

ПРИМЕНЕНИЕ ВЕБ-АНАЛИТИЧЕСКОГО ИНСТРУМЕНТАРИЯ GOOGLE TRENDS В ИССЛЕДОВАНИИ СИСТЕМЫ ОБЯЗАТЕЛЬНОГО ЭКЗЕМПЛЯРА ГЕРМАНИИ

© С. В. Соколов, 2019

*Институт научной информации по общественным наукам РАН,
Москва, Россия; e-mail: beholder73@gmail.com*

Статья посвящена веб-статистическому сервису Google Trends и его использованию в рамках исследования системы обязательного экземпляра в Германии. Подробно разобран процесс составления семантического словаря для подбора ключевых слов в поисковом запросе. Показана поэтапная методика использования сервиса Google Trends на примере анализа популярности терминов «обязательный экземпляр», «сетевые публикации», «Немецкая национальная библиотека», «диссертации».

Ключевые слова: веб-аналитика, Google Trends, обязательный экземпляр, семантический словарь, поисковый запрос

Для цитирования: Соколов С. В. Применение веб-аналитического инструментария Google Trends в исследовании системы обязательного экземпляра Германии // Библиосфера. 2019. № 1. С. 11–17. DOI: 10.20913/1815-3186-2019-1-11-17.

Applying Google Trends web-analytic tool to study the German legal deposit copy system

S. V. Sokolov

Institute of Scientific Information on Social Sciences, Moscow, Russia; e-mail: beholder73@gmail.com

The article «Applying Google Trends to study the German legal deposit copy system» discusses the use of web tools to investigate current library science problems. Using web-based statistical method the author searches the following issues:

- the dynamics of interest in the subject of a legal deposit copy from the date of adoption of the Law on the German National Library (2006) to nowadays;
- relations of the public interest peak changes in this topic to certain phenomena in the social and cultural life of Germany;
- the federal dimension of these issues when comparing interest to the topic in different regions of Germany;
- the public opinion on the popularity of legal copy among traditional and electronic sources.

The article is divided into four parts. The first one sets the work objective and main tasks, gives a general description of the chosen research method. The second part deals with the process of creating a semantic dictionary; analyzes traditional and electronic sources of synonymic dictionaries. It describes the strengths of such an online language matching service as semager.de. The third part dissects a group of keywords related to the topic of a legal deposit copy along with the most interesting and problematic, from the point of the author's view, additional keywords such as the German National Library, network publications, and dissertations. Using web statistical tools the paper shows that the most intense issues regarding the legal deposit, the problems of the German National Library and online publications were raised in the lands of West Germany. Developing the legal deposit copy system will go, first of all, through online publications, greater cooperation with academic and scholar libraries; open access of scientific data and publications related to dissertations and theses of West German lands' universities. The fourth part presents main conclusions and substantiates the method significance for library and sociological research.

Keywords: web analytics, Google Trends, legal deposit copy, semantic dictionary, search query

Citation: Sokolov S. V. Applying Google Trends web-analytic tool to study the German legal deposit copy system. *Bibliosphere*. 2019. № 1. P. 11–17. DOI: 10.20913/1815-3186-2019-1-11-17.

Развитие системы обязательного экземпляра в Германии является актуальной библиотечной, социологической, правовой и экономической проблемой, затрагивающей

целый ряд вопросов библиотечной и информационной политики. Передача экземпляра различных видов тиражированных документов в библиографирующие учреждения, крупные

библиотеки и информационные центры в порядке, установленном государственным законодательством, библиографический учет издаваемой продукции и пополнение фондов крупнейших библиотек в Германии имеет свои региональные особенности. Изучение опыта зарубежных стран по формированию системы обязательного экземпляра, в том числе и электронными изданиями, необходимо, как подчеркивают многие ведущие отечественные исследователи [12], для выработки эффективных решений при совершенствовании российского законодательства об обязательном экземпляре документов. Исследование проблемы обязательного экземпляра в Германии, особенно при рассмотрении вопросов, связанных с изучением общественного интереса к указанной проблематике, наиболее результативно с использованием веб-статистического сервиса Google Trends.

При исследовании популярности пользовательских запросов в немецком открытом веб-относительно темы обязательного экземпляра в Германии были поставлены следующие задачи:

- определение временных пиков популярности группы терминов, относящихся к выбранной тематике;
- определение колебаний популярности этого термина в различных регионах Германии;
- определение национальных особенностей восприятия этого термина по корреляции с другими терминами библиотечной тематики.

Первым шагом в проведении исследования поисковых запросов пользователей является составление семантического словаря поиска. Семантический (идеографический) словарь включает все возможные слова, которыми можно выразить искомое ключевое слово. Его составление необходимо потому, что пользователи, задавая поиск в интернете по определенной проблеме (так называемому интену запроса), могут применить не одно ключевое слово, а целый ряд близких по смыслу терминов. Google Trends обладает технологией контекстно-зависимого поиска только при составлении поисковых запросов на английском языке. Тем не менее даже для максимально автоматизированного поиска в Google существуют специалисты-«ассессоры», которые корректируют работу поисковика, внося необходимые поправки в составление отдельных англоязычных поисковых запросов. Для русского и некоторых европейских языков, например немецкого, необходимо введение дополнительных ключевых слов для наиболее полного охвата поля поиска. Семантическое поле – самая крупная смысловая парадигма, объединяющая слова различных частей речи, значения которых имеют один общий семантический признак. Использование в веб-статистическом исследовании всего се-

мантического поля запроса невозможно, поскольку оно объединяет несколько сотен терминов. Для решения поставленных задач достаточно выделить совокупность наиболее близких целевых поисковых запросов – семантическое ядро поиска. Остальные слова, соответствующие фразам с низкой частотой употребления, так называемые низкочастотные запросы (НЧ-запросы), и в особенности микро-низкочастотные (МНЧ-запросы), образуют так называемый длинный хвост поисковых запросов. МНЧ-запросы, по современным исследованиям, составляют до 33% всех запросов пользователей и игнорировать их было бы нежелательно, однако сервис Google Trends позволяет одновременно искать не более чем по пяти поисковым запросам. Официальные руководства по сервису Google Trends утверждают, что с помощью знаков препинания сервис позволяет фильтровать результаты поиска. Например, знак «+», соответствующий команде «или», якобы гарантирует одновременный поиск по нескольким взаимозаменяемым понятиям, но практика показывает, что при использовании логических операторов поиск производится некорректно. Выделение семантического ядра – пяти наиболее близких семантических единиц – является важнейшим ограничением и важным условием постановки исследовательской задачи.

Рассмотрим основные источники составления семантического словаря для понятия «обязательная поставка документов» (Pflichtablieferung).

1. Синонимические словари и тезаурусы

Исследуя перечень ключевых слов, максимально близких к понятию «поставка обязательного экземпляра в библиотеки» – Pflichtablieferung, Pflichtexemplar и т. д., автор столкнулся с тем, что ни один из современных немецких тезаурусов или синонимических словарей не дал списка хотя бы из 5 слов, перекрывающих искомую тематику. Использовались словари и тезаурусы: Openthesaurus [7]; Synonyme.woxikon [9]; Duden. Das Synonymwörterbuch. Ein Wörterbuch sinnverwandter Wörter, 2007 [4]; Duden – Das Wörterbuch der Synonyme: 100.000 Synonyme für Alltag und Beruf Bibliographisches Institut GmbH, 3. Auflage 2016 [3]; Dornseiff, Franz. Der deutsche Wortschatz nach Sachgruppen, Degruyter, 2004 [2].

2. Программы для парсинга сайтов и подбора ключевых слов

Очевидным их минусом является ориентированность на англоязычные или русскоязычные запросы пользователей, а также высокая стоимость. Наиболее популярными являются сервисы Rash-Analytics [8] (999 р. в месяц) и Just-Magic [6] (от 1000 р. в месяц).

3. Автоматизированные онлайн-сервисы, например «Планировщик ключевых слов Google»

Планировщик Google требует наличия собственного сайта и позволяет сравнить частоту использования ключевых слов, введенных пользователем самостоятельно. Этот сервис может помочь исследователю отсеять средне- и микро-низкочастотные запросы, но только на конечном этапе формирования семантического ядра поиска (рис. 1).

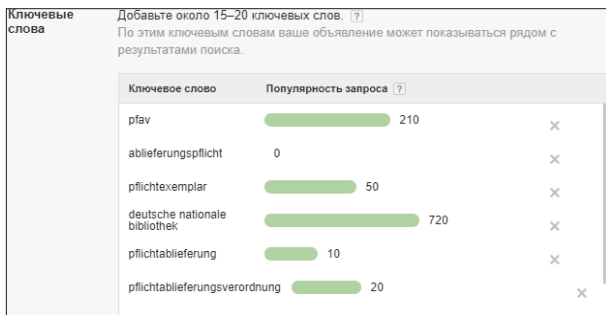


Рис. 1. Планировщик ключевых слов Google на сайте автора «Libseminion»

Fig. 1. The Google keyword planner on the author website «Libseminion»

Наиболее эффективным источником составления семантического словаря для поискового запроса по германоязычному интернету оказалась немецкая интеллектуальная база семантических соответствий *semager.de*, созданная в апреле 2011 г. В настоящее время ее рейтинг (Google PR 5) оценивается среди других семантических систем на достаточно высоком уровне [1]. На основании этого сервиса был составлен семантический словарь понятия «обязательная поставка документов» (*Pflichtablieferung*) [5]. В таблице 1 представлены ключевые слова к понятию «обязательная поставка документов» (*Pflichtablieferung*), найденные семантическим сервисом *semager*.

Таблица 1

Словарь сочетаний к понятию «Обязательная поставка документов» (*Pflichtablieferung*) по сервису *semager*

Table 1

A dictionary of combinations to the concept of «documents obligatory delivery» (*Pflichtablieferung*) on *semager* service

| Ключевое слово | Корреляция (в %) |
|---|------------------|
| Medienwerken – произведения | 85 |
| Ablieferung – поставка | 74 |
| Netzpublikationen – сетевые публикации | 74 |
| PVLAF – аббревиатура положения о поставке обязательного экземпляра | 72 |
| Ablieferungspflicht – обязательство поставлять экземпляр | 72 |
| Pflichtablieferungsverordnung – положение о поставке обязательного экземпляра | 66 |
| Pflichtexemplar – обязательный экземпляр | 57 |

Процент корреляции тождественных терминов может быть визуализирован в виде сетевой модели *semager*. Близость вершин графа к его центру показывает больший процент родства терминов. Близость вершин между собой показывает процент корреляции между ними (рис. 2).

С учетом того, что понятия, имеющие наибольшие значения в получившемся списке, слишком абстрактны и перекрывают не только исследуемое понятие, но и смежные явления, соответственно дают избыточные данные для исследования, временно исключим из поискового запроса первые три пункта (*Medienwerken* – произведения; *Ablieferung* – доставка; *Netzpublikationen* – сетевые публикации). Также в поисковый запрос не были включены излишне узкие категории, попавшие в топ-список, например фамилия Эрика Штейнхауэра (верхняя точка графа). Этот известный немецкий исследователь активно пропагандирует политику открытых данных и выступает за так называемые гибридные публикации в платном печатном и в бесплатном электронном виде. Нужно признать тот факт, что сервис оценил значимость самого термина *Pflichtexemplar* только на 12 позиции, связав его преимущественно с печатными изданиями (см. левую часть рисунка 2). Активное обсуждение в современном медийном поле Германии проблем, связанных именно с электронными экземплярами документов, определило снижение популярности терминов печатных изданий. Итак, автоматически сгруппированные в единый поисковый запрос слова по теме: *Pflichtablieferung*, *pflav*, *Ablieferungspflicht*, *Pflichtablieferungsverordnung*, *Pflichtexemplar* – образуют семантическое ядро нашего запроса.

Поиск будет производиться по всем землям Германии в период с 2004 г. (максимальная

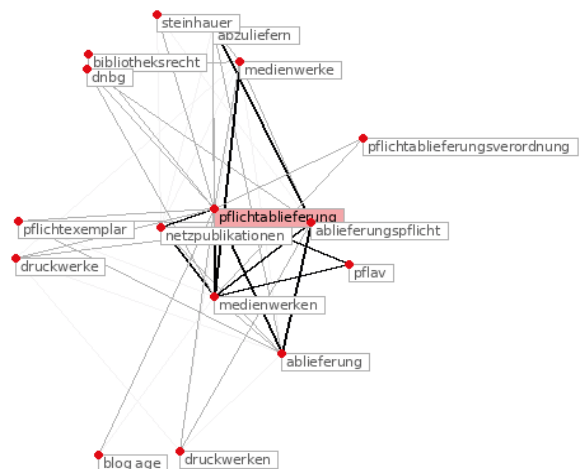


Рис. 2. Граф ключевых слов, связанных с понятием «Обязательная поставка документов» (*Pflichtablieferung*) в интернете по сервису *semager.de* [5]

Fig. 2. Keywords' graph related to the concept of «documents obligatory delivery» (*Pflichtablieferung*) on the Internet service *semager.de* [5]

нижняя граница диапазона в поиске Google Trends, альтернативная глубина выборки ранних дат возможна только за последние 5 лет) по 4 октября 2018 г. по пяти отобранным при помощи сервиса semager ключевым словам, связанным с обязательным экземпляром.

Соотношение интереса жителей Германии в сравнении всех пяти ключевых слов было по результатам Google Trends неоднозначным. Процент приоритетности того или иного ключевого слова показан на рисунке 3.

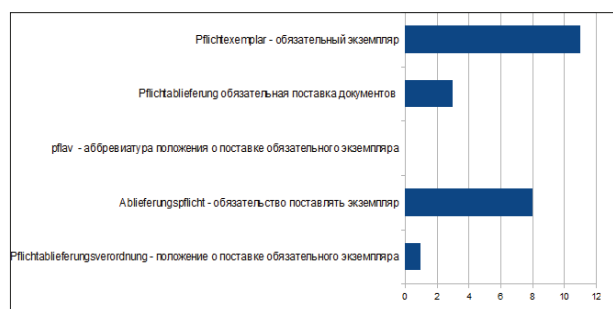


Рис. 3. Ранжирование в Google Trends по теме «Обязательный экземпляр»

Fig. 3. Ranking in Google Trends on the «legal deposit copy» topic

Общий итоговый график представлен на рисунке 4.

По горизонтали представлена временная шкала. По вертикали – уровень интереса к тому или иному запросу. Алгоритмы Google определяют точку на рисунке 4 за выбранный период, когда запрос был наиболее популярен, и принимают ее за 100. Все остальные точки на графике определяются в процентном отношении к максимуму.

На временной шкале мы видим угасание общественного интереса к теме «Обязательный экземпляр» по землям Германии. Наибольший пик интереса по тематике обязательного экзем-

пляра зафиксирован в июне 2004 г. В июне 2004 г. в Геттингенском университете проводилась федеральная конференция по электронным документам и сотрудничеству библиотек и университетов в рамках сети NESTOR. Эта сеть в Германии была создана в 2003 г. по развитию взаимодействия между федеральным центром и немецкими землями в области долговременного электронного архивирования документов библиотек, музеев и архивов. Пользуясь временной шкалой Google Trends по интенсивности общественного интереса, можно устанавливать реперные точки наиболее важных событий в истории исследуемых явлений и процессов.

Региональный интерес сразу по всем пяти ключевым словам на протяжении всего рассматриваемого периода система не смогла определить из-за недостатка данных. По отдельным ключевым словам были отобраны регионы, интерес в которых был наиболее значителен (> 50%). Геоориентирование запроса позволило определить регионы Германии, наиболее интересующиеся проблематикой обязательного экземпляра (табл. 2).

Исследование рейтинга наиболее семантически близких слов не даст представления о значимости данной проблемы в более широких проблемных полях немецкой общественной, правовой, культурной и библиотечной жизни. Для выявления позиционирования данной проблемы в более общих категориях введем дополнительные ключевые слова для поиска.

Добавив ключевое слово «Немецкая национальная библиотека», которое будет служить верхней границей пользовательского интереса, получим в сервисе Google Trends статистику изменения интереса в немецких землях к проблемам обязательного экземпляра с 2006 г. по настоящее время (рис. 5).

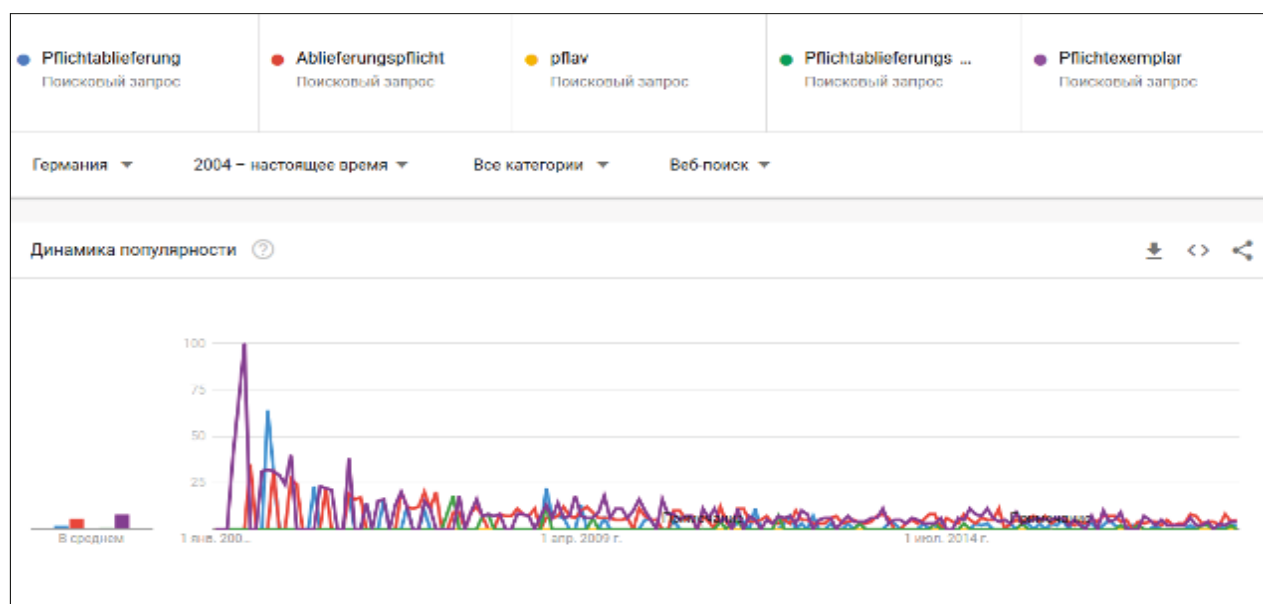


Рис. 4. Динамика популярности темы «Обязательный экземпляр» в Google Trends [11]

Fig. 4. The popularity dynamics of the «legal deposit copy» topic in Google Trends [11]

Таблица 2

Геоориентирование поискового запроса по землям Германии

Table 2

Geo-targeting a search query on German lands

| Термины | Регионы | Уровень интереса к теме (в %) |
|---|------------------------------------|-------------------------------|
| Pflichtexemplar – обязательный экземпляр | Нижняя Саксония | 100 |
| Pflichtablieferung – обязательная поставка документов | Северный Рейн-Вестфалия Бавария | 64 100 |
| Pflav – аббревиатура положения о поставке обязательного экземпляра | – | – |
| Ablieferungspflicht – обязательство поставлять экземпляры | – | – |
| Pflichtablieferungsverordnung – положение о поставке обязательного экземпляра | Северный Рейн-Вестфалия | 100 |

По итоговой таблице заметно, что пользователи интернета, составившие свои запросы с территории бывшей ГДР (за исключением Берлина и Саксонии), по сравнению с пользователями из Западной Германии, практически не интересуются как проблемами обязательного экземпляра, так и проблематикой Немецкой национальной библиотеки (ННБ). Наибольшую активность относительно максимума показало количество запросов по блоку ключевых слов, связанных с темой «Обязательный экземпляр», из немецких земель Нижняя Саксония, Северный Рейн-Вестфалия и Бавария (см. табл. 2) – это земли с наибольшим количеством крупных издательств, с наиболее развитыми научными организациями и библиотечными системами.

Также интересным нам представляется сравнить блок ключевых слов темы «Обязательный экземпляр» с понятием «сетевые публикации». Заменяв аббревиатуру pflav, практически не представленную в тренде, на Netzpublikationen, можно получить очень интересную статистику за последние пять лет (рис. 6).

При всей популярности тренда «сетевые публикации» в российском библиотечном сообществе, его доля среди поисковых запросов немецких пользователей интернета не поднимается выше 30% относительно популярности терминов группы «обязательный экземпляр». В большинстве случаев видна стопроцентная корреляция терминов группы «обязательный экземпляр» и «сетевые публикации», что позволяет говорить о значительном охвате сферы электронных публикаций системой обязательного экземпляра. Отчеты ННБ по комплектованию различными типами изданий подтверждают эту статистику. Почти 100% электронных поступлений в ННБ идет по линии обязательного экземпляра, в то время как количество печатной литературы, поступившей как обязательный экземпляр, не поднимается выше 40% [10].

Особенности системы обязательного экземпляра Германии относительно диссертационных работ и иных аттестационных произведений высшей школы отразили фундаментальные особенности развития системы национального библиографического учета, особенности развития

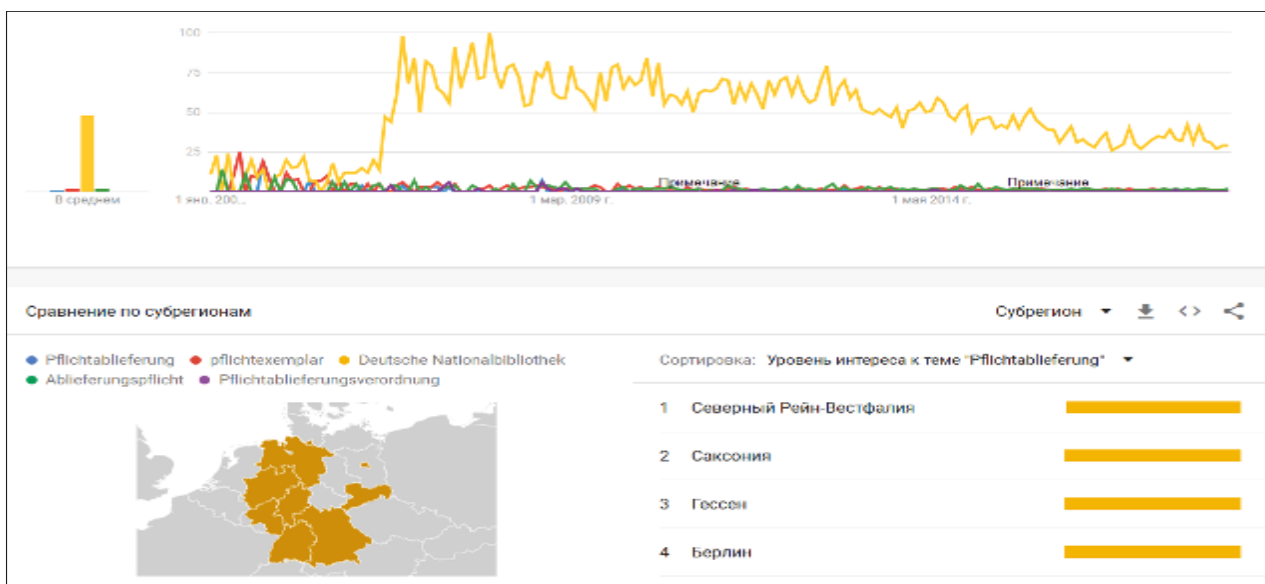


Рис. 5. Анализ географии и частотности интернет-запросов по тематике «Обязательный экземпляр – Немецкая национальная библиотека» [11]

Fig. 5. The analysis of geography and frequency of Internet requests «legal deposit copy – German National Library» [11]

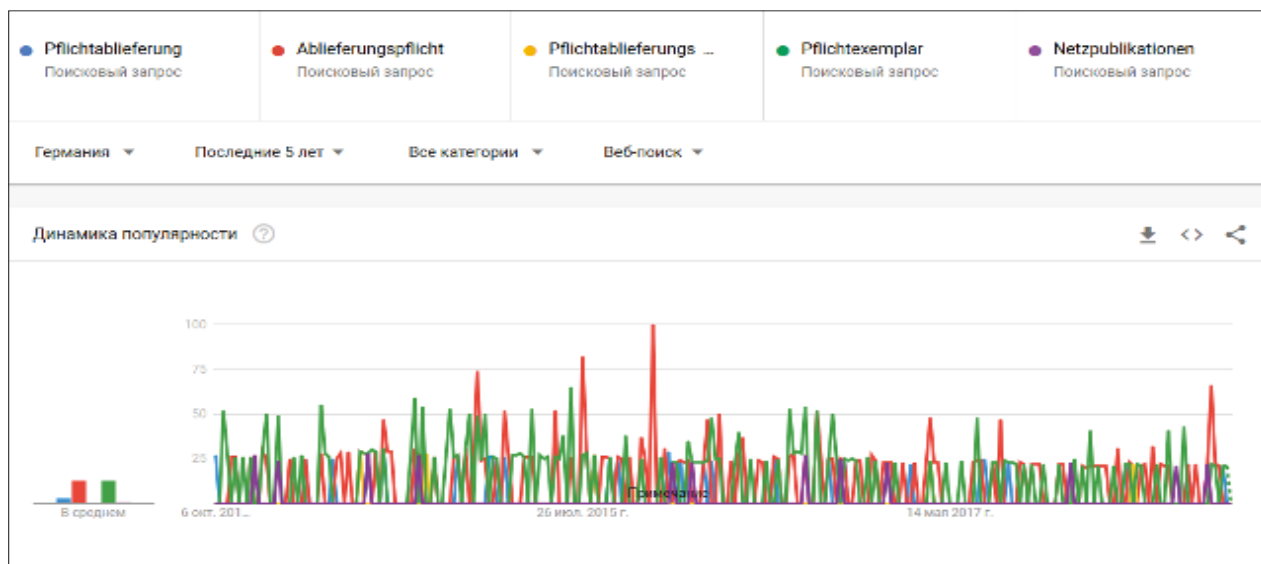


Рис. 6. Анализ географии и частотности интернет-запросов по тематике «Обязательный экземпляр – сетевые публикации» [11]

Fig. 6. The analysis of geography and frequency of Internet requests «legal deposit copy – online publications» [11]

науки и научных учреждений в тесной связи с системой высшего образования. Проиллюстрируем особенности обязательного экземпляра Германии по отношению к диссертационным работам данными веб-статистики. Обращаясь к описанному инструментарию Google Trends, вычислим географию электронного публичного пространства связей этих двух терминов. Во всех восточно-немецких землях, за исключением Берлина и Саксонии, интерес к проблеме диссертационных исследований, как и к проблеме обязательного экземпляра, практически равен нулю с 2004 по 2018 г. относительно всего количества запросов по Германии по этим тематикам (рис. 7).

Можно сделать предположение, что развитие элементов государственного регулирования библиотечной деятельности в Германии, в том числе и в дальнейшем развитии системы обязательного экземпляра, будет идти прежде всего по линии научных библиотек, научных данных и публикаций, связанных также и с диссертационными и дипломными

работами немецких университетов земель западной части Германии.

Использование статистических веб-сервисов показывает, что интерес к проблемам обязательного экземпляра с принятия закона о ННБ (2006) по настоящее время падает. Взаимосвязь пиковых изменений общественного интереса к указанной проблеме с определенными явлениями в общественно-культурной жизни Германии прослеживается с вводом дополнительных ключевых слов и искомым понятием, таких, например, как «Немецкая национальная библиотека», «сетевые публикации», «диссертации». Новые веб-аналитические инструменты библиотечного и общественно-правового исследования позволяют оценить степень влияния различных, в том числе и скрытых, процессов и явлений на позиционирование проблемы обязательного экземпляра в немецком общественно-правовом пространстве и в конечном итоге позволяют выработать новые подходы к решению проблем обязательного экземпляра в России. ■

Список источников / References

1. Countable Data Brief Semager. URL: <https://www.easycounter.com/report/semager.de> (дата обращения: 04.10.2018).
2. Dornseiff F. Der deutsche Wortschatz nach Sachgruppen, Degruyter, 2012. URL: <https://books.google.ru/books?id=GT0jAAAAQBAJ&printsec=frontcover&dq=6.%09Dornseiff,+Franz.+Der+deutsche+Wortschatz+nach+Sachgruppen,+Degruyter,+2004&hl=de&sa=X&ved=0ahUKewia79mimI3eAhVmwsKHajWAIQQ6AEIKTAA#v=onepage&q&f=false> (accessed 04.10.2018).
3. Duden – Das Wörterbuch der Synonyme: 100.000 Synonyme für Alltag und Beruf Bibliographisches Institut GmbH, 2016. URL: <https://books>.

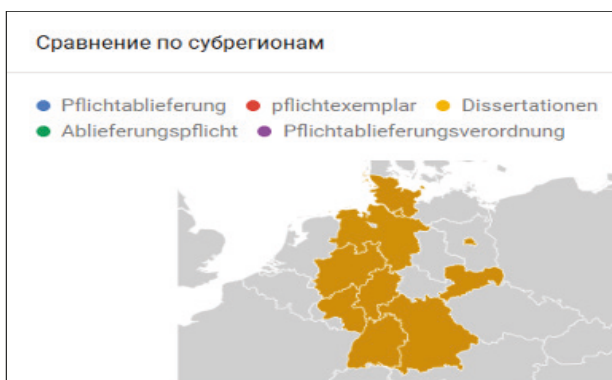


Рис. 7. Анализ географии и частотности интернет-запросов по тематике «Обязательный экземпляр – диссертации» [11]

Fig. 7. The analysis of geography and frequency of Internet requests «legal deposit copy – dissertation» [11]

google.ru/books?id=7jO4CwAAQBAJ&printsec=frontcover&dq=Duden+-+Das+Wörterbuch+der+Synonyme:+100.000+Synonyme+für+Alltag+und+Beruf+Bibliographisches &hl=de&sa=X&ved=0ahUKEwjh9Y_Wl43eAhUwqYsKHUIMDWIQ6AEIJzAA#v=onepage&q=Duden%20-%20Das%20Wörterbuch%20der%20Synonyme%3A%20100.000%20Synonyme%20für%20Alltag%20und%20Beruf%20Bibliographisches&f=false (accessed 04.10.2018).

4. Duden. Das Synonymwörterbuch. Ein Wörterbuch sinnverwandter Wörter, 2007. URL: <https://books.google.ru/books?id=Pp5iAAAAMAAJ&dq=Duden.+Das+Synonymwörterbuch+.+Ein+Wörterbuch+sinnverwandter+Wörter,+2007&hl=de&sa=X&ved=0ahUKEwjNguWil43eAhXGKywKHRtKABQQ6AEIKDAA> (accessed 04.10.2018).

5. Graph: Pflichtablieferung. URL: <https://www.semager.de/keywords/?q=pflichtablieferung&lang=de> (дата обращения: 04.10.2018).

6. Just-magic поисковый ресурс. URL: https://just-magic.org/serv/ws_freq.php (accessed 04.10.2018).

7. Openthesaurus поисковый ресурс. URL: <https://www.openthesaurus.de> (accessed 04.10.2018).

8. Rush-analytics поисковый ресурс. URL: <https://www.rush-analytics.ru/pricing-plans> (accessed 04.10.2018).

9. Synonyme.woxikon поисковый ресурс. URL: <https://synonyme.woxikon.de> (accessed 04.10.2018).

10. Zwei null eins sechs. Jahrbuch 2016. URL: <https://d-nb.info/1135316724/34> (accessed 04.10.2018).

11. Динамика популярности Google Trends ; Dinamika pop-ulyarnosti Google Trends [The dynamics of Google Trends popularity]. URL: <https://trends.google.ru/trends/explore?date=all&geo=DE&q=Pflichtablieferung,pflichtexemplar,Pflav,Ablieferungspflicht,Pflichtablieferungsverordnung> (accessed 04.10.2018). (In Russ.).

12. Соколова Е. И. Обязательный экземпляр электронных документов: международные приоритеты // Университетская книга. 2013. № 4. С. 78–80 ; Sokolova E. I. Obligatory copy of electronic documents: international priorities. *Universitetskaya kniga*, 2013, 4, 78–80. (In Russ.).

Материал поступил в редакцию 19.10.2018 г.

Сведения об авторе: *Соколов Сергей Валерьевич* – научный сотрудник Фундаментальной библиотеки ИНИОН РАН, ORCID: 0000-0002-2068-6797