

Методика формирования выборки публикаций для подготовки научных обзоров

А. В. Малышева¹, Д. В. Косяков², А. Е. Гуськов³

^{1, 2, 3}*Российский научно-исследовательский институт
экономики, политики и права в научно-технической сфере,
Москва, Российская Федерация*

¹*bag_bala@mail.ru*

²*kosyakov@sciencepulse.ru*

³*guskov.andrey@gmail.com*

Аннотация. Поиск и отбор публикаций для подготовки обзорной статьи являются задачей, решение которой находится на стыке между наукой и искусством. С увеличением публикационного потока исследователям всё сложнее выделять из него релевантные работы, имеющие определённую значимость для научного сообщества. Методологический подход, описанный в статье, позволяет систематизировать подбор публикаций для обзоров, значительно сократив число случайных ошибок и субъективных отклонений. Он упорядочивает процедуры подготовки стартовой выборки, её расширения, усечения и анализа, которые выполняются в несколько итераций, по их результатам формируются финальная выборка, её аналитические характеристики и формальный протокол подготовки.

В статье описаны содержание методики и пример применения для тематики «привлечённый библиотекарь» («embedded librarian»). На этом примере продемонстрированы преимущества предложенного подхода по сравнению с традиционным методом поиска по ключевым словам: возможность использования различных вариантов поисковых стратегий, обеспечение их воспроизводимости и большей полноты результирующей выборки. Описанная методика может применяться в научных и научно-технических библиотеках в целях развития сервисов для информационного сопровождения исследовательской деятельности.

Ключевые слова: обзор, систематический обзор, методика подготовки обзоров, анализ цитирования, поисковые стратегии, ядро из релевантных публикаций, embedded librarian

Для цитирования: Малышева А. В., Косяков Д. В., Гуськов А. Е. Методика формирования выборки публикаций для подготовки научных обзоров / А. В. Малышева, Д. В. Косяков, А. Е. Гуськов // Научные и технические библиотеки. 2022. № 11. С. 56–81. <https://doi.org/10.33186/1027-3689-2022-11-56-81>

UDC 001.818-047.44

<https://doi.org/10.33186/1027-3689-2022-11-56-81>

Methodology for selecting publications for scientific reviews

Alexandra V. Malysheva¹, Denis V. Kosyakov² and Andrey E. Guskov³

^{1, 2, 3}*Russian Research Institute of Economics, Politics and Law in Science and Technology, Moscow, Russian Federation*

¹*bag_bala@mail.ru*

²*kosyakov@sciencepulse.ru*

³*guskov.andrey@gmail.com*

Abstract. The task of searching and selecting publications for scientific reviews is located on the intersection of science and art. The increasing publication flow has made it all harder to reveal relevant and important papers for scientific community. The methodological approach being described enables to organize selecting publications for reviews while decreasing accidental mistakes and subjective declinations; it is to structure primary selection with its further expansion, reduction and analysis, that can be accomplished in several iterations resulting in the final selection, its analytical features and formalized protocol. The authors describe the method structure and provide the example for the subject “embedded librarian”. This example demonstrates the advantages of the approach as compared to the traditional search by keywords: various variants of search strategies are applicable; their repeatability and exhaustiveness of resulted selection. The method can be applied in academic and sci-tech libraries to develop their services supporting science and research.

Keywords: review, systematic review, citation analysis, methodology for the preparation of reviews, search strategies, relevant publications core, embedded librarian

Cite: Malysheva A. V., Kosyakov D. V., Guskov A. E. Methodology for selecting publications for scientific reviews / A. V. Malysheva, D. V. Kosyakov, A. E. Guskov // Scientific and technical libraries. 2022. No. 11. P. 56–81. <https://doi.org/10.33186/1027-3689-2022-11-56-81>

Введение

Обзоры являются важной компонентой исследовательского процесса и служат, в том числе, для выявления пробелов в исследованиях. Одна из наиболее часто встречающихся проблем – отбор подходящих публикаций: включение всех достаточно значимых и исключение наименее релевантных или сомнительных. Исследователи Ю. Чжэ и Т. Мензис отмечают, что «относительно простая задача – найти несколько подходящих статей для любого конкретного исследовательского запроса. Но проблема состоит не в том, чтобы найти несколько статей, а в том, чтобы найти наиболее подходящие статьи» [1]. Широко применяемый подход, основанный на выборке публикаций, ограниченной только поиском по ключевым словам, обеспечивает исследователя внушительным массивом статей. «Ручной» уточняющий поиск в нём занимает много времени и не гарантирует того, что будут выбраны наиболее значимые исследования. Набор ключевых слов, характеризующих научную тематику, может иметь региональную/языковую специфику или измениться со временем. Эти факторы приводят к исключению релевантных публикаций из выборки. Существует и проблема нерелевантных публикаций, связанная с совпадением терминологической базы в разных направлениях исследований. Поэтому стратегия поиска – важный элемент большей части литературных обзоров и неотъемлемая часть систематических.

Поиски публикаций могут быть сложными и повторяющимися, особенно когда исследователи обращаются к широко изученной или новой теме. Ознакомление с поисковой стратегией позволяет заинтересованным читателям обзора оценить полноту охвата предметной области и воспроизвести поиск. В настоящее время наибольшее количество статей публикуется в области медицинских и биологических наук, что формирует ярко выраженную потребность в обзорных мате-

риалах. Разработанные в этой сфере подходы к систематическому анализу и метаанализу публикационного потока позволяют получать более надёжные результаты, на основе которых можно делать выводы и принимать решения. В широко распространённых методиках подготовки систематических обзоров требования к отбору публикаций носят скорее рекомендательный характер [2, 3]. Так, использование при подготовке обзоров протоколов PRIZMA¹ (с их последующей регистрацией в PROSPERO² или оценке с помощью инструмента AMSTAR³) позволяет избежать дублирования в предоставлении данных, предвзятого отношения к процессу отбора публикаций, делая его более прозрачным [4, 5]. В этом случае обзоры обычно обеспечивают высокую общую информативность по рассматриваемой теме.

Авторам обзора по протоколу PRISMA для медицинского научного сообщества предлагается предоставить список используемых баз данных (БД) с хронологическим охватом (включая сформулированные поисковые запросы) и указать дату последнего поиска [5]. Одним из главных отличий данного протокола является определение критериев для включения и исключения исследований с подробным описанием. Авторы просят не только архивировать результаты поиска, но и по возможности указывать, подвергались ли поисковые стратегии рецензированию в рамках подготовки систематического обзора. Поскольку процесс поиска в медико-биологических науках достаточно трудоёмкий (из-за большого количества опубликованных научных результатов), в научной среде нередко образуются целые группы исследователей для подготовки таких систематических обзоров [6]. Но и такой подход не гарантирует охвата исследователем всех значимых публикаций. Многое зависит от индивидуальных поисковых стратегий, навыков владения работой с отдельными библиографическими БД и знания предметной области.

Специальное программное обеспечение для подготовки систематических обзоров с максимальным охватом релевантных публикаций,

¹ Preferred Reporting Items for Systematic Reviews and Meta-Analysis.

² International prospective register of systematic reviews.

³ A MeaSurement Tool to Assess systematic Reviews.

разрабатываемое в последние годы, не получило широкого распространения в научной среде.

Так, авторы статьи [7] предлагают обогащать таксономию (Discoverable Taxonomy Set) передовыми технологиями поиска информации с помощью операторов близости и усечения, а также за счёт «самонастройки» методами искусственного интеллекта. В результате пользователи выгружают данные, которые обычно невозможно получить «вручную» из-за необходимости обрабатывать очень большие объёмы текстов. После ввода пользователем определённых поисковых запросов служба автоматически вызывает различные API, например ScienceDirect, и начинает запрограммированный процесс поиска. Примечательно, что именно обзорные статьи существенно улучшили данную таксономию.

Для поиска публикаций при подготовке научных обзоров исследователи всё чаще применяют наукометрические методы и подходы. Анализ цитирования позволяет выделять исследовательские поля и связанные тематические направления, рассматривать их развитие, делать выводы о смене научных парадигм. Например, использовать при поиске репрезентативного набора документов для систематических обзоров итеративное (каскадное) [8] отслеживание цитирования (автор использует термин «расширение цитирования», имея в виду отслеживание цитирования по цитируемым и цитирующим документам) данных научных публикаций, касающихся области исследований. Оно заключается в том, что, начиная с исходного набора данных или даже с одной исходной статьи, этот набор расширяется за счёт новых публикаций, которые цитируют работы из набора данных или были процитированы в них. Для этого в качестве инструмента используют программу CiteSpace и данные по цитируемости из API Dimensions. Разработчики методики утверждают, что такой подход позволяет пользователям определять порог включения с учётом достаточного для исследователя количества цитирований. Статьи с цитированием ниже порогового значения исключаются из процессов расширения [9].

Действительно, наиболее цитируемые публикации вызывают у исследователя больше доверия, но частота цитирования не всегда значима сама по себе. Соответственно, поиск релевантных публикаций является сложной и неоднозначной наукометрической задачей. Задача

специалистов в области информационно-аналитической деятельности заключается в подготовке репрезентативной выборки, которая, с одной стороны, позволит гарантировать включение в неё всех релевантных публикаций, а с другой – исключит сомнительные и нерепрезентативные результаты. Для исследования тематической области актуален не только поиск по ключевым словам – общий рейтинг тем, журналов и авторов также является важной дополнительной информацией. Какие темы публикуются в журналах с наибольшим импакт-фактором? Какие авторы являются наиболее авторитетными и часто цитируемыми? Ответы на эти вопросы не даёт простой поиск по ключевым словам, но может дать дополнительный анализ. Обзор с тщательной выборкой источников отличается более высоким качеством, он будет чаще использоваться и цитироваться другими исследователями, что, в свою очередь, положительно повлияет на его научную значимость. Методика, представленная в данной статье, предназначена для создания уточнённых выборок релевантных публикаций в определённой тематической области при подготовке научных обзоров.

Методология

Рассматриваемая методика предлагает алгоритмизировать поиск публикаций для подготовки обзоров с помощью наукометрических показателей и экспертной оценки по определённой тематике. Для поиска предлагается использовать любую доступную для пользователей библиографическую базу данных (ББД). Используемая ББД должна предоставлять пользователям возможность оценить такие наукометрические показатели публикаций и отдельных авторов, как цитируемость, индекс Хирша, а также общую публикационную активность по тематике исследования. В научной среде для анализа публикационного потока специалисты используют следующие популярные ББД: Scopus, Web of Science, Dimensions, РИНЦ. Универсальный подход к отбору и анализу публикаций позволяет выбрать для подготовки обзора любую из них.

В табл. 1 представлены основные требования к результатам отбора публикаций при использовании предлагаемой методики. Высокая степень вариативности при работе с выборкой даёт возможность интерпретировать полученные результаты, исходя из целей и задач будущего обзора. Чтобы обеспечить прозрачность процесса, согласованность дей-

ствий экспертов и возможность повторить поиск, на всех этапах применения методики рекомендуется вести протокол отбора публикаций.

Важным моментом при использовании данной методики является выбор пороговых значений для отбора публикаций. Ответ на него неоднозначен и будет варьироваться в зависимости от субъективного мнения эксперта, который определяет релевантное «ядро» публикаций для обзора (финальную выборку). Кроме того, отбору публикаций зачастую сопутствуют некоторые условия поиска источников информации, такие, например, как полноценный охват изучаемой области или, наоборот, ограничение хронологического охвата выборки. При применении методики рекомендуется выбирать пороговое значение того или иного показателя для включения публикаций в соответствии с целевым назначением планируемого обзора и величиной исходной выборки. В финальную выборку входит лишь небольшая доля работ даже из обширной по количеству статей выборки, так как массив публикаций, отобранных в ядро, должен быть проанализирован авторами будущего обзора.

Таблица 1

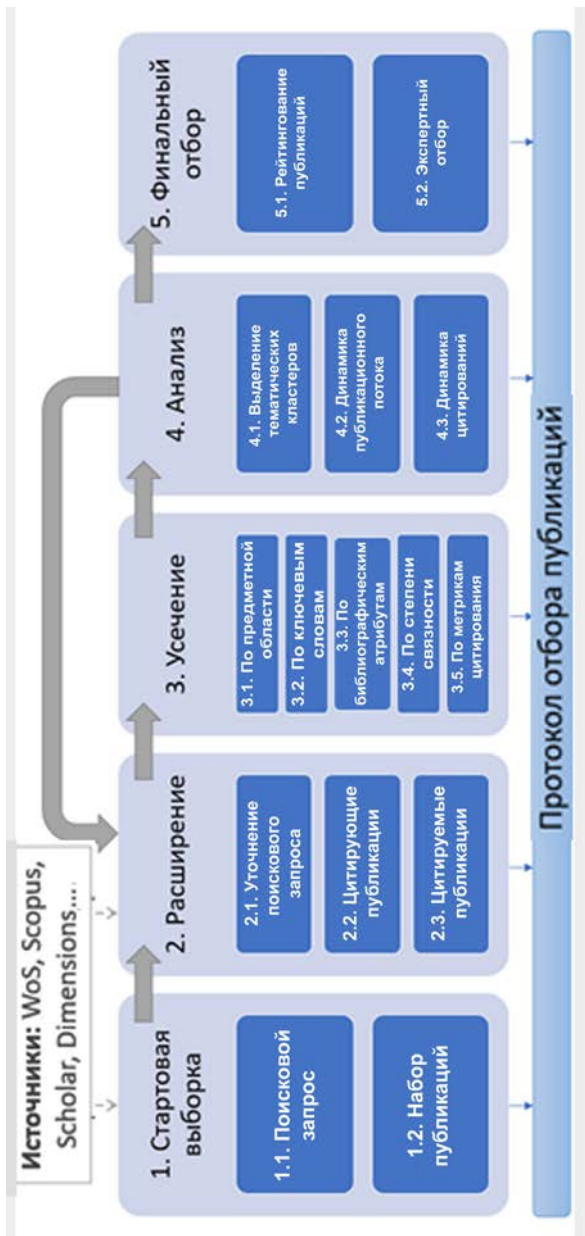
Требования к результатам отбора публикаций

Требование	Описание	Способ реализации
Полнота	Максимальное количество релевантных публикаций	Различные источники данных. Различные стратегии поиска
Релевантность	Эффективные инструменты отсека-ния нерелевантных публикаций и выделения наиболее значимых	Различные стратегии усечения выборки
Воспроизводимость	Возможность повторить методику отбора в независимых условиях	Количественные и каче-ственные критерии отбора публикаций. Ведение протокола отбора публикаций

Требование	Описание	Способ реализации
Структуризация	Выявление групп тематически связанных публикаций и хронологии их изменений	Кластеризация выборки. Анализ динамики изменений публикационного потока
Вариативность	Возможность выбора экспертом наиболее подходящей комбинации стратегий поиска и фильтрации	Цикличность методики

Алгоритм создания выборки релевантных публикаций

Алгоритм состоит из пяти основных этапов, выполняемых экспертом с привлечением инструментальных средств работы с библиографическими данными. На первом этапе готовится стартовая выборка публикаций: эксперт применяет свои исходные знания по изучаемой тематике. Затем она расширяется за счёт новых публикаций (этап 2) или сужается путём удаления нерелевантных (этап 3). Полученная промежуточная выборка подвергается анализу, задачей которого является расширение знаний эксперта о развитии тематики и её внутреннем устройстве (этап 4). Этапы расширения, усечения и анализа выполняются итерационно столько раз, сколько эксперт посчитает нужным для получения оптимального набора публикаций. На некоторых итерациях расширение или сужение выборки может не выполняться. Иными словами, допустимо применение 2-го, 3-го и 4-го этапов в произвольном порядке. На заключительном шаге из полученного массива релевантных публикаций эксперт отбирает те наиболее значимые, которые он включит в обзор (см. рис.).



Алгоритм отбора публикаций для обзора

На каждом этапе на основе ранее полученной выборки (далее – *предшествующая выборка*) формируется её новая, уточнённая версия (*текущая выборка*), а описание критериев включения новых публикаций и исключения ранее добавленных фиксируется в *протоколе отбора*.

Основные этапы

1. Стартовая выборка

Для формирования стартовой выборки на начальном этапе методики эксперт транслирует своё исходное представление об изучаемой тематике одним из двух способов.

1.1. Поисковый запрос

Первый способ – подготовка поискового запроса: выбор ББД, ключевых терминов, хронологического охвата и т. п. Оптимально сформированный поисковый запрос обеспечивает сокращение количества нерелевантных результатов и при этом сохраняет их полноту.

1.2. Набор публикаций

Второй способ – подготовка небольшого набора публикаций, непосредственно относящихся к тематике исследования. В случае большой и неоднородной выборки есть смысл сначала перейти к её усечению и анализу (этапы 3 и 4) и лишь потом к расширению (этап 2).

2. Расширение выборки

Расширение выборки необходимо для добавления тех релевантных публикаций, которые не попали в неё ранее из-за не оптимально составленного поискового запроса, применения нестандартной терминологии или недостаточных знаний эксперта. Расширение выполняется с помощью уточнения исходного поискового запроса и добавления публикаций, которые либо интенсивно цитируют публикации из предшествующей выборки, либо, наоборот, цитируются ими.

В случае получения большой и неоднородной выборки на предыдущем этапе этот шаг на текущей итерации целесообразно пропустить и перейти сначала к её усечению и анализу (этапы 3 и 4).

2.1. Уточнение поискового запроса

Уточнение поискового запроса происходит как по результатам выборки, так и анализа ключевых слов и библиографической связанности научных документов на этапах 3 и 4. Типичный пример – это выявление в предшествующей выборке новых или альтернативных терминов, применение которых является характерными признаками изучаемой тематики. Отметим, что при изменении поискового запроса выборка может не только расшириться, но и сократиться. Поэтому необходимо убедиться, что отсекаемые таким образом публикации действительно являются нерелевантными.

2.2. Цитирующие публикации

В текущую выборку добавляются публикации, которые цитируют работы из предшествующей. Порог минимально необходимого количества цитирований определяется экспертом. Этот шаг даёт возможность пополнить выборку более поздними публикациями, расширяющими и углубляющими предыдущие исследования, которые по тем или иным причинам не попали в первоначальную выборку.

2.3. Цитируемые публикации

В текущую выборку добавляются публикации, которые упоминаются в списках литературы предшествующей выборки. Порог минимально необходимого количества цитирований определяется экспертом. Этот шаг даёт возможность пополнить выборку публикациями, которые легли в основу изучаемой тематики.

Часть публикаций, отобранных на этом этапе, может оказаться нерелевантной и должна быть отброшена на этапе усечения. Кроме того, некоторые из отобранных публикаций могут быть не проиндексированы в используемой ББД. Включение их в текущую выборку позволит в значительной степени снизить ограничения, накладываемые выбором конкретной ББД и её охватом источников. Однако для этого потребуются провести дополнительный поиск библиографических данных таких документов и их обработку.

3. Усечение выборки

Усечение выборки необходимо для исключения нерелевантных (не соответствующих тематике) публикаций. Оно выполняется путём

выявления нерелевантных ключевых слов, иных предметных областей, отсутствия библиографических связей с остальными публикациями из выборки и слабо цитируемых публикаций.

3.1. По предметной области

Из выборки исключаются публикации, по данным ББД тематически относящиеся к нерелевантным предметным областям.

Этот способ следует использовать аккуратно, так как в настоящее время в ББД применяется предметная классификация статей на основе журнала, в котором они опубликованы. Этот подход имеет низкую точность и может привести к ложному исключению релевантных публикаций. Обозначить принадлежность публикации к предметной области может поиск определяющих развитие области терминов в заглавии, аннотации и ключевых словах.

3.2. По ключевым словам

С помощью анализа ключевых слов предшествующей выборки находятся термины, не относящиеся к тематике исследования. Публикации, содержащие эти термины, исключаются из рассмотрения.

Если количество исключаемых публикаций велико, то на этапе 2 можно уточнить поисковый запрос.

3.3. По библиографическим атрибутам

Из выборки исключаются публикации, «отфильтрованные» по библиографическим атрибутам, таким, как источник публикации (непрофильные журналы), её тип (всё, кроме монографий и статей), коллектив авторов с небольшим публикационным вкладом и т. д. Вклад отдельного автора в исследуемую область может оцениваться по индексу Хирша, количеству публикаций в изучаемой области, общему количеству цитирований, а также по его публикациям внутри выборки.

3.4. По степени связности

За несоответствие тематике исследования из выборки удаляются публикации или группы публикаций, сформированные на основе метода «bibliographic coupling», которые не попали ни в один из кластеров на этапе 4.

Библиографическая связь возникает, если две работы ссылаются на общую третью в списках литературы («bibliographic coupling»).

Это указывает на то, что, вероятно, эти две работы касаются одной предметной области. При кластеризации с помощью этого метода образуются группы публикаций, связанные между собой. Если при кластеризации появляется большое количество групп публикаций, не связанных с тематикой будущего обзора, возможно, следует скорректировать первоначальную стратегию поиска (этап 2.3).

3.5. По метрикам цитирований

На данном этапе выборка ранжируется по количеству цитирований, после чего исключаются статьи, которые выходят за минимальный порог цитирования, определяемый экспертом. Релевантные ценные статьи последних лет ещё не успели набрать достаточное количество, поэтому целесообразно ориентироваться на взвешенное цитирование внутри предметной области с учётом года их публикации.

Усечение выборки по метрикам цитирования рекомендуется проводить в последнюю очередь, когда большинство не соответствующих тематической области публикаций будут удалены.

4. Анализ выборки

Этот этап необходим для осмысления экспертом структуры текущей выборки, уточнения его представлений об изучаемой тематике, корректировке правил отбора публикаций и формирования стратегии подготовки обзора.

Основанием отбора научной работы для будущего обзора может являться ориентация на количество цитирований, год публикации, принадлежность к тем или иным группам публикаций (объединённым одним исследовательским вопросом), тип публикации. В протоколе по подготовке обзора должны быть количественно определены (с помощью пороговых значений показателей) и последовательно описаны критерии включения и исключения тех или иных публикаций.

Стоит отметить, что качественные обзоры агрегируют актуальную информацию по теме исследования, соответственно, они являются ценным элементом библиографического поиска. При небольшой по количеству статей тематической выборке рекомендуется добавить все имеющиеся обзоры для последующего анализа экспертами области.

4.1. Выделение тематических кластеров

Кластерный анализ выборки по семантической и библиографической близости позволяет разбить область исследования на отдельные топики. Каждый топик будет содержать список тематически близких публикаций. Анализ по кластерам позволяет убедиться в точности начального выбора стратегии поиска и обогатить словарь дополнительными поисковыми терминами, с помощью которых можно провести уточняющий поиск (этап 2.3). Кластеризацию также можно выполнять: по цитируемым публикациям, по цитирующим публикациям, по авторам, по организациям авторов, по ключевым словам⁴.

Для кластерного анализа авторы рекомендуют использовать специализированные программные средства. Например, к ним относятся бесплатно распространяемые VOSviewer и CiteSpace, которые могут работать с данными из разных источников (WoS, Scopus, Dimensions, CrossRef, Medline).

4.2. Динамика публикационного потока

Анализ динамики количества публикаций позволяет выявить временной отрезок наибольшей концентрации научных работ, а также определить момент возникновения интереса к данной проблематике. Результаты могут влиять на хронологический охват поискового запроса и определение более значимых публикаций.

4.3. Динамика цитирований

На данном этапе анализируется динамика цитирования публикаций. Все публикации ранжируются по двум показателям: общее количество цитирований и количество цитирований публикациями внутри выборки. Определяется временной отрезок, на котором сконцентрированы самые цитируемые работы. Полученные данные соотносятся с мнением эксперта. При необходимости хронологический охват поискового запроса корректируется. Дополнительно можно проанализировать цитирования внутри списков публикаций по группам с учётом общих исследовательских интересов.

⁴ При формировании кластеров по ключевым словам целесообразно удалять малоинформативные термины, что в свою очередь повысит специфичность связей между публикациями и положительно отразится на результатах кластеризации.

5. Финальный отбор

На заключительном этапе эксперт, руководствуясь некоторым набором количественных и качественных критериев, из полученной выборки релевантных публикаций отбирает те, которые войдут в итоговую выборку. Количественные критерии позволяют построить один или несколько ранжированных списков и для каждого определить порог отсеечения. Публикации, оказавшиеся выше порога в любом из рейтингов, включаются в обзор. Качественные показатели отбирает сам эксперт, добавляя на их основании в обзор те работы, которые не проходят по количественным показателям.

Может применяться и сочетание критериев. Например, если рейтинг строится по взвешенному цитированию, то первый, высокий порог отмечает публикации, которые должны попасть в обзор, а второй, более низкий порог – те, которые могут в него попасть после оценки экспертом.

Наконец, для небольших тематик возможна ситуация, когда вся полученная перед финальным этапом выборка публикаций включается в итоговый обзор.

5.1. Рейтингование публикаций

На этапе рейтингования определяется список количественных критериев, основанных на библиометрических показателях. Затем по каждому показателю строятся ранжированные списки публикаций, для которых эксперт указывает пороги отсеечения.

Чаще всего применяются библиометрические показатели публикации (например, взвешенное по году и предметной области количество цитирований, размер списка литературы) или журнала, в котором она была опубликована (например, квартиль журнала в соответствующей предметной области). Также могут применяться наукометрические показатели, относящиеся к основному автору или всему авторскому коллективу:

- количество публикаций,
- количество цитирований всех публикаций,
- количество цитирований публикаций из выборки,
- количество цитирований публикаций из выборки публикациями из выборки,
- индекс Хирша.

Могут применяться и комплексные индикаторы, состоящие из комбинации нормированных показателей, с учётом заданных весов.

Предложенные подходы к рейтингованию, как инструменту отбора, позволяют вариативно оценить массив публикаций (этапы 1–4) и значительно облегчить работу эксперта из исследуемой области за счёт структурирования результатов.

5.2. Экспертный отбор

Финальное решение о включении публикаций в обзор формируется экспертами на основании известных им качественных критериев. Все предшествующие этапы необходимы для подготовки максимально полной выборки релевантных публикаций и их ранжирования, чтобы сделать это решение предельно информированным и обоснованным. Отметим, что для экспертной оценки необходимо приглашать специалистов из рассматриваемой в обзоре тематической области.

Апробация методики

Для апробации методики была выбрана относительно узкая тематическая область – «*embedded librarianship*» (табл. 2). Российские исследователи часто переводят название данного направления на русский язык как «встроенное» библиотечное дело [10, 11]. Нам такой перевод кажется не совсем удачным. Учитывая природу этого определения и специфику данного вида деятельности, в этой работе будет использоваться термин «привлечённый библиотекарь» [13, 14].

Так как понятие «*embedded librarianship*» устоялось, поисковый запрос для выбранного тематического направления содержал кавычки и включал символ * как элемент, позволяющий модифицировать окончание: TITLE-ABS-KEY (“*embedded librar**”). В результате выборки с использованием БД Scopus мы получили 310 публикаций⁵. Расширение поисковой выборки увеличило её до 10 740 публикаций, из них в финальную выборку после анализа по авторам, ключевым словам и цитированию вошло 299 документов. Таким образом, полученная финальная выборка сократилась более чем в десять раз по сравнению с расширенной.

⁵ Данные на июль 2021 г.

Из работ ведущих учёных была сформирована основа для финального отбора релевантной группы публикаций данной области. Анализ цитируемости всех публикаций выборки позволил добавить как самые цитируемые документы, так и те, на которые чаще всего ссылаются авторы внутри выборки. В данном примере пороговое значение оказалось равным 5%, поскольку остальные публикации цитировались значительно реже. Такие значения были определены из-за относительно небольшого количества учёных, занимающихся данной тематикой. Например, максимальное количество публикаций одного автора в исходной выборке равно пяти, тогда как авторы, имеющие только одну публикацию по теме, представляют большую часть поисковой выборки (84%). Такой подход позволил включить самые значимые публикации по общему количеству цитирований и по ценности среди авторов внутри выборки. Среди небольшого количества обзоров⁶ были отобраны все релевантные тематике, которые также вошли в финальную выборку.

Таблица 2

Протокол апробации методики

Этап	Шаг	Критерии включения/исключения	Результаты этапа (число публикаций)
Этап 1. Подготовка поискового запроса	1.1. Поисковый запрос	TITLE-ABS-KEY ("embedded librar*")	310
Этап 2. Расширение выборки	2.1. Цитирующие публикации	Все документы, цитирующие поисковую выборку	2 994
	2.2. Цитируемые публикации	Все упоминаемые документы	7 436 (из них 5 583 проиндексированы в БД Scopus)
Объединение результатов шагов 1.1, 2.1 и 2.2		1.1 + 2.1 + 2.2 – дубликаты 319 + 2 884 + 7 436 – 1 752 = 8 887	8 887

⁶ В конечную выборку попало всего 18 обзоров.

Продолжение таблицы 2

Этап	Шаг	Критерии включения/исключения	Результаты этапа (число публикаций)
Цикл 1. Публикации, индексируемые в БД Scopus			
Этап 3. Усечение выборки	3.1. По предметной области	Соответствующие тематической области (определялось упоминанием в TITLE-ABS-KEY терминов "embedded" и "librar*")	1 015
	3.2. По ключевым словам	Отобраны публикации, содержащие понятие «embedded» в ключевых словах БД Scopus	220
	3.3. По библиографическим атрибутам	Публикации 5% авторов с самыми высокими значениями, показателем <i>h</i> -индекса, количества публикаций и общего количества цитирований	85
	3.5. По метрикам цитирования	Отбор 5% самых цитируемых публикаций	50
Объединение результатов шагов 3.2, 3.3 и 3.5		3.2 + 3.3 + 3.5 – дубликаты 220 + 85 + 50 – 70 = 285	285

Этап	Шаг	Критерии включения/исключения	Результаты этапа (число публикаций)
Этап 4. Анализ результатов цикла 1	4.1. Динамика публикационного потока	Весь временной отрезок (со времени введения понятия «embedded librarianship»). Исключены публикации до 2004 г. ⁷	В обзоре отражены публикации с 2004 г.
	4.2. Динамика цитирования	Весь отрезок цитирования ⁸	С 2007 г. ⁹ наблюдается стабильный рост цитируемости тематики
Цикл 2. Публикации, не индексируемые в БД Scopus			
Этап 2. Расширение выборки	2.2. Цитируемые публикации	Неиндексируемые в БД Scopus публикации, которые упоминают авторы поисковой выборки 7 436 – 5 583 = 1 853	1 853
Этап 3. Усечение выборки	3.5. По метрикам цитирования	Самые часто упоминаемые авторами поисковой выборки публикации, релевантные тематике исследования	14 ¹⁰

⁷ Первая работа по определению понятия «embedded librarianship» автора В. Dewey была опубликована в 2004 г. [15].

⁸ Наибольшее внимание исследователей привлекают классические работы по тематике «embedded librarianship».

⁹ Работа Shumaker and Tyler (2007) стала одной из основополагающих в концепции и привлекла внимание исследователей к изучению данной проблематики (не индексируется в БД Scopus) [16].

¹⁰ Пороговое значение определено на основании результатов рейтингования работ в порядке уменьшения цитирования. Публикация под номером 15 (и последующие) рейтинга цитируется на порядок ниже, чем под номером 14.

Этап	Шаг	Критерии включения/исключения	Результаты этапа (число публикаций)
Объединение результатов			
Объединение результатов, полученных в циклах 1 и 2	285 + 14 = 299		299
Этап 4. Анализ объединённой выборки по результатам циклов 1 и 2	4.1. Выделение тематических кластеров	299 публикаций разделены на отдельные тематические кластеры, отражающие ту или иную специализацию «привлечённого» библиотекаря	Выделились три наиболее крупных кластера: «библиотекарь, участвующий в образовательном процессе», «библиотекарь в медицинской среде» и «библиотекарь, привлечённый к научным исследованиям»
Этап 5. Финальный отбор	5.1. Рейтингование публикаций	Рейтингование публикаций происходило внутри тематических кластеров для оптимизации работы эксперта	Формирование рейтинга по кластерам
	5.2. Экспертный отбор	Небольшие по размеру группы шага 4.1 выборки позволили отобрать публикации для включения в обзор с помощью непосредственного анализа как текста аннотации, так и рейтинга публикации в тематической подобласти	61 публикация рекомендована для включения в обзор

Значительным преимуществом предложенной методики стало включение значимых для области публикаций, не вошедших в первоначальную выборку документов. Так, одной из наиболее частых причин

является отсутствие документа в индексе используемой ББД. Такие неиндексированные публикации, попавшие в выборку на этапе расширения, потребовали особого подхода из-за отсутствия некоторых данных, в том числе по цитированию. Поэтому публикации, отобранные на этом этапе, анализировались двумя отдельными циклами.

Коллекция из часто упоминаемых внутри выборки публикаций, по тем или иным причинам не попавшим в результаты поискового запроса, дополнила финальную выборку публикациями с высокой научной значимостью. Например, авторы расширенной выборки активно цитировали работы одного из ключевых исследователей направления «*embedded librarianship*» Д. Шумакера. Три публикации автора, содержащие основополагающие принципы концепции «привлечённого библиотекаря», оказались не проиндексированы в БД Scopus и, соответственно, по этой причине отсутствовали в результатах поискового запроса [17–19].

Кроме того, публикации могут не попасть в результаты первичной поисковой выборки из-за отсутствия в заглавии, аннотации или ключевых словах словосочетаний из поискового запроса. Приведём наиболее распространённые примеры:

Заглавие статьи в индексе БД Scopus содержит ошибку: «*The embedded libra – Rian*».

В аннотацию интересующие термины включены в соответствии с контекстом по правилам английской грамматики: «*Purpose – This paper seeks to present the findings of a series of experiments in *embedding a librarian* at a variety of levels into the undergraduate classroom*»¹¹.

В заглавии статьи поисковые термины разделены дополнительным словом: «*The embedded science librarian: Partner in curriculum design and delivery*».

Тематика исследований деятельности «привлечённого библиотекаря» для апробации предложенной методики была выбрана случайно. Такой формат общения библиотекарей и академического сообщества оптимален при реализации данной методики. Обзор публикаций по данному направлению показал, что в должностные функции такого специалиста входит умение использовать в своей деятельности

¹¹ Данная публикация не была включена в финальный отбор, так как была отфильтрована на этапе усечения выборки.

работу с наукометрическими БД для создания информационного продукта (в данном случае обзора) при тесном контакте с исследователями. Эффективное применение методики возможно только при плотном контакте специалиста библиотечной среды и исследователей из разных тематических областей на этапах выбора тематики исследования, при корректировке процесса подбора литературы и на этапе заключительной экспертной оценки результатов.

Апробация методики позволила сформировать репрезентативную выборку предметной области «*embedded librarianship*», обзор которой будет представлен в отдельной публикации.

Заключение

Преимущества применения данной методики, в отличие от традиционного поиска с использованием только ключевых терминов, очевидны. Большие по объёму результаты поисковых запросов не позволяют исследователям тщательно изучить содержимое всех публикаций. Требуется дополнительные манипуляции, самой популярной среди которых является уточнение поискового запроса, приводящее к уменьшению исходной выборки. При этом велика вероятность упустить из вида значимые исследования. Безусловно, уточнение запроса опытным специалистом, знающим ведущих учёных и основополагающие публикации научного направления, позволяют уменьшить негативные эффекты. Но далеко не всегда исследователь, приступающий к подготовке обзора (в том числе в рамках диссертационного исследования), обладает таким знанием области.

Предложенный методологический подход, предполагающий высокую степень вариативности при отборе публикаций, позволяет значительно снизить риск пропуска важных релевантных статей, сохранив при этом определённую гибкость при адаптации к специфике различных научных дисциплин.

СПИСОК ИСТОЧНИКОВ

1. **Yu Z., Menzies T.** FAST2: An intelligent assistant for finding relevant papers // *Expert Systems with Applications*. 2019. Vol. 120. P. 57–71.
2. **Egger M. et. al.** How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study // *Health Technol Assessment*. 2003. Vol. 1. № 7. P. 1–76.
3. **Song F. et al.** Dissemination and publication of research findings: an updated review of related biases // *Health Technol Assessment*. 2010. Vol. 14. № 8. P. III, IX–XI, 1–193.
4. **Sideri S., Papageorgiou S. N., Eliades T.** Registration in the international prospective register of systematic reviews (PROSPERO) of systematic review protocols was associated with increased review quality // *Journal of Clinical Epidemiology*. 2018. Vol. 100. C. 103–110. doi: 10.1016/j.jclinepi.2018.01.003
5. **Liberati A. et. al.** The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration // *Journal of Clinical Epidemiology*. 2009. Vol. 62. № 10. P. e1–e34. doi: 10.1016/j.jclinepi.2009.06.006
6. **Bagirova A. V., Kosyakov D. V., Guskov A. E.** The 50 Most Highly Cited Reviews of 2013–2017 // *Scientific and Technical Information Processing*. 2021. Vol. 48. № 3. P. 168–184.
7. **Buchkremer R. et. al.** The application of artificial intelligence technologies as a substitute for reading and to support and enhance the authoring of scientific review articles // *IEEE Access*. 2019. Vol. 7. P. 65263–65276. doi: 10.1109/ACCESS.2019.2917719
8. **Chen C.** Cascading citation expansion // *Journal of Information Science Theory and Practice*. 2018. Vol. 2. № 6. P. 6–23.
9. **Chen C., Song M.** Visualizing a field of research: A methodology of systematic scientometric reviews // *PLoS ONE*. 2019. Vol. 10. № 14. P. e0223994. <https://doi.org/10.1371/journal.pone.0223994>
10. **Каменская М. А.** Эволюция информационно библиотечного обслуживания: вопросы терминологии // *Научно-техническая информация. Серия 1: Организация и методика информационной работы*. 2016. № 8. С. 24–33.
11. **Галявиева М. С.** Научные библиотеки в информационном сопровождении и поддержке научных исследований: основные тенденции // *Труды ГПНТБ СО РАН*. 2017. № 121. С. 221–227.
12. **Kesselman M. A., Watstein S. B.** Creating opportunities: Embedded librarians // *Journal of Library Administration*. 2009. Vol. 49. № 4. P. 383–400. doi: 10.1080/01930820902832538
13. **Drewes K., Hoffman N.** Academic embedded librarianship: An introduction // *Public service quarterly*. 2010. Vol. 6. № 2–3. P. 75–82. doi: 10.1080/15228959.2010.498773

14. **Ziede E. J.** In bed with the military: First Amendment implications of embedded journalism // NYUL Rev. 2005. Vol. 80. P. 1309.
15. **Dewey B. I.** The embedded librarian: Strategic campus collaborations // Resource Sharing & Information Networks. 2004. Vol. 17. № 1–2. P. 5–17. doi: 10.1300/J121v17n01_02
16. **Shumaker D., Tyler L.** Embedded Library Services: An Initial Inquiry into Practices for Their Development, Management, and Delivery. In: Shumaker D. et al. (eds.) Special Libraries Association Annual Conference, Denver, CO (2007).
17. **Shumaker D.** The Embedded Librarian: Innovative Strategies for Taking Knowledge Where It's Needed // Medford: Information Today. 2012. XVII, 212 S.
18. **Shumaker D. et. al.** Models of Embedded Librarianship Final Report (2009). URL: https://embeddedlibrarian.files.wordpress.com/2013/04/models-of-embedded-librarianship_finalreportrev.pdf.
19. **Shumaker D., Talley M.** Models of embedded librarianship: A research summary // Information Outlook. 2010. Vol. 14. № 1. P. 26–35.

References

1. **Yu Z., Menzies T.** FAST2: An intelligent assistant for finding relevant papers // Expert Systems with Applications. 2019. Vol. 120. P. 57–71.
2. **Egger M. et. al.** How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study // Health Technol Assessment. 2003. Vol. 1. № 7. P. 1–76.
3. **Song F. et al.** Dissemination and publication of research findings: an updated review of related biases // Health Technol Assessment. 2010. Vol. 14. № 8. P. III, IX–XI, 1–193.
4. **Sideri S., Papageorgiou S. N., Eliades T.** Registration in the international prospective register of systematic reviews (PROSPERO) of systematic review protocols was associated with increased review quality // Journal of Clinical Epidemiology. 2018. Vol. 100. C. 103–110. doi: 10.1016/j.jclinepi.2018.01.003
5. **Liberati A. et. al.** The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration // Journal of Clinical Epidemiology. 2009. Vol. 62. № 10. P. e1–e34. doi: 10.1016/j.jclinepi.2009.06.006
6. **Bagirova A. V., Kosyakov D. V., Guskov A. E.** The 50 Most Highly Cited Reviews of 2013–2017 // Scientific and Technical Information Processing. 2021. Vol. 48. № 3. P. 168–184.
7. **Buchkremer R. et. al.** The application of artificial intelligence technologies as a substitute for reading and to support and enhance the authoring of scientific review articles // IEEE Access. 2019. Vol. 7. P. 65263–65276. doi: 10.1109/ACCESS.2019.2917719

8. **Chen C.** Cascading citation expansion // *Journal of Information Science Theory and Practice*. 2018. Vol. 2. № 6. P. 6–23.
9. **Chen C., Song M.** Visualizing a field of research: A methodology of systematic scientometric reviews // *PLoS ONE*. 2019. Vol. 10. № 14. P. e0223994. <https://doi.org/10.1371/journal.pone.0223994>
10. **Kamenskaia M. A.** E`voliutiia informacii bibliotchnogo obsluzhivaniia: voprosy` terminologii // *Nauchno-tekhicheskaia informatciia. Serii 1: Organizatciia i metodika informacii noi` raboty`*. 2016. № 8. S. 24–33.
11. **Galiavieva M. S.** Nauchny`e biblioteki v informacii nom soprovozhdenii i podderzhke nauchny`kh issledovani`: osnovny`e tendencii // *Trudy` GPNTB SO RAN*. 2017. № 121. S. 221–227.
12. **Kesselman M. A., Watstein S. B.** Creating opportunities: Embedded librarians // *Journal of Library Administration*. 2009. Vol. 49. № 4. P. 383–400. doi: 10.1080/01930820902832538
13. **Drewes K., Hoffman N.** Academic embedded librarianship: An introduction // *Public service quarterly*. 2010. Vol. 6. № 2–3. P. 75–82. doi: 10.1080/15228959.2010.498773
14. **Ziede E. J.** In bed with the military: First Amendment implications of embedded journalism // *NYUL Rev*. 2005. Vol. 80. P. 1309.
15. **Dewey B. I.** The embedded librarian: Strategic campus collaborations // *Resource Sharing & Information Networks*. 2004. Vol. 17. № 1–2. P. 5–17. doi: 10.1300/J121v17n01_02
16. **Shumaker D., Tyler L.** Embedded Library Services: An Initial Inquiry into Practices for Their Development, Management, and Delivery. In: Shumaker D. et al. (eds.) *Special Libraries Association Annual Conference, Denver, CO (2007)*.
17. **Shumaker D.** The Embedded Librarian: Innovative Strategies for Taking Knowledge Where It's Needed // *Medford: Information Today*. 2012. XVII, 212 S.
18. **Shumaker D. et. al.** Models of Embedded Librarianship Final Report (2009). URL: https://embeddedlibrarian.files.wordpress.com/2013/04/models-of-embedded-librarianship_finalreportrev.pdf.
19. **Shumaker D., Talley M.** Models of embedded librarianship: A research summary // *Information Outlook*. 2010. Vol. 14. № 1. P. 26–35.

Информация об авторах / Information about the authors

Мальшева Александра Валерьевна – младший научный сотрудник лаборатории наукометрии и научных коммуникаций Российского научно-исследовательского института экономики, политики и права в научно-технической сфере, Москва, Российская Федерация
bag_bala@mail.ru

Косьяков Денис Викторович – заместитель заведующего лаборатории наукометрии и научных коммуникаций Российского научно-исследовательского института экономики, политики и права в научно-технической сфере, Москва, Российская Федерация
kosyakov@sciencepulse.ru

Гуськов Андрей Евгеньевич – канд. техн. наук, заведующий лабораторией наукометрии и научных коммуникаций Российского научно-исследовательского института экономики, политики и права в научно-технической сфере, Москва, Российская Федерация
guskov.andrey@gmail.com

Aleksandra V. Malysheva – Junior Researcher, Laboratory for Scientometrics and Scholarly Communications, Russian Research Institute of Economics, Politics and Law in Science and Technology, Moscow, Russian Federation
bag_bala@mail.ru

Denis V. Kosyakov – Deputy Head of the Laboratory for Scientometrics and Scholarly Communications, Russian Research Institute of Economics, Politics and Law in Science and Technology, Moscow, Russian Federation
kosyakov@sciencepulse.ru

Andrey E. Guskov – Cand. Sc. (Engineering), Head of the Laboratory for Scientometrics and Scholarly Communications, Russian Research Institute of Economics, Politics and Law in Science and Technology, Moscow, Russian Federation
guskov.andrey@gmail.com