

Применение методов количественной обработки цифровых данных в науковедческих исследованиях (на примере изучения влияния научных идей Г.Ф. Гаузе на развитие науки)

Д.Е. Прокудин, Г.С. Левит.

Аннотация — Настоящее исследование нацелено на применение методов извлечения метаданных из цифровых электронных ресурсов и последующей количественной их обработки. Оно основано на результатах проведённого тематического исследования по оценке возможностей цифровых информационных ресурсов для применения в науковедческих исследованиях на примере научного наследия Г.Ф. Гаузе, в котором были отобраны приоритетные цифровые ресурсы. Нами рассмотрены цифровые информационные ресурсы, предоставляющие метаданные в структурированном виде. Произведены сбор и количественная обработка метаданных, полученных из информационных ресурсов, представляющих метаданные в структурированном виде. Полученные результаты представлены в наглядном виде. На основе анализа полученных данных сделаны соответствующие выводы. Одним из основных выводов является вывод об эффективности использования цифровых информационных ресурсов в исследованиях по истории науки. Также нами оценена возможность извлечения метаданных из цифровых информационных систем, в которых они представлены в неструктурированном виде. Намечено продолжение исследования в двух направлениях: сбор и последующее применение методов количественной обработки метаданных, полученных из международных и российских цифровых информационных ресурсов, представляющих метаданные в неструктурированном виде и не обладающие инструментами выгрузки метаданных в общепринятых форматах.

Ключевые слова — научная информация, цифровые информационные ресурсы, извлечение структурированных данных, методы количественной обработки данных, Digital Humanities, дигитализация научного наследия, Г.Ф. Гаузе

I. ЦЕЛЬ И ЗАДАЧИ ИССЛЕДОВАНИЯ

Основной целью данного исследования является применение количественных методов обработки метаданных научных публикаций, отобранных из

различных цифровых информационных ресурсов. Результаты количественной обработки будут использованы для оценки возможности применения цифровых информационных ресурсов для качественного анализа влияния идей на развитие науки.

В рамках предлагаемого исследования решаются следующие задачи:

1. Извлечение структурированных метаданных научных публикаций из различных цифровых информационных ресурсов.
2. Количественная обработка полученных метаданных в соответствии с целями исследования.
3. Визуализация полученных результатов.
4. Интерпретация полученных результатов.

При этом также ставилась задача использовать такие методы, которые не предполагают применение сложных технологических решений и доступны подавляющему большинству исследователей (как по стоимости, так и по простоте использования).

II. ИСПОЛЪЗУЕМЫЕ ПОДХОДЫ И МЕТОДЫ

Основной подход, которого мы придерживаемся, — это доступность предлагаемых методов и инструментов для любого исследователя. В настоящем исследовании были использованы методы и подходы активно развивающегося междисциплинарного направления Digital Humanities, которое всё чаще используется и среди российских исследователей [1]-[3][7][10]:

- поиск данных в цифровых информационных ресурсах [5];
- извлечение данных из цифровых информационных ресурсов (Data Mining) [6][11]-[13][18-20];
- обработка больших данных (Big Data) [8][14][16].

При этом используемые методы можно разбить на две группы:

- методы извлечения данных из цифровых информационных ресурсов;
- методы количественной обработки полученных данных.

К первой группе мы относим:

- автоматическое извлечение структурированных

Статья получена 22.10.2018

Прокудин Дмитрий Евгеньевич, докт. филос. наук, СПбГУ, доцент; Университет ИТМО, аналитик, Россия; (e-mail: hogben.young@gmail.com).

Левит Георгий Семёнович, докт. биол. наук, доцент, Университет ИТМО, Россия (e-mail: gslevit@corp.ifmo.ru).

данных средствами цифровых информационных ресурсов;

- ручное извлечение данных из цифровых информационных систем, не предоставляющих средств экспорта.

Ко второй группе относятся:

- усечение наборов метаданных до минимально необходимых для достижения целей исследования (минимизация наборов метаданных);

- приведение наборов метаданных, извлечённых из разных цифровых информационных ресурсов, к единому стандарту представления;

- объединение метаданных, извлечённых из разных цифровых информационных ресурсов, в единый массив;

- устранение дубликатов;

- количественная обработка метаданных;

- наглядное представление полученных результатов;

- интерпретация полученных результатов.

III. БАЗА ИССЛЕДОВАНИЯ

На первом этапе исследования мы изучили возможность применения цифровых информационных ресурсов для их использования в исследованиях по истории науки. Оценка проводилась по разным критериям. Одним из основных критериев была выбрана репрезентативность данных, полученных по сформулированным поисковым запросам [9][17].

В качестве предмета исследования было выбрано влияние идей выдающегося советского биолога Г.Ф. Гаузе на развитие науки [4]. Гаузе был выбран по ряду причин. Во-первых, он был выдающимся ученым сталинской и пост-сталинской эпохи и при этом остался абсолютно непричастным к антинаучным течениям, таким как лысенковщина [15]. Он также не был репрессирован и его научная карьера развивалась непрерывно. Гаузе был научно активен в течение всей жизни и пользовался авторитетом как в СССР, так и за рубежом. Во-вторых, он публиковался на английском языке в ведущих западных издательствах и журналах. Его главная книга «Борьба за существование» была впервые опубликована на английском языке в США. Эти факты биографии Гаузе позволяют поставить вопрос о влиянии учёного, жившего в тоталитарном изолированном обществе на развитие науки внутри этого общества и за его пределами. В-третьих, Гаузе известен сразу в нескольких научных областях. Наиболее известен его вклад в эволюционную теорию и экологию. Он был первым, кто экспериментально исследовал борьбу за существование. Закон, выведенный Гаузе и известный как «принцип конкурентного исключения», доказывает, что никакие два вида, конкурирующие за ограниченный ресурс, не могут сосуществовать в рамках единой экологической ниши. Менее известно, что Гаузе был выдающимся исследователем в области медицины, и в период Великой Отечественной Войны разработал первый советский антибиотик Gramacidin S.

Научоведческий аспект нашего исследования

заключается в сравнении влияния Гаузе в СССР и за рубежом с целью нахождения закономерностей, которые позднее могли бы быть объединены традиционными средствами истории науки.

Мы произвели оценку различных цифровых информационных ресурсов для возможности извлечения из них структурированных данных. При этом были отобраны следующие ресурсы, которые стали базой исследования: SCOPUS, Academic Search Complete (EBSCO) и SpringerLink.

Выбор этих ресурсов обоснован тем, что:

- они обладают инструментами выгрузки метаданных публикаций в структурированном виде;

- в них было найдено достаточное количество записей.

Также в настоящем исследовании мы использовали информационно-поисковую систему Google Scholar, которая, хотя и не позволяет получить результаты поиска в структурированном виде, обладает возможностью использовать различные фильтры.

IV. ИЗВЛЕЧЕНИЕ СТРУКТУРИРОВАННЫХ ДАННЫХ

При извлечении метаданных для исследования мы использовали инструменты импорта привлекаемых в исследование цифровых информационных ресурсов. Во всех рассматриваемых системах мы использовали поисковый запрос “Gause AND competitive exclusion principle”. В информационной системе Scopus по этому запросу было получено 575 записей. Эти записи были сохранены в текстовый файл формата CSV (Comma-Separated Values).

Затем в информационной системе EBSCO было получено 103 библиографических записи. Эти записи были размещены на 3 страницах (максимальное число записей на страницу - 50). В EBSCO не предусмотрена выгрузка всех результатов поиска. Поэтому пришлось сначала сохранить результаты с каждой страницы в специальную папку, а затем произвести импорт метаданных публикаций и объединить их в один файл формата CSV.

Поиск в системе SpringerLink дал результат в 462 записи. Эти записи также были сохранены в файле формата CSV.

V. КОЛИЧЕСТВЕННАЯ ОБРАБОТКА ПОЛУЧЕННЫХ ДАННЫХ

Для количественной обработки полученных результатов нами были выбраны электронные таблицы Microsoft Excel. Этот инструмент сочетает в себе как мощные вычислительные средства для обработки данных, так и возможности визуализации. Это необходимо для презентации полученных результатов и их интерпретации.

Мы импортировали полученные данные из цифровых электронных ресурсов отдельно на разные листы. Оказалось, что все наборы данных имеют разную структуру и порядок следования полей. В связи с этим мы определили минимально необходимый для дальнейшей обработки набор данных. В него вошли

следующие поля:

- Item Title (название публикации)
- Authors (авторы)
- Publication Year (год публикации)
- Publication Title (название издания)
- Journal Volume (том)
- Journal Issue (выпуск)
- Content Type (тип публикации)
- Item DOI

Все три набора данных мы вручную расположили в этом порядке следования полей. Затем все наборы данных были скопированы на отдельный лист. Результирующий набор данных составил 1140 записей.

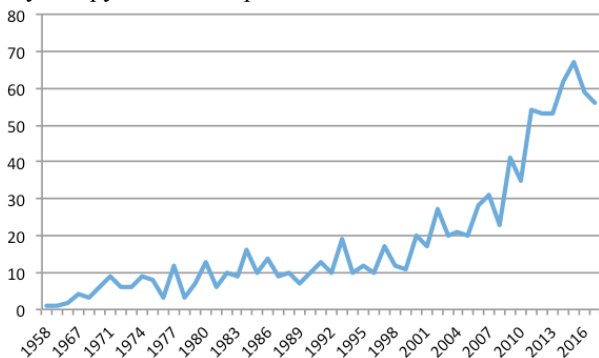


Рис. 1. Распределение публикаций по годам

После этого мы произвели сортировку записей по полю "Item Title". Оказалось, что в результирующем наборе присутствуют дубликаты некоторых записей. В результате ручной обработки дубликаты были удалены. После устранения дубликатов результирующий набор данных составил 996 записей.

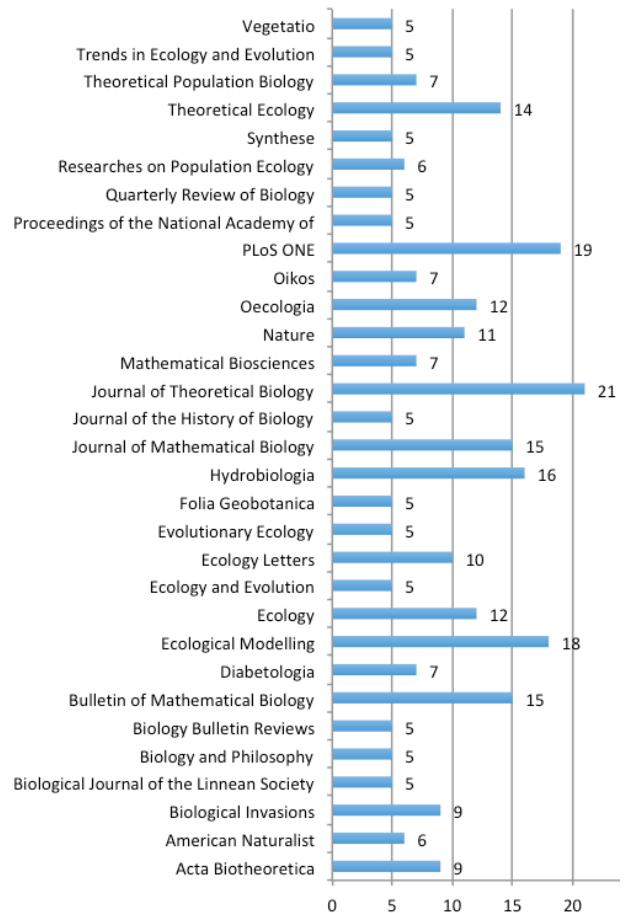


Рис. 2. Распределение публикаций по журналам

Для анализа динамики числа публикаций мы использовали инструмент «Сводная таблица». В сводную таблицу мы разместили данные из поля "Publication Year". Затем мы построили график, в котором просуммировали поля с одинаковым значением (рис. 1).

С использованием сводной таблицы были получены данные распределения публикаций по полю "Publication Title". При этом мы отобрали только статьи их журналов (значение поля "Content Type" = "Article"), а в результирующий набор включили журналы начиная с 5 публикаций. Для наглядности данные были представлены в виде гистограммы (рис. 2).

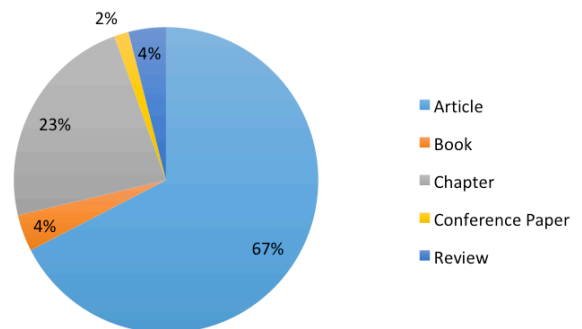


Рис. 3. Распределение публикаций по типам

В заключении таким же методом были получены

результаты распределения публикаций по типам (поле "Content Type"). Результаты были оформлены в виде круговой диаграммы (рис. 3).

VI. ВОЗМОЖНОСТЬ ИСПОЛЬЗОВАНИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ, ПРЕДСТАВЛЯЮЩИХ НЕСТРУКТУРИРОВАННЫЕ ДАННЫЕ

Большим исследовательским потенциалом обладают цифровые информационные ресурсы, предоставляющие свои данные в неструктурированном виде. Первоначальная оценка показала, что по использованному поисковому запросу "Gause AND competitive exclusion principle" из этих ресурсов были получены следующие результаты:

OpenDOAR - 6 520 записей;

Google Scholar – 1 710 записей без учёта цитирований.

OpenDOAR использовать для дальнейшего исследования не представляется возможным, так у этого цифрового информационного ресурса отсутствуют инструменты расширенного поиска и дополнительной фильтрации результатов, а также отсутствуют возможности выгрузки результатов поиска.

В отличие от этой системы Google Scholar обладает возможностью фильтрации результатов поиска. К тому же эта система представляет результаты поиска с использованием ссылок из «полнотекстовых журнальных статей, технических отчётов, препринтов, диссертаций, книг и других документов, в том числе выбранных веб-страниц, которые считаются „научными“» [20], которые собираются и индексируются из всех доступных открытых источников.

Мы произвели поиск в Google Scholar по базовому для нашего исследования запросу. В результате было получено 1930 ссылок на публикации (с цитатами) начиная с 1960 года. При помощи временного фильтра были получены ссылки по каждому году в интервале с 1960 по 2017 год. При этом было использовано 58 запросов. Полученные данные были внесены в электронную таблицу Microsoft Excel. Затем был построен соответствующий график (рис. 4, А).



Рис. 4. Распределение публикаций по годам по результатам Google Scholar (А – запрос "Gause AND competitive exclusion principle", В – запрос "* конкурентного исключения AND Гаузе")

Так как в Google Scholar индексируются все доступные сетевые цифровые информационные ресурсы научных публикаций, то мы использовали эту систему для поиска русскоязычных научных публикаций по теме исследования. Для этого был сформирован поисковый запрос "* конкурентного исключения" Гаузе. Мы использовали в запросе подстановочный символ "*", вместо которого может быть использовано любое слово. Это было сделано в связи с тем, что в русскоязычном научном дискурсе в качестве понятия "principle" могут использоваться слова "принцип" или "закон". По этому запросу было получено 252 записи начиная с 1984 г. При помощи фильтрации данные были распределены по годам. Они были занесены в электронную таблицу Microsoft Excel. Затем был построен соответствующий график (рис. 4, В).

VII. ВЫВОДЫ

Проведённое нами пилотное исследование на выбранных цифровых информационных ресурсах позволяет сделать следующие основные выводы.

1. Для получения больших объёмов данных необходимо использовать как можно больше цифровых информационных ресурсов. Так, например, при использовании всего лишь трёх ресурсов в рамках настоящего исследования оказалось менее 13% дублей.

2. Полученные результаты количественной обработки метаданных публикаций показывают постоянную положительную динамику. При этом результаты распределения публикаций по годам по рассмотренным трём цифровым информационным ресурсам коррелируют с аналогичными результатами, полученными из Google Scholar.

3. Количественный анализ распределения публикаций по журналам позволяет сделать вывод, что максимальное число статей опубликовано в журналах по следующим тематикам: экология, биология.

4. Количественный анализ распределения публикаций по типам позволяет сделать вывод, что подавляющее большинство из них представляют научные статьи (67%). Но, скорее всего, этот процент ещё больше, так как многие публикации, которые в поле "Content Type" имеют значение "Chapter", представляют собой научные статьи из сборников материалов конференций (proceedings), или коллективных монографий.

5. Сравнительный анализ распределения публикаций по годам по результатам данных, полученных из Google Scholar, говорит о том, что русскоязычные публикации по исследуемой теме также имеют положительную динамику. Эта динамика начинает наблюдаться с середины 2000-х годов. Связано это прежде всего с тем, что с этого времени начинают активно развиваться российские цифровые информационные ресурсы. До этого времени русскоязычные публикации встречаются крайне редко (первая публикация датируется 1981 годом). При этом между ними существуют лакуны (отсутствие публикаций): 1982-83, 1985-89, 1993-96, 2001 гг. Отрицательную динамику начиная с 2014 года можно объяснить тем, что русскоязычные публикации

попадают в цифровые информационные ресурсы с некоторой задержкой. Однако, это предположение можно будет проверить только через некоторое время (1-3 года).

Дальнейшее развитие данного исследования предполагает изучение возможностей извлечения данных из российских цифровых информационных ресурсов (например, Научная электронная библиотека, Истина, Киберленинка). Последующая количественная обработка этих данных позволит сравнить развитие возможностей мировых и российских ресурсов для использования в исследованиях по истории науки.

БИБЛИОГРАФИЯ

- [1] Борисов Н.В. Мультимедиа-реконструкция театрального события. Премьера спектакля «Чайка» на Александринской сцене 17 октября 1896 г. / Борисов Н.В., Никитин А.В., Смолин А.А., Трушин В.А., Чепуров А.А., Чепурова О.А. // Культура и технологии. 2016. Том 1. Вып. 1. С. 15-23. URL: <http://cat.ifmo.ru/2016/v1-i1/65>
- [2] Бородин Л. И. Виртуальные реконструкции исторической городской застройки: новые технологии презентации музейных экспозиций // Роль музеев в информационном обеспечении исторической науки: сборник статей. М.: Этерна, 2015. С. 386–395.
- [3] Бородин Л.И. Виртуальная реконструкция монастырских комплексов Москвы: проекты в контексте Digital Humanities // Вестник Пермского университета. Серия: История]. 2014. № 3 (26). С. 107-112.
- [4] Галл Я.М. Георгий Францевич Гаузе (1910-1986). СПб: Нестор-История, 2012. 234 с.
- [5] Гарскова И.М. Информационное обеспечение гуманитарных исследований в цифровую эпоху: модели формирования и развития // Вестн. Перм. ун-та. Сер. История. 2014. №3 (26). С. 76-86.
- [6] Донина О.В. Применение методов Data Mining для решения лингвистических задач // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2017. № 1. С. 154-160.
- [7] Куковякин А.В., Ляпин С.Х. Пространство Ломоносова: опыт функциональной интеграции виртуального атласа и полнотекстовой библиотеки // Культура и технологии. 2017. Том 2. Вып. 1. С. 6-12. URL: <http://cat.ifmo.ru/2017/v2-i1/98>
- [8] Майер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. М.: Манн, Иванов и Фербер. 2014. 240 с.
- [9] Прокудин Д.Е., Левит Г.С. Методы отбора цифровых информационных ресурсов на примере исследовании 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. С. 389-399. DOI: 10.20948/abrau-2017-75
- [10] Borisov N.V., Volkov O.G., Nikitina L.L., Nikolaev A.O., Smolin A.A., Stolyarov D.A. Application of Video 360deg Technology for the Presentation of the Solovetsky Monastery Cultural Heritage // Культура и технологии. 2016. Том 1. Вып. 1. С. 24-31 URL: <http://cat.ifmo.ru/en/2016/v1-i1/88>
- [11] Damerow J., Peirson B.R.E., Laubichler M.D. The Giles Ecosystem – Storage, Text Extraction, and OCR of Documents // Journal of Open Research Software. 2017. 5 (1). P. 26. DOI: 10.5334/jors.164
- [12] Huijnen P., Laan F., de Rijke M., Pieters T. A Digital Humanities Approach to the History of Science Eugenics Revisited in Hidden Debates by Means of Semantic Text Mining // SOCIAL INFORMATICS, SOCINFO 2013, Lecture Notes in Computer Science. 2014. Vol. 8359. Pp. 70 – 84.
- [13] Jockers M.L., Underwood T. Text-Mining the Humanities, in A New Companion to Digital Humanities /eds S. Schreibman, R. Siemens and J. Unsworth. John Wiley & Sons, Ltd, Chichester, 2015. DOI: 10.1002/9781118680605.ch20
- [14] Kaplan F. A map for big data research in digital humanities // Front. Digit. Humanit. 2015. Vol. 2 (1). DOI: 10.3389/fdigh.2015.00001
- [15] Kolchinsky E.I., Hossfeld U., Kutschera U., Levit G.S. The revival of Lysenkoism in Russia and epigenetics // Current Biology. 2017. Vol. 27 (19). P. 1042-1047. DOI: 10.1016/j.cub.2017.07.045
- [16] Manning P. Big data in history. Palgrave, 2013.
- [17] 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) / Eds Radomir V. Bolgov, Nikolai V. Borisov, Leonid V. Smorgunov, Irina I. Tolstikova, Victor P. Zakharov. ACM International Conference Proceeding Series. N.Y.: ACM Press, 2017. ix, 291 p. P. 69-74. DOI: 10.1145/3143699.3143739
- [18] Sinn D. Impact of digital archival collections on historical research // J. Am. Soc. Inf. Sci. 2012. Vol. 63. P. 1521–1537. DOI: 10.1002/asi.22650
- [19] Sinn D., Soares N. Historians' use of digital archival collections: The web, historical scholarship, and archival research // J Assn Inf Sci Tec. 2014. Vol. 65. P. 1794–1809. DOI: 10.1002/asi.23091
- [20] Toon E., Timmermann C., Worboys M. Text-Mining and the History of Medicine: Big Data, Big Questions? // Medical History. 2016. Vol. 60 (2). P. 294–296. DOI: 10.1017/mdh.2016.18
- [21] Vine R. "Google Scholar" // Journal of the Medical Library Association. 2006. Vol. 94 (1). P. 97–99. PMC 1324783

Methods of Digital Data Quantitative Processing in Scientific Heritage Studies (A Case of Georgy F. Gause's scientific ideas)

D.E. Prokudin, G.S. Levit

Abstract — The methods of newly appeared field of Digital Humanities are getting more and more popular in the history of science. These methods influence the establishing of digital information resources accumulating and aggregating huge amount of metadata and full text publications. In a previous publication, we used an example of a Russian evolutionary biologist and ecologist Georgy F. Gause to preliminary estimate the potential of digital resources for the science studies including history of science. We selected prioritized resources to be used in further research.

The present study explores the methods of selection, processing and quantitative analysis of data extracted from digital information resources. Our concentration is on the digital information resources offering structured metadata. We selected processed and visualized extracted metadata. Based on the analysis of the achieved results we concluded on the potential of using digital information resources in the history of science. Besides, the possibility of extracting unstructured metadata has been explored.

Keywords — Scientific information, Digital Information Resources, Extraction of Structured Data, Methods for Quantitative Data Processing, Digital Scientific Heritage, Georgy F. Gause

REFERENCES

- [1] Borisov N.V. Mul'timedia-rekonstrukcija teatral'nogo sobytija. Prem'era spektaklja «Chajka» na Aleksandrinskoj scene 17 oktjabrja 1896 g. / Borisov N.V., Nikitin A.V., Smolin A.A., Trushin V.A., Chepurov A.A., Chepurova O.A. // Kul'tura i tehnologii. 2016. Tom 1. Vyp. 1. S. 15-23. URL: <http://cat.ifmo.ru/ru/2016/v1-i1/65>
- [2] Borodkin L. I. Virtual'nye rekonstrukcii istoricheskij gorodskoj zastrojki: novye tehnologii prezentacii muzejnyh jekspozicij // Rol' muzeev v informacionnom obespečenii istoricheskij nauki: sbornik statej. M.: Jetera, 2015. S. 386–395.
- [3] Borodkin L.I. Virtual'naja rekonstrukcija monastyrskih kompleksov Moskvy: proekty v kontekste Digital Humanities // Vestnik Permskogo universiteta. Serija: Istorija]. 2014. # 3 (26). S. 107-112.
- [4] Gall Ja.M. Georgij Francevich Gauze (1910-1986). SPb: Nestor-Istorija, 2012. 234 s.
- [5] Garskova I.M. Informacionnoe obespečenje gumanitarnyh issledovanij v cifrovuju jepohu: modeli formirovanija i razvitija // Vestn. Perm. un-ta. Ser. Istorija. 2014. #3 (26). S. 76-86.
- [6] Donina O.V. Primenenie metodov Data Mining dlja reshenija lingvisticheskij zadach // Vestnik Voronezhskogo gosudarstvennogo universiteta. Serija: Sistemnyj analiz i informacionnye tehnologii. 2017. # 1. S. 154-160.
- [7] Kukovjakin A.V., Ljapin S.H. Prostranstvo Lomonosova: opyt funkcional'noj integracii virtual'nogo atlasa i polnotekstovoj biblioteki // Kul'tura i tehnologii. 2017. Tom 2. Vyp. 1. S. 6-12. URL: <http://cat.ifmo.ru/ru/2017/v2-i1/98>
- [8] Majer-Shenberger V., Kuk'er K. Bol'shie dannye. Revoljucija, kotoraja izmenit to, kak my zhivem, rabotaem i myslim. M.: Mann, Ivanov i Ferber. 2014. 240 s.
- [9] Prokudin D.E., Levit G.S. Metody otbora cifrovijh informacionnyh resursov na primere issledovanij vlijanija nauchnyh idej G.F. Gauze na razvitie nauki // Nauchnyj servis v seti Internet: trudy XIX Vserossijskoj nauchnoj konferencii (18 – 23 sentjabrja 2017 g., g. Novorossijsk). — M.: IPM im. M. V. Keldysha, 2017. S. 389-399. DOI: 10.20948/abrau-2017-75
- [10] Borisov N.V., Volkov O.G., Nikitina L.L., Nikolaev A.O., Smolin A.A., Stolyarov D.A. Application of Video 360deg Technology for the Presentation of the Solovetsky Monastery Cultural Heritage // Kul'tura i tehnologii. 2016. Tom 1. Vyp. 1. S. 24-31 URL: <http://cat.ifmo.ru/en/2016/v1-i1/88>
- [11] Damerow J., Peirson B.R.E., Laubichler M.D. The Giles Ecosystem – Storage, Text Extraction, and OCR of Documents // Journal of Open Research Software. 2017. 5 (1). P. 26. DOI: 10.5334/jors.164
- [12] Huijnen P., Laan F., de Rijke M., Pieters T. A Digital Humanities Approach to the History of Science Eugenics Revisited in Hidden Debates by Means of Semantic Text Mining // SOCIAL INFORMATICS, SOCINFO 2013, Lecture Notes in Computer Science. 2014. Vol. 8359. Pp. 70 – 84.
- [13] Jockers M.L., Underwood T. Text-Mining the Humanities, in A New Companion to Digital Humanities /eds S. Schreibman, R. Siemens and J. Unsworth. John Wiley & Sons, Ltd, Chichester, 2015. DOI: 10.1002/9781118680605.ch20
- [14] Kaplan F. A map for big data research in digital humanities // Front. Digit. Humanit. 2015. Vol. 2 (1). DOI: 10.3389/fdigh.2015.00001
- [15] Kolchinsky E.I., Hossfeld U., Kutschera U., Levit G.S. The revival of Lysenkoism in Russia and epigenetics // Current Biology. 2017. Vol. 27 (19). P. 1042-1047. DOI: 10.1016/j.cub.2017.07.045
- [16] Manning P. Big data in history. Palgrave, 2013.
- [17] 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) / Eds Radomir V. Bolgov, Nikolai V. Borisov, Leonid V. Smorgunov, Irina I. Tolstikova, Victor P. Zakharov. ACM International Conference Proceeding Series. N.Y.: ACM Press, 2017. ix, 291 p. P. 69-74. DOI: 10.1145/3143699.3143739
- [18] Sinn D. Impact of digital archival collections on historical research // J. Am. Soc. Inf. Sci. 2012. Vol. 63. P. 1521–1537. DOI: 10.1002/asi.22650
- [19] Sinn D., Soares N. Historians' use of digital archival collections: The web, historical scholarship, and archival research // J Assn Inf Sci Tec. 2014. Vol. 65. P. 1794–1809. DOI: 10.1002/asi.23091
- [20] Toon E., Timmermann C., Worboys M. Text-Mining and the History of Medicine: Big Data, Big Questions? // Medical History. 2016. Vol. 60 (2). P. 294–296. DOI: 10.1017/mdh.2016.18
- [21] Vine R. "Google Scholar" // Journal of the Medical Library Association. 2006. Vol. 94 (1). P. 97–99. PMC 1324783