективность. Этот же метод предлагается использовать для более эффективного проведения науковедческих исследований по истории науки. Для обоснования этого предположения синтетический метод применён при проведении тематического исследования научного наследия Г.Ф. Гаузе.

## **3. Поиск и экспликация научных публикаций из цифровых научных ресурсов**

В предыдущих исследованиях было выявлено, что для поиска и отбора научных публикаций из русскоязычного научного дискурса наилучшим электронным ресурсом является Научная электронная библиотека (НЭБ, <https://www.elibrary.ru>). Это связано как с тем, что в ней наиболее полно представлены отечественные научные публикации не только из научных журналов, но и из монографий, материалов конференций, сериальных изданий, сборников научных статей и пр., так и с тем, что публикации сопровождаются подробными метаданными, включающими в себя аннотации, списки ключевых слов, тематические направления научных исследований и т.п. Важным обстоятельством при

выборе НЭБ в качестве источника научных публикаций явилось наличие поискового механизма с мощными возможностями. При этом применение многочисленных фильтров и поисковых условий позволяет решать достаточно сложные задачи отбора научных публикаций по заданным критериям.

Основными недостатками НЭБ являются: 1) для загрузки метаданных публикаций и разнообразной библиометрической информации предоставляется интерфейс программирования API, но только на платной основе [4]; 2) отсутствует возможность пакетной загрузки текстов отобранных публикаций. Поэтому в найденные и отобранные в рамках данного исследования публикации пришлось скачивать вручную, что не является эффективным способом.

Для поиска научных публикаций были использованы следующие запросы: «Гаузе & (принцип конкурентного исключения)»; «Гаузе & антибиотики»; «Гаузе & Лотка & Вольтерра». Последний запрос отражает тот факт, что Г.Ф. Гаузе экспериментально проверил положения Лотки-Вольтерра, которые независимо друг от друга разработали математическую модель взаимодействия двух видов типа «хищник — жертва» [19, 23]. Поэтому часто результаты их исследований обсуждаются в научной литературе. По этим запросам были сформированы подборки публикаций. Затем все три подборки были объединены в результирующую, так как в них предполагалось получить дублирующие публикации, что и было подтверждено (таб. 1).

Таб. 1. Статистика публикаций и полных текстов, найденных по соответствующим запросам в НЭБ (<https://www.elibrary.ru>).

<b>Подборка публикаций</b>	<b>Число публикаций</b>	<b>Число доступных текстов публикаций</b>
Гаузе & (принцип конкурентного исключения)	720	120
Гаузе & антибиотики	17	10
Гаузе & Лотка & Вольтерра	5	5
Результирующая подборка	736	132

После этого были скачаны 131 текст в формате PDF. Одна ссылка на текст оказалась «битой». Затем пришлось в ПО Adobe Acrobat Pro распознать 5 текстов публикаций, которые были представлены без текстового слоя. Из этих текстов был сформирован корпус.

#### 4. Анализ текстовых данных научных публикаций с применением информационно-коммуникационных технологий

Для дальнейшего анализа полученных научных публикаций был выбран инструмент Sketch Engine (<https://www.sketchengine.eu>), который представляет собой веб-ориентированную систему обработки текстов на естественном языке с функциями дистрибутивно-статистического анализа (тезаурус), выделения ключевых слов, определения частотности единиц и др. В основном он используется в исследованиях по корпусной лингвистике [2, 11, 13, 16].

В Sketch Engine был сформирован корпус из загруженных текстов научных публикаций. Затем были последовательно использованы следующие инструменты для анализа корпуса текстов.

На всём корпусе документов выявлена частота коллокаций слов (существительных), являющихся модификаторами по отношению к термину «Гаузе» (инструмент Word Sketch). Оказалось, что именно «принцип конкурентного исключения» в различных вариациях максимально часто употребляется в анализируемой литературе (рис. 1). Это подтверждает тот факт, что максимальное число публикаций в НЭБ было получено по соответствующему поисковому запросу.



Рис. 1. Визуализация частоты коллокаций существительных, являющихся модификаторами по отношению к термину «Гаузе» (инструмент Word Sketch).

Для более детального анализа полученных результатов можно рассмотреть отображение микроконтекстов, в которых встречаются коллокации термины «Гаузе» с её модификаторами (существительными),

для чего было произведено переключение в инструмент Concordance (рис. 2).

The screenshot shows the CONCORDANCE tool interface. At the top, there is a search bar containing 'Gause (Elibrary)'. Below the search bar, there is a status bar indicating 'CQL Гаузе (gen\_modifies) • 348' and '138.6 per million tokens • 0.014%'. The main area displays a list of search results, each with a checkbox, a document ID, and a snippet of text. The results are as follows:

Doc ID	Text Snippet
1 doc#0	<s> Подробное и достаточно критичное описание <b>опытов Гаузе</b> приводят В.Н.Тутубалин с соавторами (1999). </s>
2 doc#0	<s> К выводу о неудовлетворительности "обычного анализа экспери <b>ментов Гаузе</b> " и "более сложных экспериментов по конкуренции" прихо дит и М.Уильямсон (1975, с. 169-172). </s>
3 doc#0	<s> Реализуется он в условиях ограниченности какого-либо ресурса, что и было обязательным в <b>экспериментах Гаузе</b> с инфузориями. </s>
4 doc#0	<s> Явление викариата, взаимоотношения викарирующих видов соответствуют некоторым экологическим законам, наиболее общий из которых — принцип конкурентного исключения (или <b>теорема Гаузе</b> ). </s>
5 doc#0	<s> Реализация данного принципа (" <b>закон Гаузе</b> ") маловероятна в водохранилищах Волжского бассейна, которые характеризуются значительной гетерогенностью за счет наличия обильных мелководий, заливов и плесов, а также смешения вод различного генезиса и связаны единым стоком с бассейнами южных и северных морей. </s>
6 doc#0	<s> Концепция паразитарной системы обосновывает экстраполяцию на явления паразитизма в практическом плане многих общебиологических принципов, наиболее существенными и практически полезными в числе которых являются следующие: концепция паразитарной системы популяционная концепция концепция экологической ниши правило усиления интеграции биосистем И.И. Шмальгаузена принцип оптимальной группировки (принцип Олли) принцип внезапного усиления патогенности принцип конкурентного исключения ( <b>принцип Гаузе</b> ). </s>

Рис. 2. Контексты использования термина «Гаузе» с модификаторами – словами существительными (инструмент Concordance).

Далее для выявления контекстов по направлению исследований Г.Ф. Гаузе в области разработки антибиотиков был применён инструмент коллокаций для двух терминов (Word Sketch Difference) – «Гаузе» и «антибиотики». Из визуализации полученных результатов видно, что у этих терминов есть устойчивые сочетания и чем ближе термин к центру графа, тем более часто он встречается совместно с обоими терминами, выбранными для анализа. Из анализа совместных коллокаций видно, что вместе с Гаузе над разработкой антибиотиков работали Алпатов и Бражникова (рис. 3). В рамках синтетического метода «Гаузе» и «антибиотики» являются опорными термин-концептами, а связанные с ними – зависимыми термин-концептами, которые составляют с опорными терминами устойчивые пары и участвуют в формировании терминологического ландшафта исследуемой тематики.

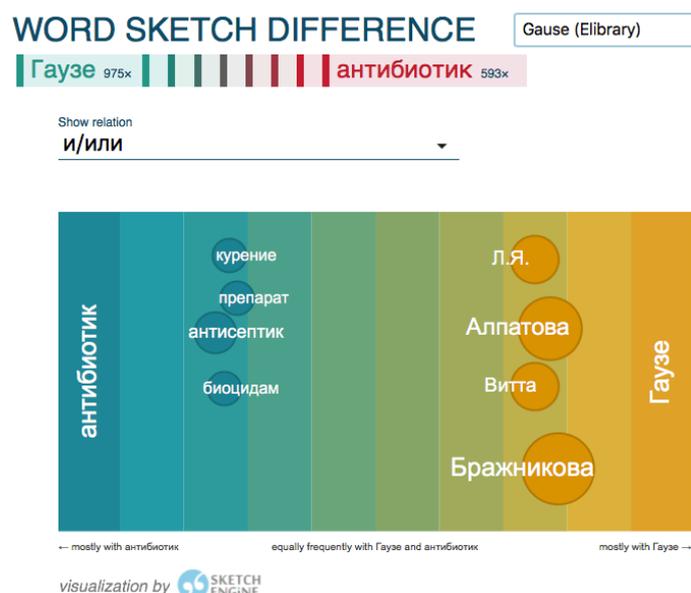


Рис. 3. Визуализация коллокаций пары терминов «Гаузе» и «антибиотики» (инструмент Word Sketch Difference).

Выявленные пары термин-концептов «Гаузе – Алпатова» и «Гаузе – Бражникова» позволяют сделать вывод о том, что разработка антибиотиков велась Гаузе совместно с этими коллегами. Поэтому второй каскадный запрос включил в себя пару терминов «Гаузе» и «Алпатова». Совместная коллокация этих терминов выявила помимо Бражниковой ещё несколько фамилий (рис. 4).

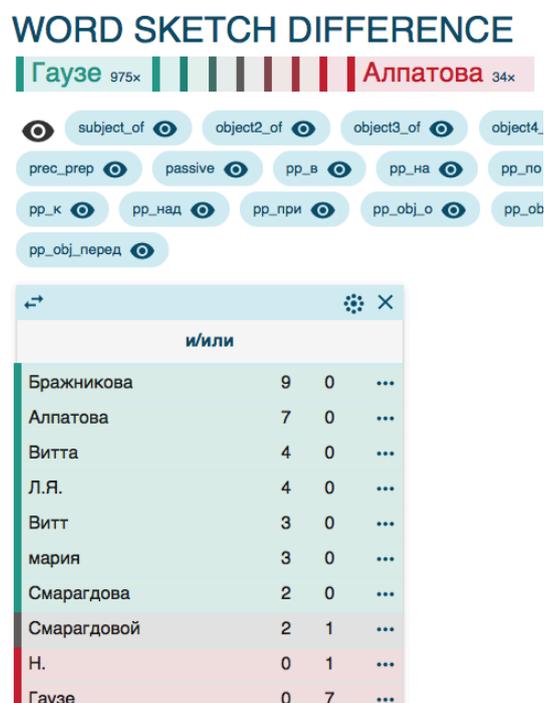


Рис. 4. Результаты совместной коллокации по паре терминов «Гаузе» и «Алпатова» (инструмент Word Sketch Difference).

Переключение в режим визуализации наглядно показывает, что в исследованиях совместно с Алпатовой принимала активное участие Смарагдова – нахождение термина в середине, что означает устойчивое сочетание термина «Смарагдовой» как с термином «Гаузе». Так и с термином «Алпатова» (рис. 5).

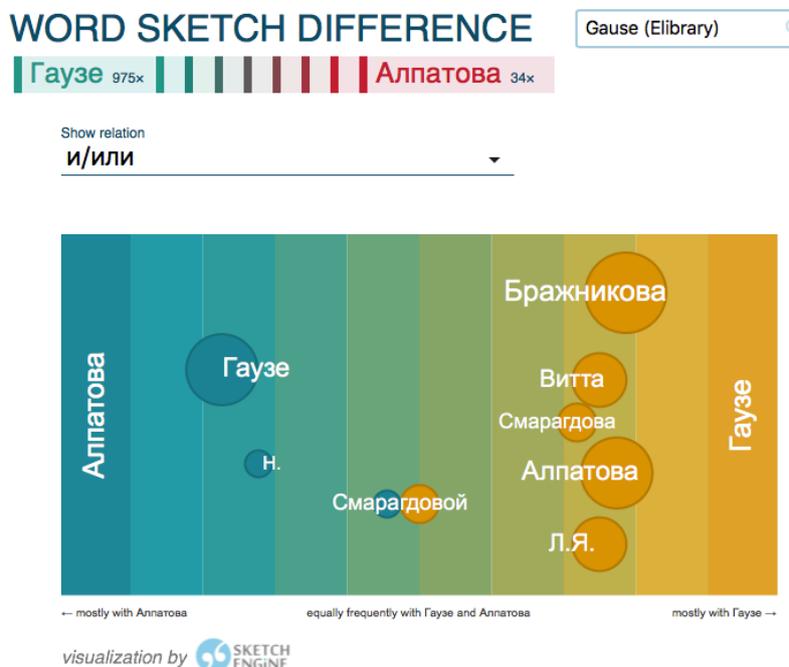


Рис. 5. Визуализация результатов совместной коллокации по паре терминов «Гаузе» и «Алпатова» (инструмент Word Sketch Difference).

При переключении в режим Concordance мы получаем всего 7 контекстов совместного использования терминов «Гаузе» и «Алпатова» (рис. 6). Анализ этих микроконтекстов позволяет сделать вывод о том, что Алпатов является мужчиной (в текстах фамилия встречается только в родительном падеже). При этом Гаузе и Алпатов проводили совместные исследования в области биологии (экологии), относящиеся к первой половине 1930-х годов.

Details		sentence
1	<input type="checkbox"/> ⓘ doc#0	<s> Летом 1933 г. Вернадский пригласил Гаузе и Алпатова к себе в санаторий Академии наук «Узкое», где тогда отдыхал. </s>
2	<input type="checkbox"/> ⓘ doc#0	<s> До поступления на работу в лабораторию Алпатова, Гаузе около года, в 1931–1932 гг., работал в биохимической лаборатории Политехнического музея в Москве, где в то время работали известный биохимик проф. </s>
3	<input type="checkbox"/> ⓘ doc#0	<s> Летом 1933 г. В.И. Вернадский (1863–1945) пригласил Гаузе и Алпатова посетить его в санатории Академии наук «Узкое», где он отдыхал. </s>
4	<input type="checkbox"/> ⓘ doc#0	<s> Междисциплинарный научный и прикладной журнал «Биосфера», 2011, т.3, №3 Я.М. галл Исходной теоретической установкой для выполнения экспериментальных работ служило положение Гаузе и Алпатова «Об обратной зависимости между приобретенными и врожденными свойствами организмов» [26]: чем сильнее какойлибо признак закреплен генетически, тем меньшей оказывается возможная модификация этого признака при изменении условий среды, и наоборот, чем меньшей является величина генотипической детерминации признака, тем значительнее оказывается величина модификации. </s>
5	<input type="checkbox"/> ⓘ doc#0	<s> Благодаря работам Гаузе, Алпатова и Смарагдовой было установлено, что простейшие, как и другие животные, подвержены географической изменчивости. </s>
6	<input type="checkbox"/> ⓘ doc#0	<s> До поступления на работу в лабораторию Алпатова, Гаузе около года, в 1931– 1932 гг., трудился в биохимической лаборатории Политехнического музея в Москве, где в то время работали профессор А.Р. Кизель и его ученики А.Н. Белозерский и В.Л. Кретович. </s>

Рис. 6. Контексты коллокаций терминов «Гаузе» и «Алпатова» (инструмент Concordance).

Анализ совместных коллокаций терминов «Гаузе» и «Бражникова» показывает иную картину – Гаузе и Бражникова совместно занимались разработкой антибиотиков и их основные исследования проводились в период Великой Отечественной Войны (рис. 7).

Details		sentence
1	<input type="checkbox"/> ⓘ doc#0	<s> Уже летом 1942 г. Гаузе и Бражникова изолировали из линии Bacillus brevis (var. </s>
2	<input type="checkbox"/> ⓘ doc#0	<s> В 1949 г. Гаузе и Бражникова открыли альбомуцин — антибиотик с очень низкой токсичностью, который широко применялся при лечении воспаления легких даже у детей [12]. </s>
3	<input type="checkbox"/> ⓘ doc#0	<s> В 1949 г. Гаузе и Бражникова открыли альбомуцин, антибиотик с очень низкой токсичностью, и он широко применялся при лечении воспаления легких даже у детей. </s>
4	<input type="checkbox"/> ⓘ doc#0	<s> Здесь у него сложились дружеские отношения с А.Н. Белозерским, и в самый ранний период работы над антибиотиками, в 1942–1943 гг., он предложил Белозерскому выполнить биохимический анализ грамицидина S (Гаузе, Бражникова, Белозерский, Пасхина, 1944; Belozerskii and Passhina, 1944; Белозерский, Пасхина, 1945). </s>
5	<input type="checkbox"/> ⓘ doc#0	<s> Гаузе и Бражникова уже летом 1942 г. получили первый оригинальный отечественный антибиотик грамицидин S (грамицидин советский). </s>
6	<input type="checkbox"/> ⓘ doc#0	<s> Его удалось внедрить в практику здравоохранения, и уже в 1943–1944 гг. он широко использовался в госпиталях и в военно-полевых условиях ( Гаузе, Бражникова, 1943а; Гаузе Бражникова, 1943b; Цируленко, Кост, Стенко, 1943; Гуляева, 1943; Бурденко, 1946). </s>
7	<input type="checkbox"/> ⓘ doc#0	<s> В 1949 г. Гаузе и Бражникова открыли альбомуцин — антибиотик с очень низкой токсичностью, он широко применялся при лечении воспаления легких даже у детей (Гаузе, Бражникова, 1951; Доброхотова, 1951; Лурье, Линяева, 1951). </s>

Рис. 6. Контексты коллокаций терминов «Гаузе» и «Бражникова» (инструмент Concordance).

Также представляют собой интерес и результаты совместных коллокаций терминов «Гаузе» и «асимметрия». На всём корпусе текстов выявлено 11 таких коллокаций, а анализ микроконтекстов позволяет сделать вывод, что в этих фрагментах идёт речь об исследованиях Гаузе асимметрии протоплазмы (Рис. 8).

The screenshot shows the CONCORDANCE interface with the search term 'Gause (Elibrary)'. The results are displayed in a table-like format with 7 rows. Each row contains a document ID (doc#0), a search term in red, and a snippet of text. The search term 'асимметрия' is highlighted in red in the original image. The interface also shows a search bar, a toolbar with various icons, and a 'sentence' dropdown menu.

Doc ID	Search Term	Text Snippet
1 doc#0	асимметрии	<s> Впервые Гаузе заинтересовался биологическим действием оптических изомеров и <b>проблемой асимметрии</b> протоплазмы (аминокислоты, сахара) еще в 1931—1932 гг., когда работал в биохимической лаборатории Поли технического музея в Москве, где в то время трудились известный биохимик А.Р.Кизель* и его ученики А.Н.Белозерский и В.Л.Кротович. </s>
2 doc#0	проблемой асимметрии	<s> В небольшой лаборатории Кизеля Гаузе впервые заинтересовался биологическим действием оптических изомеров и <b>проблемой асимметрии</b> протоплазмы (аминокислоты, сахара). </s>
3 doc#0	проблемы асимметрии	<s> Обычно Междисциплинарный научный и прикладной журнал «Биосфера», 2011, т.3, №3 Я.М. галл раз в месяц Вернадский приглашал Гаузе к себе домой, и они обсуждали <b>проблемы асимметрии</b> протоплазмы, а также проблемы экологии, которые очень волновали Вернадского, так как его исследования были связаны с закономерностями роста живого вещества в биосфере. </s>
4 doc#0	происхождении асимметрии	<s> Теоретические рассуждения Пастера о <b>происхождении асимметрии</b> протоплазмы носили расплывчатый характер и явно отставали от его экспериментальных исследований. </s>
5 doc#0	Проблема асимметрии	<s> <b>Проблема асимметрии</b> протоплазмы имела длительную историю, и изучалась она в основном химиками-органиками и физиками; Гаузе внес в нее свой оригинальный вклад. </s>
6 doc#0	проблемой асимметрии	<s> В этой лаборатории Гаузе впервые заинтересовался биологическим действием оптических изомеров и <b>проблемой асимметрии</b> протоплазмы (аминокислоты, сахара). </s>
7 doc#0	проблемы асимметрии	<s> Обычно раз в месяц тот приглашал Гаузе к себе домой, и они обсуждали <b>проблемы асимметрии</b> протоплазмы, а также проблемы экологии, которые очень волновали Вернадского, так как его исследования

Рис. 8. Контексты коллокаций терминов «Гаузе» и «ассимметрия» (инструмент Concordance).

В целом, полученные списки микроконтекстов коллокаций, полученных через разные инструменты Sketch Engine, возможно экспортировать в одном из общепринятых форматов (txt, csv, xls, xml). Это позволяет либо обрабатывать их в других системах, либо проводить более качественный экспертный анализ.

## 5. Заключение

Проведённое исследование позволяет сделать следующие основные выводы:

- для формирования корпуса текстовых документов научных публикаций, полученных из НЭБ, необходима предварительная обработка. Например, распознавание текста и сохранение документа с текстовым слоем;
- различные инструменты системы обработки текстов на естественном языке Sketch Engine позволяют формировать термины, необходимые для анализа при проведении научных исследований;

- использование систем обработки текстов на естественных языках позволяет быстро и эффективно находить нужные контексты, релевантные проводимому исследованию;

- визуализация в Sketch Engine полученных результатов даёт наглядное представление зависимостей анализируемых терминов и помогает выбрать более релевантные тематике термин для дальнейшего анализа;

- возможность построения списков коллокаций анализируемых терминов позволяет создавать коллекции релевантных исследованию макроконтекстов. А возможность их выгрузки в файлы распространённых форматов даёт возможность обрабатывать их в других системах или проводить качественный экспертный анализ.

В общем, проведённое исследование показало эффективность использования синтетического метода, основанного на комплексном применении информационно-коммуникационных технологий, для поиска, экспликации и анализа текстовых данных научных публикаций в науковедческих исследованиях.

Дальнейшее развитие данного исследования предполагает изучение возможностей других общедоступных инструментов для экспликации и анализа сформированного корпуса научных публикаций. Например, будут рассмотрены такие системы как T-Libra и Voyant-tools. Помимо этого будут исследованы возможности поиска и экспликации данных научных публикаций мирового научного дискурса из зарубежных цифровых научных ресурсов с последующим анализом в различных инструментах и информационных системах.

Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований, проект 18-011-00923-а.

## Литература

1. Алексеева С.В., Кольцов С.Н., Кольцова О.Ю. Linis-crowd.org: лексический ресурс для анализа тональности социально-политических текстов на русском языке // Компьютерная лингвистика и вычислительные онтологии. Сборник научных статей XVIII Объединенной конференции «Интернет и современное общество» IMS-2015, Санкт-Петербург, 23-25 июня 2015 г. — ИТМО, 2015. — С. 25-34. — URL: <https://openbooks.itmo.ru/ru/file/2203/2203.pdf>
2. Базарова Б.Б. Sketch Engine в лингвистических исследованиях // Казанская наука. — 2018. — №5. — С. 28-30
3. Галл Я.М. Георгий Францевич Гаузе (1910-1986). — СПб. Нестор-История. — 2012. — 234 с.
4. Интерфейс программирования API [электронный текст] // Elibrary. — URL: [https://www.elibrary.ru/projects/api/api\\_info.asp](https://www.elibrary.ru/projects/api/api_info.asp)

5. Кононова О.В., Ляпин С.Х., Прокудин Д.Е. Синтетический метод извлечения контекстного знания в русскоязычной социально-гуманитарной сфере: комплексный подход // Информационное общество: образование, наука, культура и технологии будущего. — 2017. — Вып. 1. — С. 52-67. — DOI: 10.17586/2587-8557-2017-1-52-67
6. Кононова О.В., Прокудин Д.Е. Подход к извлечению, экспликации и представлению контекстного знания при изучении развивающихся междисциплинарных направлений исследований // International Journal of Open Information Technologies. — 2020. — Том 8. — № 1. — С. 90-101. — URL: <http://injoit.org/index.php/j1/article/view/882/844>
7. Кононова О.В., Ляпин С.Х., Прокудин Д.Е. Исследование терминологической базы междисциплинарного научного направления "цифровая экономика" с использованием инструментов контекстного анализа // International Journal of Open Information Technologies. — 2018. — Т. 6. — № 12. — С. 57-66
8. Кононова О.В., Прокудин Д.Е. Геймификация и социокультурные контексты цифрового урбанизма: общий подход к анализу и прогнозу научного дискурса [электронный текст] // Культура и технологии. — 2019. — Том 4. — Вып. 3. Серьезные компьютерные игры: где, как и почему они работают. — С. 83-105. — DOI: 10.17586/2587-800X-2019-4-3-83-105
9. Коршунов А., Гомзин А. Тематическое моделирование текстов на естественном языке // Труды Института системного программирования РАН. — 2012. — С. 215-242. — URL: <https://elibrary.ru/item.asp?id=18361454>
10. Манаев О.Т. Контент-анализ как метод исследования // — URL: <http://psyfactor.org/lib/content-analysis3.htm>
11. Масевич А.Ц., Захаров В.П. Семантические поля обозначений цвета в русском языке // Компьютерная лингвистика и вычислительные онтологии. — Выпуск 3 (Труды XXII Международной объединенной научной конференции «Интернет и современное общество», IMS-2019, Санкт-Петербург, 19–22 июня 2019 г. Сборник научных трудов). — СПб: Университет ИТМО, 2019. — С. 61–76. — DOI: 10.17586/2541-9781-2019-3-61-76
12. Москвина А.Д., Митрофанова О.А., Ерофеева А.Р., Харabet Я.К. Автоматическое выделение ключевых слов и словосочетаний из русскоязычных корпусов текстов с помощью алгоритма RAKE // Труды международной конференции «Корпусная лингвистика–2017». — СПб. 2017. — С. 268–277
13. Новикова А. А. Сравнение инструментов Sketch Engine и Termostat для извлечения терминологии // International Journal of Open Information Technologies. — 2020. — №11. — С. 73-79. — URL: <http://injoit.org/index.php/j1/article/view/1029>

14. Прокудин Д.Е., Левит Г.С. Методы отбора цифровых информационных ресурсов на примере исследования влияния научных идей Г.Ф. Гаузе на развитие науки // Научный сервис в сети Интернет: труды XIX Всероссийской научной конференции (18-23 сентября 2017 г., г. Новороссийск). — М.: ИПМ им. М. В. Келдыша, 2017. — 480 с. — DOI: 10.20948/abrau-2017-75
15. Прокудин Д.Е., Левит Г.С. Применение методов визуализации геоданных, полученных из неструктурированных цифровых данных, в науковедческих исследованиях (на примере научного наследия Г.Ф. Гаузе) // Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23-28 сентября 2019 г., г. Новороссийск). — М.: ИПМ им. М.В.Келдыша, 2019. — С. 580-592. — DOI:10.20948/abrau-2019-23
16. Чекменева А.В. Метафора в музыковедческом Дискурсе // Компьютерная лингвистика и вычислительные онтологии. Выпуск 3 (Труды XXII Международной объединенной научной конференции «Интернет и современное общество», IMS-2019, Санкт-Петербург, 19–22 июня 2019 г. Сборник научных трудов). — СПб: Университет ИТМО, 2019. — С. 142–152. — DOI: 10.17586/2541-9781-2019-3-142-152
17. Effendi D.N. et al. Bibliometric analysis of scientific literacy using VOS viewer: Analysis of science education // J. Phys.: Conf. Ser. — 2021. — Vol. 1796. — 012096. — DOI: 10.1088/1742-6596/1796/1/012096
18. Gause G.F. The Struggle for Existence. — Baltimore, 1934
19. Lotka A. J. Analytical note on certain rhythmic relations in organic systems // Proc. Nat. Acad. — 1920. — No. 6. — P. 410-415
20. Prokudin D., Levit G., Hossfeld U. Selection Methods of Digital Information Resources for Scientific Heritage Studies: A Case Study of Georgy F. Gause // Internet and Modern Society: Proceedings of the International Conference IMS-2017 (St. Petersburg; Russian Federation, 21-24 June 2017). Radomir V. Bolgov, Nikolai V. Borisov, Leonid V. Smorgunov, Irina I. Tolstikova, Victor P. Zakharov (eds). — ACM International Conference Proceeding Series, ACM Press, N.Y., 2017. — P. 69-74. — DOI: 10.1145/3143699.3143739
21. Prokudin D., Levit G., Hossfeld U. Selection Methods for Quantitative Processing of Digital Data for Scientific Heritage Studies // Alexandrov D., Boukhanovsky A., Chugunov A., Kabanov Y., Koltsova O. (eds). Digital Transformation and Global Society. DTGS 2018. Communications in Computer and Information Science, 2018, vol. 859. P. 134-145. — DOI: 10.1007/978-3-030-02846-6\_11
22. Sood S.K., Kumar N., Saini M. Scientometric analysis of literature on distributed vehicular networks: VOSViewer visualization techniques // Artif Intell Rev. — 2021. — DOI: 10.1007/s10462-021-09980-4

23. Volterra V. Fluctuations in the abundance of a species considered mathematically // *Nature*. — 1926. — No. 118. — P. 558-560

### References

1. Alexeeva S., Kolcov S., Koltsova O. Linis-crowd.org: A lexical resource for Russian sentiment analysis of social media // // *Komp'yuternaya lingvistika i vychislitel'nye ontologii. Sbornik nauchnykh statey XVIII Ob"edinennoy konferentsii «Internet i sovremennoe obshchestvo» IMS-2015, Sankt-Peterburg, 23-25 iyunya 2015 g.* — ITMO, 2015. — P. 25-34. — URL: <https://openbooks.itmo.ru/ru/file/2203/2203.pdf>
2. Bazarova B.B. Sketch Engine v lingvisticheskikh issledovaniyakh // *Kazanskaya nauka*. — 2018. — №5. — S. 28-30
3. Gall Ya.M. Georgiy Frantsevich Gause (1910-1986). — SPb. Nestor-Istoriya. — 2012. — 234 p.
4. Interfeys programirovaniya API // *Elibrary*. — URL: [https://www.elibrary.ru/projects/api/api\\_info.asp](https://www.elibrary.ru/projects/api/api_info.asp)
5. Kononova O.V., Lyapin S.Kh., Prokudin D.E. Synthetic Method of Contextual Knowledge Extraction in the Russian socio-humanitarian sphere: an integrated approach // *Information Society: Education, Science, Culture and Technologies of the Future*. — 2017. — Vol. 1. — P. 52-67. — DOI: 10.17586/2587-8557-2017-1-52-67
6. Kononova O.V., Prokudin D.E. An approach to extraction, explication and presentation of contextual knowledge in the study of developing interdisciplinary research areas // *International Journal of Open Information Technologies*. — 2020. — Vol 8. — № 1. — P. 90-101. — URL: <http://injoit.org/index.php/j1/article/view/882>
7. Kononova O.V., Lyapin S.Kh., Prokudin D.E. Studying the Interdisciplinary Terminological Landscape of Digital Economy with the Use of Contextual Analysis Tools // *International Journal of Open Information Technologies*. — 2018. — Vol. 6. — № 12. — P. 57-66. — URL: <http://injoit.org/index.php/j1/article/view/648>
8. Kononova O.V., Prokudin D.E. Gamification and Socio-cultural Contexts of Digital Urbanism: the General Approach to the Analysis and Prediction of Scientific Discourse // *International Culture & Technology Studies*. — 2019. — Vol. 4. — Iss. 3. — P. 83-105. — DOI: 10.17586/2587-800X-2019-4-3-83-105
9. Korshunov A., Gomzin A. Topic modeling in natural language texts // *Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS)*. — 2012. — Vol. 23. — DOI: 10.15514/ISPRAS-2012-23-13
10. Manaev O.T. Kontent-analiz kak metod issledovaniya // — URL: <http://psyfactor.org/lib/content-analysis3.htm>

11. Masevich A., Zakharov V. Semantic fields of color designations in Russian // *Computer Linguistics and Computing Ontologies*. Vol. 3 (Proceedings of the XXII International Joint Scientific Conference «Internet and Modern Society», IMS-2019, St. Petersburg, June 19-22, 2019). — St. Petersburg: ITMO University, 2019. — P. 61–76. — DOI: 10.17586/2541-9781-2019-3-61-76
12. Moskvina A.D., Mitrofanova O.A., Erofeeva A.R., Kharabet Ya.K. Automatic extraction of key words and phrases from Russian text corpora by means of RAKE algorithm // *Proceedings of the international conference «Corpus Linguistics–2017»*, June 27–30, 2017, St. Petersburg. — СПб. 2017. — P. 268–277
13. Novikova A. A. Sketch Engine and TermoStat tools for automatic term extraction // *International Journal of Open Information Technologies*. — 2020. — Vol. 8. — №11. — С. 73-79. — URL: <http://injoit.org/index.php/j1/article/view/1029>
14. Prokudin D.E., Levit G.S. Metody otbora tsifrovyykh informatsionnykh resursov na primere issledovaniy vliyaniya nauchnykh idey G.F. Gauze na razvitiye nauki // *Nauchnyy servis v seti Internet: trudy XIX Vserossiyskoy nauchnoy konferentsii (18-23 sentyabrya 2017 g., g. Novorossiysk)*. — M.: IPM im. M. V. Keldysha, 2017. — 480 c. — DOI: 10.20948/abrau-2017-75
15. Prokudin D.E., Levit G.S. Methods of GEODATA Visualization Extracted from Unstructured Digital Data in Scientific Research // *Nauchnyy servis v seti Internet: trudy XXI Vserossiyskoy nauchnoy konferentsii (23-28 sentyabrya 2019 g., g. Novorossiysk)*. — M.: IPM im. M.V.Keldysha, 2019. — С. 580-592. — DOI:10.20948/abrau-2019-23
16. Chekmeneva A. V. Metaphor in Musicological Discourse // *Computer Linguistics and Computing Ontologies*. Vol. 3 (Proceedings of the XXII International Joint Scientific Conference «Internet and Modern Society», IMS-2019, St. Petersburg, June 19-22, 2019). — St. Petersburg: ITMO University, 2019. — P. 142–152. — DOI: 10.17586/2541-9781-2019-3-142-152
17. Effendi D.N. et al. Bibliometric analysis of scientific literacy using VOS viewer: Analysis of science education // *J. Phys.: Conf. Ser.* — 2021. — Vol. 1796. — 012096. — DOI: 10.1088/1742-6596/1796/1/012096
18. Gause G.F. *The Struggle for Existence*. — Baltimore, 1934
19. Lotka A. J. Analytical note on certain rhythmic relations in organic systems // *Proc. Nat. Acad.* — 1920. — No. 6. — P. 410-415
20. Prokudin D., Levit G., Hossfeld U. Selection Methods of Digital Information Resources for Scientific Heritage Studies: A Case Study of Georgy F. Gause // *Internet and Modern Society: Proceedings of the International Conference IMS-2017 (St. Petersburg; Russian Federation, 21-24 June 2017)*. Radomir V. Bolgov, Nikolai V. Borisov, Leonid V. Smorgunov, Irina I. Tolstikova, Victor P. Zakharov (eds). — ACM International Conference Proceeding

- Series, ACM Press, N.Y., 2017. — P. 69-74. — DOI: 10.1145/3143699.3143739
21. Prokudin D., Levit G., Hossfeld U. Selection Methods for Quantitative Processing of Digital Data for Scientific Heritage Studies // Alexandrov D., Boukhanovsky A., Chugunov A., Kabanov Y., Koltsova O. (eds). Digital Transformation and Global Society. DTGS 2018. Communications in Computer and Information Science, 2018, vol. 859. P. 134-145. — DOI: 10.1007/978-3-030-02846-6\_11
  22. Sood S.K., Kumar N., Saini M. Scientometric analysis of literature on distributed vehicular networks: VOSViewer visualization techniques // Artif Intell Rev. — 2021. — DOI: 10.1007/s10462-021-09980-4
  23. Volterra V. Fluctuations in the abundance of a species considered mathematically // Nature. — 1926. — No. 118. — P. 558-560