

Историк в потоке информации: особенности работы с международными библиографическими базами данных

Научный руководитель – Полежаева Татьяна Владимировна

Гуменюк Кристина Алексеевна

Студент (бакалавр)

Национальный исследовательский Томский государственный университет, Исторический факультет, Томск, Россия

E-mail: kristinagumenuk553@gmail.com

Сегодня цифровая революция полностью изменила методологию исследований в гуманитарных науках, предоставив новые инструменты и возможности для более глубокого и всестороннего изучения. Действительно, сейчас стало возможным работать с большими данными. Все чаще в анализе историографии и становления научных направлений применяются библиометрические методы, методы сетевого анализа, способные выявить не только передовых исследователей, но и ряд проблем и нерешенных вопросов в определенной тематической категории [1, 2, 3, 4, 5, 6]. При этом качество полученных результатов напрямую зависит от массива анализируемых данных. Цель данной работы – показать на примере темы старообрядчества, с какими сложностями может столкнуться исследователь-историк при анализе историографии методами библиометрии. В работе анализируются публикации, вышедшие в период 2010-2024 гг. по теме старообрядчества и проиндексированные международными базами данных цитирования (Scopus, OpenAlex и Dimensions [10,8, 9,]).

База данных (БД) Scopus (разработчик компания Elsevier) начала свою работу в марте 2004 года. Сегодня это одна из крупных баз данных, содержит более 94 млн. записей о научных публикациях, в том числе социогуманитарного направления. Кроме того, Scopus является одной из БД, по которой оценивается в РФ публикационная активность ученых. На данный момент несмотря на санкции, доступ к Scopus в РФ открыт в программном интерфейсе API.

База OpenAlex – это новый бесплатный сервис компании OurResearch запущенный в 2022 году. Тем не менее, на сегодняшний день БД агрегирует свыше 248 млн. записей, однако данный ресурс не соревнуется с базой Scopus из-за отсутствия строгих процедур отбора. В виду большого объема данных OpenAlex активно используется наукометристами со всего мира.

Еще одна база данных Dimensions, запущенная в 2018 году компанией Digital Science, сегодня собрала уже свыше 140 млн. записей о публикациях. Ее особенностями является работа с открытыми источниками данных, а также охват информации о грантах, стратегических документах, наборах данных и патентов.

На основе выгруженных данных из указанных БД была предпринята попытка создать облако ключевых слов с помощью инструмента VoSViewer [11] для сравнения результатов выгрузки.

В результате работы были сформулированы следующие критерии для сравнения указанных БД: объем данных, полученных по поисковому запросу “old believer”; объем данных, доступных для выгрузки; качество метаданных и соответствие ключевых слов предметной области.

API Scopus выдал 738 публикаций с указанным ключевым словом. При лимитах выгрузки в 5 000 публикаций это оказалось вполне достаточным количеством для последующей визуализации в VoSViewer. Однако API Scopus выгружает неполные данные: отсутствуют ключевые слова, аннотации, соавторы. Кроме того, в получившейся выгрузке

из 738 публикаций релевантными запросу оказались 361 (49% выгрузки). Данное обстоятельство стало результатом ряда региональных и технических особенностей индексации данных, таких как качество предоставляемых издателями метаданных, тематическая рубрикация Scopus и др.

За обозначенный период в OpenAlex было найдено 45 580 публикаций, что позволило при лимитах в 100 000 записей построить карту ключевых слов с помощью VosViewer. Полученные данные содержали 11 612 ключевых слов, было выявлено 8 крупных тематических кластеров. Карту ключевых слов можно рассматривать как удобный инструмент анализа научных направлений, поскольку она позволяет увидеть неочевидные межпредметные связи. В качестве ограничений работы с OpenAlex (помимо ограничений в 50 000 результатов, анализируемых в VosViewer и 100 000 доступных для выгрузки) является качество предоставляемых метаданных.

Результат в базе Dimensions составил 257 799 публикаций, однако в виду платного доступа к расширенным данным, выгрузка из бесплатного интерфейса составляет не более 2 500 публикаций (менее 1% от результатов поиска). Это ограничение не позволило перейти к следующему шагу по визуализации облака ключевых слов.

Таким образом, на примере темы старообрядчества было показано, что историк, работая с международными базами данных, может столкнуться с рядом трудностей: ограничения в выгрузке больших объемов данных, платный доступ к расширенным сервисам, некорректность индексации российских журналов, метаданных публикаций, отсутствие контекстного поиска. Кроме того, такая особенность гуманитарного знания как существенная доля книжных публикаций [7] требует тщательного подбора БД для анализа публикаций по теме, т.к. существующие крупные информационные агрегаторы ориентированы больше на журнальные публикации. На сегодняшний день мы можем утверждать, что не из специализированных тематических баз для историков наибольший интерес представляет OpenAlex.

Источники и литература

- 1) 1. Бородкин Л. И. От библиографии к библиометрии: к вопросу об оценке публикационной активности историков // Роль библиографии в информационном обеспечении исторической науки: сб. ст. М., 2018. С. 347-353.
- 2) 2. Виноградова Т. В. Библиометрия и социогуманитарные науки не совместимы? // Наукоедческие исследования. 2016. № 2016. С. 90–106.
- 3) 3. Гарскова И. М. Методы анализа тематической библиографии и историографии // Роль библиографии в информационном обеспечении исторической науки : сб. ст. М., 2018. С. 326–347.
- 4) 4. Гарскова И. М. Сетевой анализ историографии: динамика формирования региональных центров исторической информатики // Историческая информатика. 2017. № 3. С. 94–115. DOI: 10.7256/2585-7797.2017.3.24566.
- 5) 5. Гарскова И.М. Сетевой анализ историографии: динамика формирования межрегиональной компоненты сети АИК // Историческая информатика. 2017. № 4. С. 112 - 129. DOI: 10.7256/2585-7797.2017.4.25078.
- 6) 6. Сильнича А. В., Гуменюк Л. Г. Российские геополитические исследования через призму библиометрии // Региональные исследования. 2020. № 1(67). С. 76–88. DOI 10.5922/1994-5280-2020-1-7 – EDN GWAGUA.
- 7) 7. Москалева О.В. Научные публикации как средство анализа и оценки научной деятельности // Руководство по наукометрии: индикаторы развития науки и технологии. ЕКБ : Изд-во Уральского ун-та, 2021. С. 221-229.

- 8) 8. Dimensions: <https://www.dimensions.ai/>
- 9) 9. OpenAlex: <https://openalex.org/>
- 10) 10. Scopus: <https://www.scopus.com/home.uri>
- 11) 11. VoSViewer: <https://www.vosviewer.com/>