

УДК 004.65:025.4.03:330.14:331.101.262:61

ПОСТРОЕНИЕ КАРТЫ СОПРИСУТСТВИЯ КЛЮЧЕВЫХ СЛОВ ПО ТЕМЕ «КАПИТАЛ ЗДОРОВЬЯ» В ПРОГРАММЕ VOSVIEWER

Павлова Ирина Анатольевна,

iapav@mail.ru

Национальный исследовательский университет «Высшая школа экономики»,
Россия, 101000, г. Москва, ул. Мясницкая, 20

Павлова Ирина Анатольевна, кандидат экономических наук, старший научный сотрудник Международной лаборатории прикладного сетевого анализа Национального исследовательского университета «Высшая школа экономики».

Актуальность. Библиометрический анализ с применением методов сетевого анализа представляет собой активно развивающееся научное направление, методы и программные инструменты которого помогают понять структуру и эволюцию знаний научных дисциплин. VOSviewer – открытый программный продукт, при помощи которого строятся библиометрические карты (сети) разных типов (соавторство, сопричастие, со-цитирование и т. д.). Библиометрический анализ сопричастия ключевых слов предполагает исследование одновременного присутствия ключевых слов в разных публикациях и графически позволяет визуализировать карту ключевых слов, которую при определённых допущениях можно рассматривать как модель структуры знаний в анализируемой области. **Цель:** построение карты ключевых слов исследовательского поля по теме «Капитал здоровья» на основе материалов базы данных Web of Science; анализ процесса построения библиометрической сети сопричастия ключевых слов при помощи программы VOSviewer. **Методы:** библиометрический анализ с методами сетевого анализа. Эмпирической базой исследования выступают материалы сети Web of Science (57853 публикации, найденные по словам health и capital и выгруженные весной 2022 г. файлами по 500 публикаций в каждом в формате full record). Для построения карты ключевых слов использована программа VOSviewer, версия 1.6.19. **Результаты.** На первом этапе, при построении карты ключевых слов на основе всего массива из 57 853 публикаций без предварительной обработки данных (очистки данных), выделено и проанализировано пять кластеров. На втором этапе построена карта ключевых слов на основе 3000 максимально релевантных публикаций с обработкой данных (очистка данных через построение тезауруса). В результате выделены и проанализированы ключевые слова, которые характеризуют разные типы капиталов (социальный, человеческий, культурный, психологический). **Выводы.** Построение карты ключевых слов может реализовываться до проведения систематического обзора, поскольку такой предварительный анализ поля исследования направляет и ускоряет процесс литературного обзора по нужной научной тематике. Такая карта – сеть ключевых слов – способствует расширению понимания за счет охвата теорий, концепций, подходов и визуализации ключевых слов проблемного поля исследования. В представленном примере по теме «Капитал здоровья» показано, что даже с необработанными данными можно построить кластеры, которые благодаря визуальному представлению проблемного поля через ключевые слова способствуют формированию дальнейших шагов исследования по выявленным направлениям. Библиометрический анализ сопричастия ключевых слов позволил представить модель структуры области знаний по теме капитала здоровья, а именно: 1) картировать тематические кластеры; 2) построить темпоральные сети для анализа эволюции знаний по выбранному направлению; 3) выявить разные типы капиталов, которые можно анализировать в дальнейшем, и на их основе проводить систематический литературный обзор.

Ключевые слова: библиометрический анализ, капитал здоровья, VOSviewer, сетевой анализ, сопричастие ключевых слов.

Введение

Библиометрия и наукометрия – важные инструменты, которые можно использовать для измерения научного вклада. При этом учитываются различные аспекты, такие как количество опубликованных статей, цитируемость и импакт-факторы. Это помогает ученым понять свой вклад в научную область, а также сравнивать результативность других исследователей. Библиометрия является одним из самых простых и естественных способов определения научного вклада исследователя, что делает ее бесценным инструментом для любого академического сообщества, исследовательского учреждения и экспертного сообщества.

Библиометрический анализ научных сетей – это исследования, которые применяют методологию библиометрического сетевого анализа для изучения различных научных полей [1]. Эта методология подразумевает анализ большого объема данных из базы научного цитирования Web of Science или других научных баз данных [2] и предполагает построение сетей, связывающих разные библиометрические объекты (публикации, авторов, журналы, ключевые слова и т. д.) отношениями различного типа (соприсутствие, цитирование, соавторство, со-цитирование, библиографическое сочетание (bibliographic coupling) и др.). Анализ таких сетей предоставляет возможности для изучения эволюции различных научных отраслей и дисциплин, их тематической структуры и научных коллективов.

Задачи исследователя при построении библиометрических карт (графической визуализации результатов библиометрического анализа) связаны с получением информации и знаний о том, как устроена библиометрическая сеть: 1) каковы основные темы/направления исследований в рамках научной области; 2) как эти темы/направления связаны друг с другом; 3) как со временем развивалась проблематика научной области (каков темпоральный контекст развития научной области).

В последнее время активное развитие получило применение библиометрического подхода в анализе ключевых слов. Подход предполагает формирование списка наиболее часто используемых связанных друг с другом ключевых слов, которые далее можно систематически исследовать в контексте выбранных групп авторов и журналов. «В современном академическом мире ключевые слова стали важной частью информации о публикациях, так как обычно они являются обязательными для предоставления статьи или книги. Однако, если ключевые слова не указаны автором, они могут быть присвоены статье журналом или базой данных или автоматически извлечены из заголовка. Таким образом, тематическая идентичность любого поля может быть легко построена на основе метаданных академических работ [3, с. 358]». Анализ ключевых слов помогает построить сеть – сформировать поле (карту) ключевых слов, которые характерны для конкретного направления исследований или научной проблематики. В настоящее время формирование такой карты ключевых слов часто реализуется при помощи множества программных продуктов Pajek, VOSviewer, Gephi и т. д. Такая автоматизация работы исследователя облегчает задачу анализа литературы и проведения качественного систематического литературного обзора.

Методология исследования

Целью данной работы является построение карты ключевых слов исследовательского поля по теме «Капитал здоровья», а также анализ процесса построения библиометрической сети соприсутствия ключевых слов при помощи программы VOSviewer. В качестве эмпирической базы исследования были использованы материалы сети Web of Science. Поиск публикаций осуществлялся простым способом по словам health и

capital. Всего найдено 57 853 публикации. Выгрузка данных производилась весной 2022 г. файлами по 500 публикаций в каждом (без пристатейных списков литературы) в формате full record. Для построения карты ключевых слов использована программа VOSviewer, версия 1.6.19. В статье представлено построение карт ключевых слов – сетей из библиометрических объектов на основе отношений сопричастия (co-occurrence, совместное появление ключевых слов).

VOSviewer – программа, разработанная в Лейденском университете (Королевство Нидерланды). Это программный инструмент для построения и визуализации библиометрических сетей (сети журналов, сети исследователей, сети публикаций). Эти сети могут быть построены на основе цитирования, библиографической связи, совместного цитирования или отношений соавторства. VOSviewer также предлагает функциональные возможности анализа текста, которые применяются для создания и визуализации сетей при совпадении важных терминов, извлеченных из массива научной литературы [4].

В отличие от большинства компьютерных программ, используемых для проведения библиометрического анализа, VOSviewer уделяет особое внимание графическому представлению библиометрических сетей. Таким образом, в программе реализуется библиометрический анализ при помощи инструментов сетевого анализа через построение сетей. Функциональность VOSviewer особенно полезна для отображения больших библиометрических карт в удобном для понимания пользователя виде [5]. Для построения сетей – картирования данных – VOSviewer использует метод отображения VOS [6] (visualization of similarities – визуализация сходств, далее метод VOS). Разработчики программы VOSviewer выделяют два подхода к картированию в библиометрических исследованиях – построение карт на базе расстояний между объектами и карт на основе графов (табл. 1).

VOSviewer строит библиометрическую карту на основе матрицы совпадений (co-occurrence matrix) – матрицы совместной встречаемости. Построение карты представляет собой процесс, состоящий из трех этапов [5]:

- на первом этапе вычисляется матрица подобия (similarity matrix) на основе матрицы совпадений;
- на втором этапе строится карта путем применения метода отображения VOS к матрице подобия;
- на третьем этапе карта перемещается, поворачивается и отражается.

Следует отметить важный момент при визуализации с применением метода VOS: данный метод использует в качестве входных данных матрицы подобия, которые могут быть получены из матриц совпадения путем нормализации, то есть путем корректировки матрицы на различия в общем количестве появлений или совпадений элементов. Наиболее популярными мерами сходства для нормализации данных являются близость косинусов угла и индекс подобия Жаккара (Jaccard index). Однако в VOSviewer вместо этого используется мера сходства, известная как сила связи/ассоциации (association strength) [7], также известная как индекс близости (proximity index) [8] или вероятностный индекс соответствия (probabilistic affinity index) [9]. В работе Van Eck и Waltman [10] анализ мер подобия, применяемых в наукометрических исследованиях (scientometrics), показал, что существуют два принципиально различных типа мер подобия: 1) теоретико-множественные меры (set-theoretic), которые можно интерпретировать как меры относительного перекрытия двух множеств; 2) вероятностные (probabilistic) меры, которые можно интерпретировать как меры отклонения наблюдаемых частот совпадений от ожидаемых частот совпадений при предположении о независимо-

сти. Сила связи/ассоциации является вероятностной мерой, тогда как близость косинусов угла и индекс Жаккара представляют собой теоретико-множественные меры. Каждый из этих типов мер подобия решает свои задачи; в VOSviewer применяется вероятностная мера подобия – сила связи/ассоциации.

Таблица 1. Сравнение подходов к построению библиометрических карт [5]
Table 1. Comparison of approaches to build bibliometric maps [5]

Подход к построению библиометрических карт Approach to build bibliometric network (bibliometric mapping)	Карты, основанные на расстоянии Distance-based maps	Карты на основе графа Graph-based maps
Характеристика Characteristics	Карты, в которых расстояние между двумя элементами отражает силу связи между элементами. Меньшее расстояние обычно указывает на более сильную связь Maps where the distance between two elements reflects the strength of the connection (link strength) between the elements. A smaller distance usually indicates a stronger connection	Карты, в которых расстояние между двумя элементами не обязательно должно отражать силу связи между элементами. Вместо этого между элементами рисуются линии для обозначения отношений Maps in which the distance between two elements does not have to reflect the strength of the relationship between the elements. Instead, lines are drawn between elements to indicate relationships.
Распределение элементов на библиометрической карте Distribution of elements on the bibliometric map	Во многих случаях элементы распределены на картах, основанных на расстоянии, неравномерно In many cases, the elements are distributed quite unevenly in maps based on distance	На картах на основе графов элементы часто распределяются равномерно In graph-based maps, elements are often fairly evenly distributed
Преимущества Advantages	Упрощается идентификация кластеров связанных элементов Identification of clusters of related elements is simplified	Наблюдается меньшее количество проблем с перекрывающимися метками элементов при визуализации There are fewer issues with overlapping item labels when rendering visuals
Недостатки Disadvantages	Иногда на библиометрической карте затруднена маркировка всех элементов, перекрывающих друг друга Sometimes it is difficult to mark on a bibliometric map all the elements that overlap each other	Обычно сложнее определить силу связи между двумя элементами, а также обнаружить кластеры связанных элементов It is usually more difficult to see the strength of the relationship between two elements as well as to detect clusters of related elements
Примеры методов (техник) реализации библиометрических карт Examples of methods (techniques) for the implementation of bibliometric maps	<ul style="list-style-type: none"> • Multidimensional scaling; • VOS (visualization of similarities); • VxOrd; • Kopcsa-Schiebel 	<ul style="list-style-type: none"> • Kamada-Kawai; • Fruchterman-Reingold; • Pathfinder networks
Компьютерные программы Computer programs	VOSviewer	<ul style="list-style-type: none"> • Pajek • CiteSpace • Network Workbench Tool

Источник: составлено автором.

Source: compiled by the author.

В программе VOSviewer реализуется построение карт как на основе полного счета (full counting), так и фракционного – дробного счета (fractional counting). Полный счет предполагает, что все связи между библиографическими объектами имеют равный вес; фракционный счет предполагает дробный учет участия: например, если у публикации имеется 10 соавторов, то участие каждого из соавторов будет учитываться как 1/10. В этом случае дробного счета каждая публикация будет иметь одинаковый общий вес. В определенном смысле в случае полного счета высокоцитируемые публикации или публикации с большим количеством авторов рассматриваются и будут представлены на карте как более репрезентативные, чем малоцитируемые публикации или публикации с меньшим количеством авторов, что может быть не всегда желательно. Perianes-Rodriguez, Waltman и Van Eck отмечают, что построение библиометрических сетей на основе полного и фракционного счета для цитирований, со-цитирований, библиографических сочетаний может привести к получению разных результатов [11]. Различные методы счета позволяют по-разному работать с данными, что в целом зависит от цели анализа – какой именно подход может быть более полезным для исследовательской задачи. Для анализа вышеуказанных сетей рекомендуется использовать фракционный счет. Однако, основываясь на теоретических соображениях и эмпирическом анализе, авторы отмечают, что в некоторых случаях результаты построения сетей на основе фракционного счета бывает сложнее интерпретировать.

В данном исследовании в рамках библиометрического анализа реализовано построение карты ключевых слов (сети ключевых слов) при помощи типа co-occurrence – соприсутствия ключевых слов. Такой тип анализа оценивает, насколько соотносятся друг с другом библиометрические объекты (ключевые слова) исходя из того количества документов, в которых они присутствуют одновременно (соприсутствуют). При корректном описании элементов такой сети (в нашем случае это ключевые слова) и взаимосвязей между ними (как эти ключевые слова одновременно присутствуют в публикациях) мы получаем сеть (карту) из ключевых слов. Этот подход мы можем, в том числе, определить как картирование структуры знаний по соответствующему научному направлению. Любой тип визуализации никогда не будет совершенным, поскольку реальная «структура знаний» представляет собой исключительно сложное явление и лежит за пределами любой двух- или трехмерной системы визуализации [12]. Тем не менее карту ключевых слов можно рассматривать как модель структуры знаний (модель структуры научной темы или научного направления). Описанное направление библиометрического анализа получило в литературе собственное название keyword co-occurrence networks (KCNs) – сети соприсутствия ключевых слов. Сеть, построенная таким образом, представляет собой кумулятивное знание предметной области и помогает раскрыть значимые компоненты знаний на основе взаимодействия элементов и силе связей между ключевыми словами, которые встречаются в литературе [13].

С целью построения карты ключевых слов по направлению «Капитал здоровья» проведен анализ в два этапа:

- этап 1. Построение карты ключевых слов на основе 57 853 публикаций, выгруженных из базы Web of Science по ключевым словам health и capital. Анализ проводился без предварительной обработки данных (очистка данных);
- этап 2. Построение карты ключевых слов на основе 3000 максимально релевантных публикаций, выгруженных из базы Web of Science по ключевым словам health и capital. Анализ проводился с обработкой данных (очистка данных через построение тезауруса).

В табл. 2 представлен алгоритм проведения анализа в программе VOSviewer.

Таблица 2. Алгоритм построения карты ключевых слов по теме «Капитал здоровья» в программе VOSviewer

Table 2. Algorithm of mapping keywords on the topic «Health capital» in VOSviewer

Этап/Step	Действия/Actions
Создать Create	Выбираем опцию «Создать» Choose an option «Create»
Выбрать тип данных Choose data type	Выбираем опцию «Создать карту на основе библиографических данных» Choose an option «Create a map based on bibliographic data»
Выбрать источник данных Choose data source	Выбираем опцию «Загрузить библиографические данные из файлов из базы данных» Choose an option «Read data from bibliographic database files»
Выбрать файлы Select files	Выбираем файлы с расширением ciw Choose all necessary files with ciw extension
Выбрать тип анализа и метод счета Choose type of analysis and counting method	Выбираем варианты/Choose options: <ul style="list-style-type: none"> • Совместное присутствие ключевых слов/Co-occurrence; • Полный или фракционный счет/Full/Fractional Counting; • Все ключевые слова/All keywords

Источник: составлено автором.
Source: compiled by the author.

Результаты и обсуждение

Этап 1. Построение карты ключевых слов на основе 57 853 публикаций

В загруженном наборе данных в общей сложности было выделено 126 084 ключевых слова. Если установить ограничение на минимальное количество сопричастующих ключевых слов – 15 слов, то только 4519 ключевых слов соответствуют порогу и участвуют в построении карты. Далее для всех отобранных 4519 ключевых слов рассчитывается общая сила связи (total link strength) и в нашу визуализацию включаются только 1000 ключевых слов (1000 слов установлено по умолчанию, хотя количество можно менять по желанию исследователя).

Create Map					Create Map				
Verify selected keywords					Verify selected keywords				
Selected	Keyword	Occurrences	Total link strength		Selected	Keyword	Occurrences	Total link strength	
<input checked="" type="checkbox"/>	health	5313	29363		<input checked="" type="checkbox"/>	health	5313	5188.00	
<input checked="" type="checkbox"/>	mortality	2973	17633		<input checked="" type="checkbox"/>	risk	3154	3097.00	
<input checked="" type="checkbox"/>	prevalence	3071	17625		<input checked="" type="checkbox"/>	prevalence	3071	3023.00	
<input checked="" type="checkbox"/>	social capital	2725	17377		<input checked="" type="checkbox"/>	mortality	2973	2946.00	
<input checked="" type="checkbox"/>	risk	3154	17198		<input checked="" type="checkbox"/>	social capital	2725	2670.00	
<input checked="" type="checkbox"/>	association	2023	12774		<input checked="" type="checkbox"/>	children	2477	2376.00	
<input checked="" type="checkbox"/>	impact	2403	12750		<input checked="" type="checkbox"/>	impact	2403	2362.00	
<input checked="" type="checkbox"/>	children	2477	12050		<input checked="" type="checkbox"/>	disease	2051	2008.00	
<input checked="" type="checkbox"/>	disease	2051	10304		<input checked="" type="checkbox"/>	association	2023	2006.00	
<input checked="" type="checkbox"/>	depression	1303	8883		<input checked="" type="checkbox"/>	management	1960	1876.00	
<input checked="" type="checkbox"/>	risk-factors	1509	8746		<input checked="" type="checkbox"/>	expression	1792	1768.00	
<input checked="" type="checkbox"/>	management	1960	8578		<input checked="" type="checkbox"/>	epidemiology	1582	1542.00	
<input checked="" type="checkbox"/>	population	1407	8492		<input checked="" type="checkbox"/>	outcomes	1551	1518.00	
<input checked="" type="checkbox"/>	outcomes	1551	8092		<input checked="" type="checkbox"/>	care	1560	1493.00	
<input checked="" type="checkbox"/>	epidemiology	1582	8074		<input checked="" type="checkbox"/>	risk-factors	1509	1485.00	
<input checked="" type="checkbox"/>	expression	1792	7752		<input checked="" type="checkbox"/>	population	1407	1389.00	
<input checked="" type="checkbox"/>	care	1560	7745		<input checked="" type="checkbox"/>	depression	1303	1298.00	
<input checked="" type="checkbox"/>	obesity	1179	7646		<input checked="" type="checkbox"/>	women	1251	1224.00	
<input checked="" type="checkbox"/>	women	1251	7048		<input checked="" type="checkbox"/>	china	1257	1215.00	

Рис. 1. Список ключевых слов для полного и фракционного счета по теме «Капитал здоровья»

Fig. 1. List of keywords for full and fractional counting on the topic «Health capital»

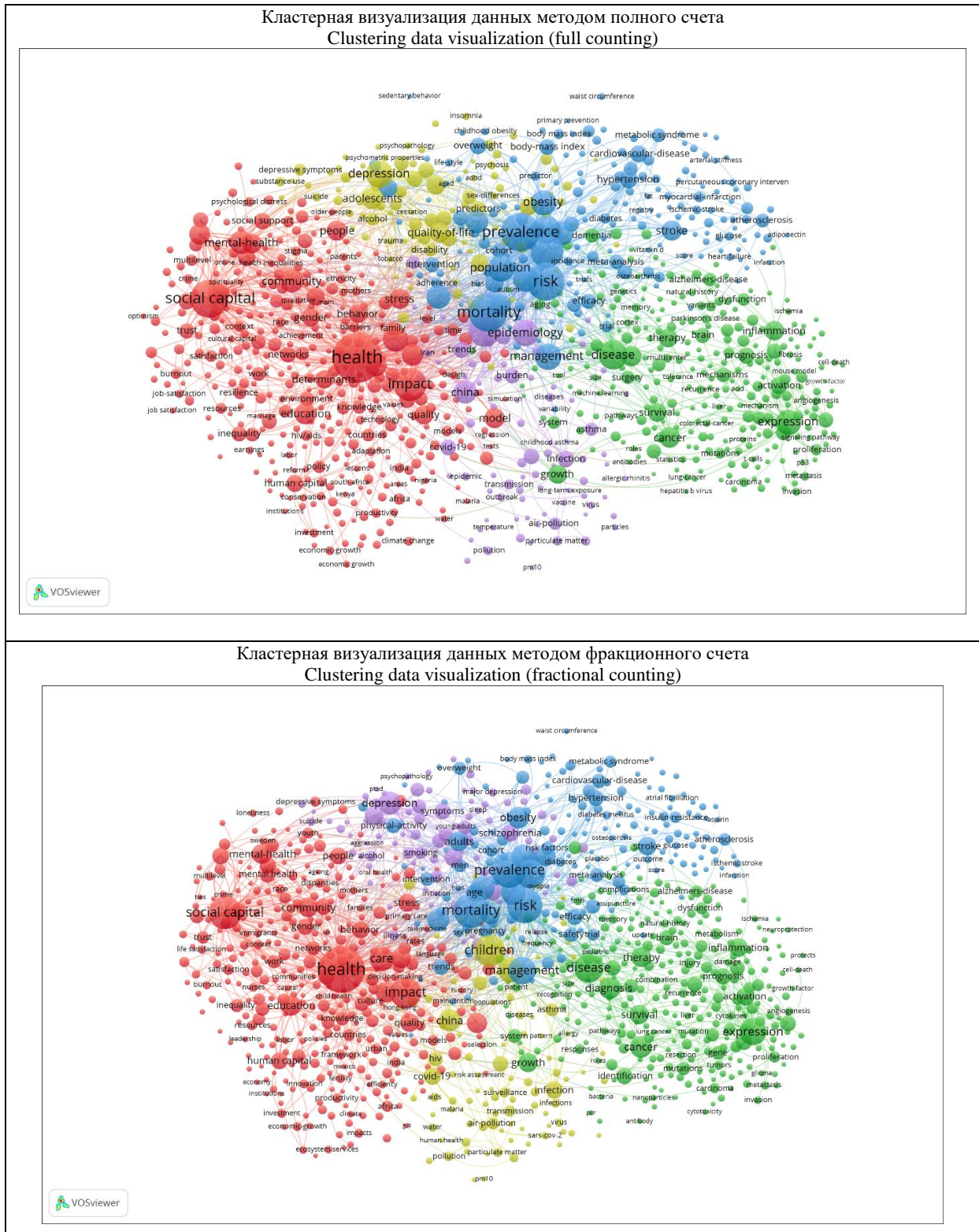
При выборе полного или фракционного счета списки ключевых слов отличаются (рис. 1) – наблюдаются отличия в порядке отображения некоторых ключевых слов. При выводе данных на визуализацию автоматически в программе VOSviewer метод нормализации выставляется как association strength. При построении разных карт в блоке Weights выбираем total link strength для сохранения единообразия методики при визуализации данных как для полного, так и фракционного счета.

В библиометрическом и наукометрическом анализе построение карт и кластеризация часто используются вместе и имеют аналогичную цель – дать представление о структуре сети. Однако, несмотря на их тесную связь, методы сетевого анализа и кластеризации обычно разрабатывались отдельно друг от друга [14]. В программе VOSviewer реализована возможность одновременного картирования (построения сети) и кластеризации (выделение кластеров в этой сети). Таким образом, одновременно визуализируется не только структура сети, но и выделяются кластеры внутри этой сети. В VOSviewer интерактивно можно менять параметры, в том числе предложенные методы кластеризации для выбора наиболее подходящего метода, который позволит наилучшим образом интерпретировать выделяемые кластеры.

В работе вне зависимости от выбора метода счета – полный или фракционный – при построении карты выделяются пять кластеров, которые очень похожи (рис. 2). Можно сказать, что при визуальном исследовании они практически идентичны. На рис. 2 разные кластеры представлены разными цветовыми решениями. Цветовые решения для полного и фракционного счета немного отличаются (некоторые идентичные кластеры в полном и фракционном счете отличаются цветовыми решениями). При визуализации библиометрической сети в программе VOSviewer размер объекта (ключевое слово на карте) обозначает его полную силу связей, а толщина линии и расстояние между двумя ключевыми словами – силу связей между этими объектами (терминами).

Содержательно пять кластеров в первую очередь представлены ключевыми словами по тематике разнообразных аспектов здоровья:

1. Самый крупный кластер (выделен красным цветом) содержит ключевые слова «здоровье» (health), «социальный капитал» (social capital), «ментальное здоровье» (mental health), «сообщество» (community), «воздействие» (impact), «образование» (education), «детерминанты» (determinants) и др. Данный кластер демонстрирует наличие связи между исследованиями здоровья как такового и социального капитала. Ключевые слова в данном кластере позволяют его охарактеризовать с позиций исследования социоэкономических детерминант (факторов), предопределяющих здоровье.
2. Следующий по размеру кластер (выделен синим цветом) включает ключевые слова «смертность» (mortality), «распространенность» (prevalence), «связь» (association), «риск» (risk), «ожирение» (obesity), «сердечно-сосудистые заболевания» (cardiovascular disease), «индекс массы тела» (body mass index), «удар» (stroke), «атеросклероз» (atherosclerosis), «управление» (management), «гипертензия» (hypertension), «мета-анализ» (meta-analysis) и т. д. Данный кластер представлен исследованиями рисков и причин смертности, ухудшения здоровья из-за сердечно-сосудистых заболеваний, а также рекомендациями, как снизить эти риски.
3. «Зеленый» кластер содержит ключевые слова «болезнь» (disease), «воспаление» (inflammation), «рак» (cancer), «терапия» (therapy), «болезнь Альцгеймера» (Alzheimer's disease) и т. д. Данный кластер можно охарактеризовать как исследования в сфере онкологических и нейродегенеративных заболеваний.



Источник: составлено автором.
Source: compiled by the author.

Рис. 2. Кластерная визуализация методом полного и фракционного счета по теме «Капитал здоровья»
Fig. 2. Clustering data visualization (full and fractional counting) on the topic «Health capital»

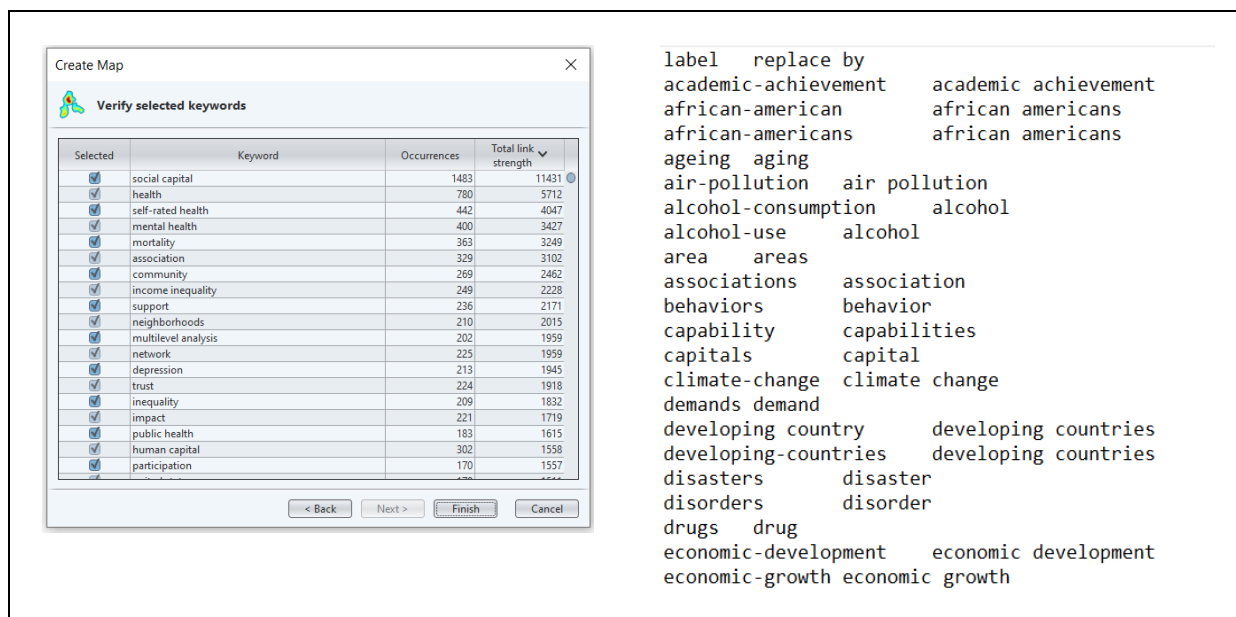
4. Кластер с ключевыми словами «депрессия» (depression), «пожилые люди» (older adults), «симптомы» (symptoms), «подростки» (adolescents), «расстройством» (disorder), «качество жизни» (quality of life), «инвалидность» (disability), «алкоголь» (alcohol), «биполярное расстройство» (bipolar disorder) и т. п. Данный кластер включает исследования психоэмоциональных расстройств у разновозрастных групп населения; влияние употребления алкоголя и курения на состояние здоровья; влияние депрессии и бессонницы на состояние здоровья и т. д.
5. Самые значимые ключевые слова самого маленького кластера – «дети» (children), «эпидемиология» (epidemiology), «подверженность» (exposure), «инфекция» (infection), «астма» (asthma), «тренды» (trends) и т. д. – позволяют предположить, что этот кластер представлен эпидемиологическими исследованиями. Ключевое слово «дети» характеризует эту возрастную категорию как наиболее уязвимую в эпидемиологическом контексте. Интересное наблюдение для этого кластера – это появление таких значимых слов, как «Китай» (China), COVID-19, что, конечно же, связано с пандемией коронавируса. Именно эти два ключевых слова в данном кластере на темпоральной карте являются самыми «молодыми» хронологически (рис. 3).

Отметим, что для целей построения карты ключевых слов были использованы необработанные (неочищенные) данные. Тем не менее на этом этапе вполне успешно в программе были выделены кластеры из взаимосвязанных ключевых слов, которые достаточно легко подвергаются интерпретации. В основе операционализации понятия «капитал здоровья» лежат два самостоятельных термина – «здоровье» и «капитал». В данном случае в результате библиометрического анализа соприсутствия ключевых слов выстраивается логическая взаимосвязь здоровья с группами заболеваний и дисфункций. Отметим тот факт, что никакие другие коннотации термина «капитал» не присутствуют в списке. Все варианты «капитала» содержательно связаны в смысловом поле с термином «здоровье». Также кластерный анализ показал, что теория социального капитала (первый кластер) является важным рамочным понятием для концептуальной проработки литературы по исследованию капитала здоровья. Даже в таком виде с использованием необработанных выгруженных данных при реализации очень простых шагов в программе VOSviewer построение карты ключевых слов помогает выявить научные направления и подходы для проведения систематического литературного обзора и обозначить шаги для дальнейшей работы в построении концептуальной модели объекта исследования. На рис. 3 представлены карты для полного и фракционного счета в темпоральном разрезе – ключевые слова связываются с датой публикации. Отличия в данном случае также незначительны.

Этап 2. Построение карты ключевых слов на основе 3000 максимально релевантных публикаций

Для этого этапа построение карты ключевых слов реализовывалось по 3000 максимально релевантным публикациям из первоначального списка. Всего в данном наборе данных было выделено 8333 ключевых слова, при выставлении барьера в минимум пять слов (ключевое слово в наборе данных появлялось минимум 5 раз) количество ключевых слов сокращалось до 902.

На этом этапе исследования проводилась очистка данных – объединение одинаковых ключевых слов с разным написанием в один семантический конструкт. Например, mental-health и mental health объединялись в mental health, generalised trust и generalized trust – в generalized trust, older-people, older-adults, older people, elderly-people, elderly, older adults – в older adults. Применение тезауруса ключевых слов сократило набор данных до 8242 слов всего и 822 слов с барьером минимум пять слов.



Источник: составлено автором.

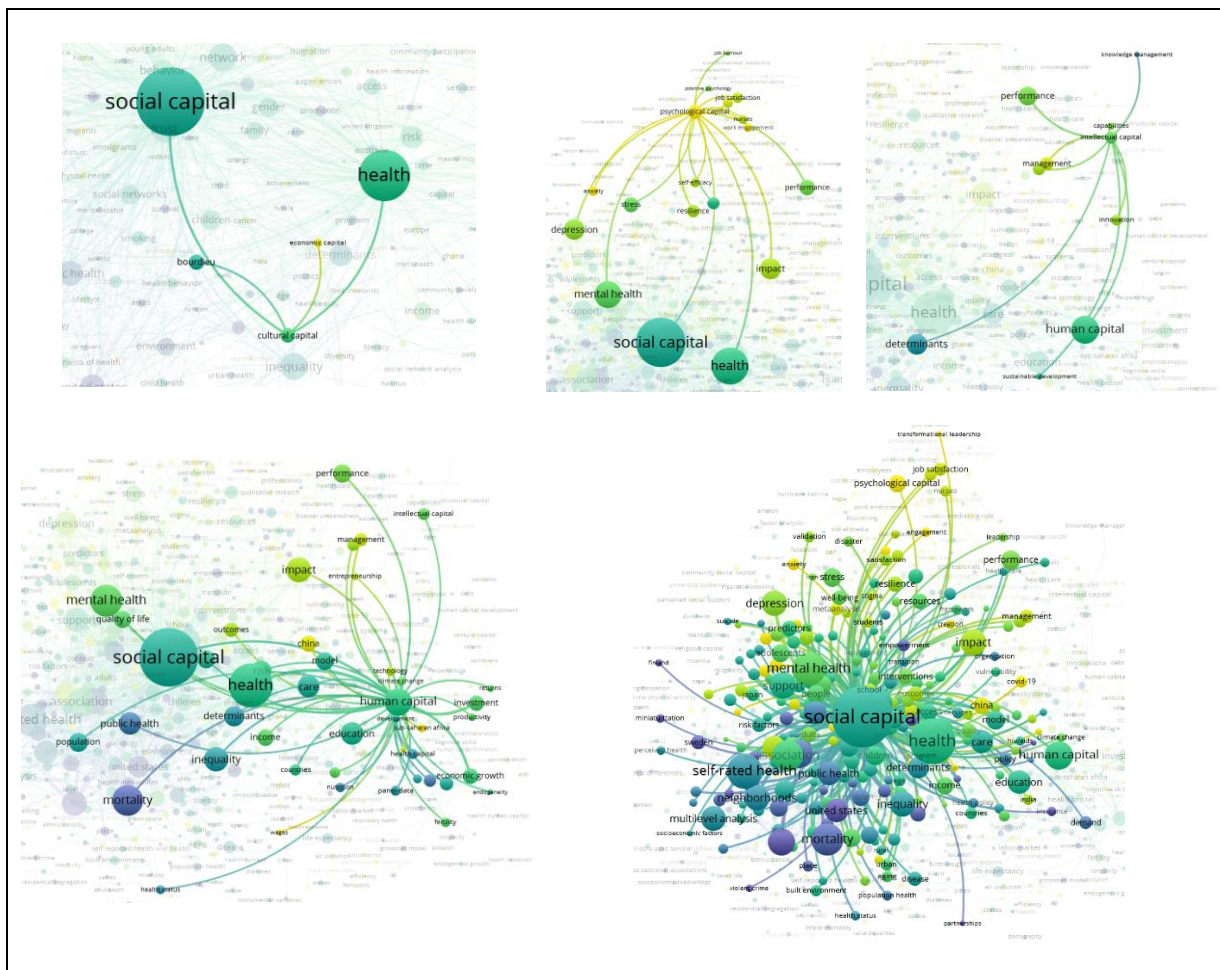
Source: compiled by the author.

Рис. 4. Топ ключевых слов после обработки данных и тезаурус ключевых слов
Fig. 4. Top of the keywords list after data cleaning and thesaurus of terms

При построении карты ключевых слов по методу полного и дробного счета значительной разницы в визуализации также не наблюдалось, как и на первом этапе. Данная карта позволяет определить термин «социальный капитал» (social capital) как наиболее значимый среди всех ключевых слов (рис. 5). Дополнительно на этом этапе исследования пул ключевых слов, связанных с капиталом здоровья, можно расширить за счет следующих терминов: человеческий капитал (human capital), психологический капитал (psychological capital), интеллектуальный капитал (intellectual capital), культурный капитал (cultural capital), социально-экономический статус (socio-economic status), самооценка здоровья (self-rated health), ментальное здоровье (mental health) и др. В темпоральном разрезе на представленной карте ключевых слов обращает на себя направление исследований психологического капитала как наиболее молодое (рис. 5).

На рис. 6 представлена связь ключевых слов для разных видов капитала. В качестве основного узла выбраны термины, обозначающие разные виды капитала (социальный капитал, человеческий, психологический, интеллектуальный и культурный). Для каждого из представленных видов капитала показана связь с другими ключевыми словами в темпоральном разрезе. Синий цвет на рисунках обозначает устоявшиеся в исследованиях ключевые слова (традиционные, «старые» работы), желтый цвет обозначает более «молодые» ключевые слова в публикациях.

По результатам представленного выше библиометрического анализа соприсутствия ключевых слов можно предположить, что систематический обзор литературы требует анализа публикаций по нескольким научным направлениям – социология, экономика и управление, медицина, что делает операционализацию понятия капитала здоровья непростой задачей. Если капитал в целом рассматривать как набор определенных ресурсов, которые могут быть использованы индивидом и обществом для реализации целей, то результат построения карты ключевых слов по теме «Капитал здоровья» показал, что в анализе можно задействовать несколько форм капиталов.



Источник: составлено автором.
Source: compiled by the author.

Рис. 6. Связи ключевых слов (терминов, характеризующих разные виды капитала) по теме «Капитал здоровья»

Fig. 6. Links of keywords (terms specifying different types of capital) on the topic «Health capital»

Так, например, изначально социальный капитал представлен как объект исследования в социологии. В социологической теории социальный капитал изучался такими авторами, как Патнэм и Бурдьё, чьи взгляды на данное явление не совпадали. Концепция социального капитала Патнэма состоит из трех компонентов: моральных обязательств и норм, социальных ценностей (особенно доверия) и социальных сетей (особенно добровольных ассоциаций). Такое представление социального капитала и доверия связано с механизмами, укрепляющими интеграцию ценностей общества, что создает консенсус и поддерживает стабильное развитие [15]. Концепция Бурдьё связана с его теоретическими представлениями о классе. Он выделяет три измерения капитала, каждое из которых имеет свое собственное отношение к классу: социальный капитал, экономический капитал и культурный капитал. Эти три ресурса становятся социально эффективными, а право собственности на них узаконивается через посредничество символического капитала. Ключевой особенностью символического капитала является воплощение в нем других форм капитала, которое воспринимается другими индивидами как отдельная смысловая единица. Социальный капитал подразумевает наличие у

индивида слабых и сильных социальных связей [16]. Бурдьё отмечал, что объем социального капитала рассчитывается как объем других форм капитала, который индивид может привлечь для достижения поставленных целей, благодаря социальным связям. На современном этапе исследований отмечается, что социальный капитал следует рассматривать в его неотъемлемой связи с социально-экономическим статусом индивида и средовыми социально-экономическими условиями, в которых живут люди [17].

Также концептуализация капитала здоровья может быть построена на противопоставлении теории человеческого капитала, получившей развитие в рамках Чикагской школы экономики, и концепции Бурдьё с его пониманием социального поля и социального капитала [18]. Таким образом, сравнение капитала здоровья реализуется в рамках парадигм двух научных дисциплин – социологической и экономической теории. В теории человеческого капитала G. Becker рассматривает здоровье как неотъемлемую часть человеческого капитала [19], так как здоровье является ключевым компонентом благосостояния и уровня жизни человека. Однако, по мнению A. Schneider-Kamp, точка зрения Бурдьё потенциально более плодотворна для отражения важности социальных и культурных аспектов в построении индивидуального здоровья, в связи с чем капитал здоровья можно рассматривать как совокупность фактических или потенциальных ресурсов, которыми обладает агент и которые способны влиять на положение агентов в социальной сфере здоровья [18].

Выводы и заключение

Программа VOSviewer представляет собой бесплатный удобный программный продукт, облегчающий работу исследователя. При минимальном освоении программы с минимальными временными затратами возможно собрать данные и построить карту ключевых слов, что поможет в проведении расширенного систематического литературного обзора. Построение карты ключевых слов может реализовываться как предварительный этап до проведения систематического обзора, поскольку такой предварительный анализ поля исследования через ключевые слова направляет и ускоряет процесс литературного обзора по нужной научной тематике. Такая карта – сеть ключевых слов – способствует расширению охвата теорий, концепций, подходов за счет визуализации ключевых слов проблемного поля. Этот подход является довольно популярным для изучения знаниевой структуры разных научных дисциплин и исследовательских тематик. Например, в работе [12] представлен анализ соприсутствия ключевых слов для картирования публикаций в сфере технологического форсайта (technology foresight) – визуализация структуры знаний, включающей в себя такие элементы, как распределение ресурсов НИОКР, анализ результатов НИОКР и т. д. В работе [20] используется картирование ключевых слов для оценки структуры региональной инновационной системы – какие элементы региональной инновационной системы анализируются в публикациях, в каких журналах представлены работы, в каких странах работают авторы публикаций. В статье [21] обсуждается анализ соприсутствия ключевых слов для изучения литературы по применению метода DEA (data envelopment analysis, оболочечный анализ), который является одним из наиболее популярных методов для оценки эффективности организационных единиц. В указанной работе моделируется эволюция отношений между понятиями в области использования метода DEA, оценивается динамика структурных изменений в данной области знаний и анализируются приоритетные направления исследований за 2008–2017 гг. VOSviewer является далеко не единственным способом построения карты ключевых слов. Например, для анализа соприсутствия ключевых слов по применению метода DEA в указанной выше работе [21] использует-

ся программный язык R (пакеты *igraph* и *powerLaw*) для расчета показателей и Gephi для визуализации полученных сетей.

В случае представленного в данной работе примера с построением карты ключевых слов по теме «Капитал здоровья» показано, что даже с необработанными данными при помощи VOSviewer можно построить кластеры, которые благодаря визуальному представлению проблемного поля через ключевые слова способствуют формированию шагов исследования по выявленным направлениям.

На необработанных данных из 57853 публикаций удалось сформировать пять терминологических кластеров, один из которых показал связь здоровья и социального капитала. На втором этапе при анализе обработанных данных из 3000 наиболее релевантных публикаций удалось выделить термины, обозначающие разные виды капиталов (социальный капитал, человеческий, культурный и т. д.), а также предположить эволюцию исследований по теме капитала здоровья через темпоральные характеристики построенных сетей – дифференцию «старых» и «новых» ключевых слов. Построение карт методами прямого и фракционного счета не показало существенных отличий при визуализации данных. Применение разных методов счета может быть важным и обоснованным в случае построения сетей для цитирований, со-цитирований, библиографических сочетаний, так как может привести к получению разных результатов. Однако, очевидно, что в случае анализа соприсутствия – совместного использования ключевых слов – результаты по теме «Капитал здоровья» графически практически не отличались.

Исследование выполнено при финансовой поддержке Российского научного фонда в рамках научно-исследовательского проекта «Институты реализации ресурсного потенциала старшего поколения в экономике старения» (проект № 19-18-00300, <https://rscf.ru/project/19-18-00300/>).

СПИСОК ЛИТЕРАТУРЫ/REFERENCES

1. *Bibliometric analysis of scientific networks. International laboratory for Applied Network Research, Higher School of Economics.* Available at: https://anr.hse.ru/en/soc_prj3/ (accessed 20 March 2023).
2. Batagelj V., Cerinšek M. On bibliographic networks. *Scientometrics*, 2013, no. 96, pp. 845–864. Available at: <https://doi.org/10.1007/s11192-012-0940-1> (accessed 20 March 2023).
3. Maltseva D., Batagelj V. Towards a systematic description of the field using keywords analysis: main topics in social networks. *Scientometrics*, 2020, no. 123 (1), pp. 357–382.
4. Highlights. VOSviewer – Visualizing scientific landscapes. Available at: <https://www.vosviewer.com/features/highlights> (accessed 20 March 2023).
5. Van Eck N.J., Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*. 2010, no. 84, pp. 523–538. Available at: <https://doi.org/10.1007/s11192-009-0146-3> (accessed 20 March 2023).
6. Van Eck N.J., Waltman L. VOS: a new method for visualizing similarities between objects. *Advances in data analysis. Proceedings of the 30th annual conference of the German Classification Society.* Eds. H.-J. Lenz, R. Decker. Heidelberg, Springer, 2007. pp. 299–306.
7. Van Eck N.J., Waltman L. Bibliometric mapping of the computational intelligence field. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2007, no. 15 (5), pp. 625–645.
8. Peters H.P.F., Van Raan A.F.J. Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling. *Research Policy*, 1993, no. 22 (1), pp. 23–45.
9. Zitt M., Bassecoulard E., Okubo Y. Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics*, 2000, no. 47 (3), pp. 627–657.
10. Van Eck N.J., Waltman L. How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American society for information science and technology*, 2009, no. 60 (8), pp. 1635–1651.
11. Perianes-Rodriguez A., Waltman L., Van Eck N.J. Constructing bibliometric networks: a comparison between full and fractional counting. *Journal of informetrics*, 2016, no. 10 (4), pp. 1178–1195.

12. Su H.N., Lee P.C. Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight. *Scientometrics*, 2010, no. 85 (1), pp. 65–79.
13. Radhakrishnan S., Erbis S., Isaacs J.A., Kamarthi S. Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. *PloS one*, 2017, no. 12 (3), e0172778. Available at: <https://doi.org/10.1371/journal.pone.0185771> (accessed 20 March 2023).
14. Waltman L., Van Eck N.J., Noyons E.C. A unified approach to mapping and clustering of bibliometric networks. *Journal of informetrics*, 2010, no. 4 (4), pp. 629–635.
15. Siisaiainen M. Two concepts of social capital: Bourdieu vs. Putnam. *International journal of contemporary sociology*, 2003, no. 40 (2), pp. 183–204.
16. Granovetter M.S. The strength of weak ties. *American journal of sociology*, 1973, vol. 78, no. 6, pp. 1360–1380.
17. Carpiano R.M. Toward a neighborhood resource-based theory of social capital for health: can Bourdieu and sociology help? *Social science & medicine*, 2006, no. 62 (1), pp. 165–175.
18. Schneider-Kamp A. Health capital: toward a conceptual framework for understanding the construction of individual health. *Social Theory & Health*, 2021, no. 19 (3), pp. 205–219.
19. Becker G.S. Health as human capital: synthesis and extensions. *Oxford economic papers*, 2007, no. 59 (3), pp. 379–410.
20. Lee P.C., Su H.N. Investigating the structure of regional innovation system research through keyword co-occurrence and social network analysis. *Innovation*, 2010, no. 12 (1), pp. 26–40.
21. Lozano S., Calzada-Infante L., Adenso-Díaz B., García S. Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature. *Scientometrics*, 2019, no. 120, pp. 609–629.

Поступила: 12.04.2023 г.

Принята после рецензирования: 07.06.2023 г.

UDC 004.65:025.4.03:330.14:331.101.262:61

BUILDING A KEYWORDS CO-OCCURRENCE MAP ON THE TOPIC «HEALTH CAPITAL» IN THE VOSVIEWER PROGRAM

Irina A. Pavlova,
iapav@mail.ru

National Research University «Higher School of Economics»,
20, Myasnitskaya street, Moscow, 101000, Russia

Irina A. Pavlova, Cand. Sc., senior researcher, National Research University «Higher School of Economics».

Relevance. Bibliometric analysis using network analysis methods is an actively developing scientific research area. Its methods and software tools contribute to progress in understanding how scientific disciplines develop. VOSviewer is an open software product that can be used to build bibliometric maps (networks) of various types (citation, co-authorship, co-occurrence, co-citation, etc.). Bibliometric analysis of the co-presence (co-occurrence) of keywords (simultaneous presence of keywords in different publications) allows building a map of keywords, which can be considered, under certain assumptions, as a model of the knowledge structure in the analyzed area. **The objective** of this study is to build a keyword map of the research field on the topic «health capital» based on the data of the Web of Science database, as well as to analyze the process of building a bibliometric network of keywords co-occurrence in the VOSviewer program. **Methods:** bibliometric analysis with methods of network analysis. The materials of the Web of Science database were used as an empirical basis for the study (57853 publications found by the words of health and capital and downloaded in spring 2022 in files of 500 publications each in full record format). To build a keyword map, the VOSviewer program, version 1.6.19, was used. **Results.** At the first stage, when constructing a keyword map based on the entire array of 57853 publications without preliminary data processing (data cleaning), five clusters were identified and analyzed. At the second stage, a keyword map was built based on 3000 most relevant publications with data processing (data cleaning through the construction of a thesaurus). As a result, keywords that characterize different types of capital (social capital, human capital, cultural capital, psychological capital) are identified and analyzed. **Conclusions.** Keyword mapping can be done as a preliminary step before conducting a systematic review, since such a preliminary analysis of the field of study through keywords guides and accelerates the process of literature review on the desired scientific topic. Such a map – a network of keywords – helps to enhance the coverage of theories, concepts, approaches by visualizing the keywords of the research field. The example presented in the work on the topic «Health capital» shows that even with «raw» (uncleaned) data, clusters can be built, which, thanks to the visual representation of the research field through keywords, contribute to the building further research steps. The bibliometric analysis of the keywords co-occurrence made it possible to present a structural model of the knowledge in the domain of the research topic of health capital, i. e. (1) to map thematic clusters; (2) build temporal networks to analyze the evolution of knowledge in the chosen scientific field; (3) to identify different types of capitals, within which the health capital is further analyzed and a systematic literature review is elaborated.

Key words: bibliometric analysis, health capital, VOSviewer, network analysis, keyword co-occurrence.

The study was financially supported by the Russian Science Foundation as part of the research project «Institutions for Realizing the Resource Potential of the Older Generation in the Aging Economy» (project No. 19-18-00300, <https://rscf.ru/project/19-18-00300/>).

Received: 12 April 2023.

Reviewed: 7 June 2023.