

Аспектный анализ данных систематизации документного фонда библиотек и научный ландшафт

Сысоев А.Н.,

главный специалист НМО, Всероссийский институт научной и технической информации Российской академии наук (ВИНИТИ РАН), Москва, Россия, vims-sysoev@mail.ru

Белоозеров В.Н.,

кандидат филологических наук, доцент, ведущий научный сотрудник, Всероссийский институт научной и технической информации Российской академии наук (ВИНИТИ РАН), Москва, Россия, systemling@narod.ru

Аннотация. В работе обсуждается метод аспектного анализа классификационных индексов документного фонда библиотек. Метод позволяет установить связи наук и информационные потоки между ними, что определяет ландшафт научного знания. Объектом анализа служит совокупность составных классификационных индексов содержащих указатели на релевантные области знания в качестве аспектов. Предложенный подход может быть использован в наукометрических исследованиях.

Ключевые слова: иерархическая классификация, многоаспектное индексирование, классификационный индекс, связи наук, междисциплинарность

Aspect Analysis of Data Systematization of Library's Document Collection, and Scientific Landscape

Sysoev A.N.,

Russian Institute for Scientific and Technical Information (VINITI RAS)

Moscow, Russia, vims-sysoev@mail.ru

Beloozerov V.N.,

Russian Institute for Scientific and Technical Information (VINITI RAS)

Moscow, Russia, systemling@narod.ru

Annotation. The paper discusses the method of aspect analysis of classification indexes of the document collection of libraries. The method allows you to establish links between the Sciences and information flows between them, which determines the landscape of scientific knowledge. The object of analysis is a set of composite classification indexes containing pointers to relevant areas of knowledge as aspects. The proposed approach can be used in scientometric research.

Keywords: hierarchical classification, multi-aspect indexing, classification index, connections of Sciences, interdisciplinarity.

DOI: 10.31432/1994-2443-2020-15-3-63-70

Цитирование публикации: Сысоев А. Н., Белоозеров В. Н. Аспектный анализ данных систематизации документного фонда библиотек и научный ландшафт // Информация и инновации. 2020, Т. 15, № 3. с. 63–70. DOI: 10.31432/1994-2443-2020-15-3-63-70

Citation: Sysoev A.N., Beloozerov V.N. Aspect analysis of data systematization of library's document collection, and scientific landscape // Information and Innovations 2020, T. 15, № 3. p. 63–70. DOI: 10.31432/1994-2443-2020-15-3-63-70

Связь наук и научные связи — понятия, за которыми скрывается исследовательский интерес к состоянию науки, но это не одно и то же. Под **связью наук** предполагается существование характеризующейся устойчивостью тен-

денции к диалогу между отдельными областями знания и дисциплинами, составляющими науку в целом. Этот диалог осуществляется на уровне процессов обмена научным знанием, методами и накопленными фактами. Управляющие этими

процессами законы лежат в области гносеологии, эпистемологии, а осуществляются путем смены парадигм, концепций и во многих отношениях сегодня не изучены. Персонификация названного феномена на этом уровне нелепа: она выразима лишь в образах отдельных научных дисциплин.

Другое дело **научные связи**, под которыми понимают тенденцию к диалогу между отдельными исследователями, научными коллективами и предприятиями. Однако, эти связи не могут характеризоваться устойчивостью поскольку диктуются прагматическими и сиюминутными сообщениями, зачастую прямо или опосредованно определяют карьерный рост, заработную плату и доступ к грантам для множества ученых [1]. Процесс такого диалога хотя и лежит в плоскости интересов научного сообщества, но протекает на уровне отдельных исследователей и обсуждения текущих научных проблем. Управляется он законами социологии и научной коммуникации, а осуществляется через печатные издания, личные контакты, научный диалог. Процессы на этом уровне всегда персонифицированы и соотносятся с научными школами, предприятиями и отдельными исследователями.

Через статистические исследования текстов научных работ, исследуя структуру научного окружения ученого и динамику развития «незримых» научных коллективов [2; 3], исследователи пытаются оценивать тенденции в развитии различных научных направлений, идентифицируют персоны, научные предприятия и организации, научные школы, устанавливая взаимосвязи между отдельными научными сообществами. Двигаясь в направлении от отдельных статей и изданий к характеристике областей знания, исследователи с помощью сложных пересчетных коэффициентов пытаются выявить связи между науками [4] получая в результате «карты» науки, научный «ландшафт» и возвышаются до предсказания развития науки. Исходными данными для этих умозаключений служат, накопленные и систематизированные информационными ритейлерами, библиографические ссылки на изданные статьи, монографии и иные естественнонаучные и научно-технические издания. Однако, изначально эти документы предопределенно размещаются в поле отдельных областей знания, отражая тематику, декларированную издательствами

и журналами. Издателями, для обоснования этой предопределенности создаются упрощенные рубрикаторы областей знания. Так в WoS Subject Categories, каждый журнал приписывается к одной или нескольким тематическим рубрикам, а публикации, находящиеся в нем, автоматически получают все рубрики, приписанные журналу. Такое неоправданное упрощение структуры научного знания при общении с каталогом не может не вызывать курьёзов при предметном поиске. Всё чаще высказывается мнение, что на сегодняшний день междисциплинарные сравнения по абсолютным значениям импакт-факторов (числа цитирований) ни в WOS, ни в SCOPUS не только неправильны, но попросту невозможны [5; 6;].

Но не озабоченный вопросами освоения международных классификаций для непонятной ему цели международной межбиблиотечной коммуникации информационный ритейл оставляет эти проблемы библиотекам, подменяя проблему коммуникации электронных каталогов проблемой совершенствования собственных систем и оставаясь нацеленным на достижение коммуникации непосредственно с потребителем распространяемой информации, минуя библиотеки. Однако потенциал библиотечных фондов не исчерпан и содержит нереализованные сегодня возможности. Так, для естественнонаучных библиотек свойственно индексирование документов составными индексами, которые включают коды двух или более классов Универсальной десятичной классификации (УДК), представляющие разные аспекты содержания документа, относя его к разным научным сферам. Процесс создания составного индекса и применения аспектных ссылок подробно изложен в литературе по библиотечному делу [7;8;9] и регламентирован соответствующими стандартами [10]. Анализируя документ и определяя его тему, систематизатор должен найти все классы УДК, по тематике которых документ содержит существенные сведения. И если найденные индексатором классы относятся к разным иерархическим ветвям классификации, он принимает решение об индексировании документа двумя или несколькими классами, коды которых соединяются в составной индекс, указывающий смежные области знаний, перечисляя их через двоеточия (рис. 1).

аспектные ссылки

ТЕМА:АСПЕКТ:АСПЕКТ:АСПЕКТ

Рис. 1 Структура составного индекса при многоаспектном индексировании

Крайний слева индекс отмечает главную тему документа и его положение в классификации общепринятых областей знания, тогда как последующие индексы, отделённые от него двоеточием, характеризуют отдельные аспекты знания, к которым он адресуется. Многоаспектное индексирование относится к приёмам углублённого анализа документов и осуществляется систематизаторами-предметниками естественнонаучных библиотек при обработке входящих изданий, когда при индексировании систематизатор прямо указывает смежные области знаний. Например,

1405; **615:001.4:811.124(07)**; Шиканова Татьяна Аркадьевна. Латинский язык и основы фармацевтической терминологии : Учебное пособие для мед. вузов / Шиканова Т.А.; Сиб. гос. мед. ун-т. Кафедра лат. яз. и мед. терминологии. — Томск : Печат. мануфактура, 2004. — 128 с. — ISBN 5-94476-043-5.

144084; **612.6:575.17:004.9**; Handbook on analyzing human genetic data : computational approaches and software / Lin Shili, Zhao Hongyu ed. — Heidelberg etc. : Springer, cop. 2010. — XIV, 333 с. : ил., табл. — Библиография в конце ст. указ.: с. 331-333. — ISBN 978-3-540-69263-8; Руководство по анализу генетических данных человека: методы вычислений и программное обеспечение.

Анализ составных индексов накопленных в документном фонде библиотек, составляет сущность аспектного анализа. Рассмотрим его применение на примере корпуса данных открытого библиотечного каталога БЕН РАН использующего в практической работе Универсальную Десятичную Классификацию (УДК). По состоянию на конец 2019 г., каталог содержал 374957 индексов систематизации, из которых 296568 не использовали аспекты,

а 78389 индексов были составлены с использованием приёмов многоаспектного индексирования. Для систематизации 224748 документов в аспектных ссылках было использовано 81853 индексов УДК. Отношение общего количества документов к количеству аспектных ссылок позволяет количественно оценить долю их междисциплинарности. В нашем случае (81853 / 374957) она составляет 20% и, учитывая естественнонаучный профиль библиотеки за 30-летний период, можно говорить о том, что 80% естественнонаучных изданий не характеризуются ярко выраженной тенденцией междисциплинарности. Примечательно, что полученные цифры повторяют эмпирически утверждённый для многих прикладных областей принцип Парето.

Чтобы избежать заблуждений, которые могут возникнуть при обработке данных каталога из-за недокомплектованности по географии, медицине, истории и другим дисциплинам комплементарным естественным наукам мы не будем выходить за пределы УДК индекса **5 Естественные науки и Математика** профильного исследуемому каталогу. Данные для представления связей наук мы будем черпать непосредственно из индексов систематизации электронного каталога, которые заключены в библиографических записях. В ходе анализа мы будем опираться на уровневые законы иерархической классификационной системы УДК и последовательно покажем шаги аспектного анализа от нулевого до первого уровня этой классификационной системы. Здесь следует заметить, что приложение метода возможно в рамках любой универсальной иерархической классификационной схемы любого библиотечного электронного каталога, применяющего многоаспектное индексирование.

Анализ следует начинать с рассмотрения распределения корпуса данных, чтобы установить профильность используемых выборок и их соответствие профилю анализируемого каталога. Тогда степень рассеяния профильной информации по мере погружения в иерархическую структуру может служить мерой углублённости индексирования. Так распределение на рис.2 характеризует не только профильность тематики документов генеральной совокупности естественнонаучного каталога БЕН РАН, но также даёт представление об интенсивности

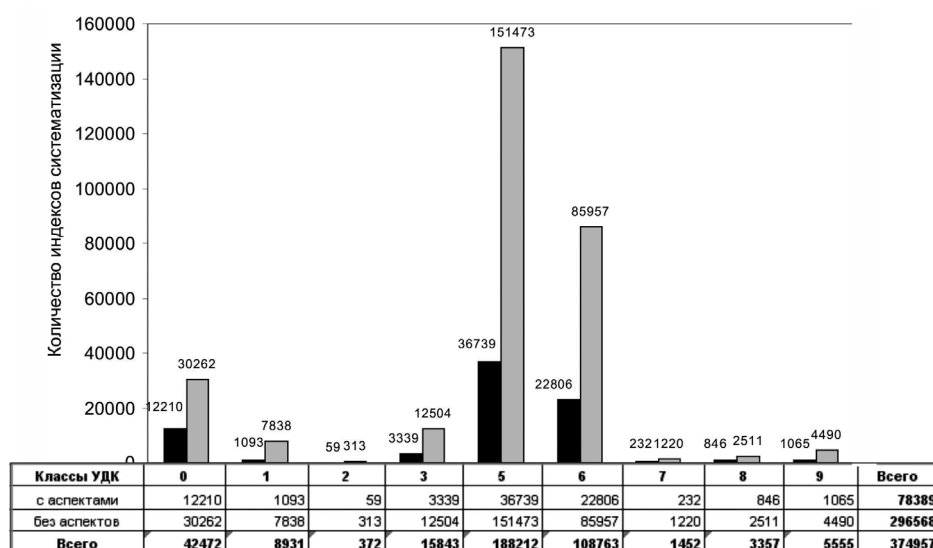


Рис. 2 Распределение индексов УДК по классам нулевого (верхнего) уровня классификации. (Тёмные столбцы — составные индексы; светлые — безаспектные индексы; цифры над столбцами — абсолютное число использованных индексов из классов УДК, номера которых указаны на оси абсцисс)

обмена информацией для каждого класса анализируемого корпуса данных.

Следующим шагом является составление матрицы перекрёстных ссылок (табл.1), в которой участвуют данные подсчёта количества аспектов обнаруженных в составных индексах по каждому из классов УДК нулевого уровня. Наукометрический смысл матрицы заключается в детализации содержания составных индексов (тёмные столбцы) на рис. 2 и служит ос-

новой для выводов о направленности и интенсивности информационных потоков на нулевом иерархическом уровне в объёме всего корпуса данных. Такая обобщённая оценка призвана характеризовать динамику информации в рамках областей знания и комплексных научных направлений.

Главная диагональ матрицы показывает количество ссылок не выходящих за пределы тематики класса.

Таблица 1

Матрица перекрёстных аспектных ссылок нулевого (верхнего) уровня. Выделена главная диагональ и данные по классу УДК 5 Естественные науки и Математика. (Цифры — абсолютное количество аспектных ссылок)

Количество аспектных ссылок по классам											
Классы	0	1	2	3	5	6	7	8	9	Всего	
0	8387	268	6	877	1844	1369	215	117	125	13208	
1	133	415	20	194	121	133	18	48	28	1110	
2	9	14	0	8	6	1	4	11	8	61	
3	1146	270	11	791	506	579	11	59	81	3454	
5	6361	757	10	1366	23465	5526	55	20	672	38232	
6	4152	319	4	986	8454	9505	125	25	83	23653	
7	83	8	5	7	47	71	9	1	7	238	
8	163	86	9	96	137	292	6	41	32	862	
9	391	14	6	63	336	96	17	32	80	1035	
Всего	20825	2151	71	4388	34916	17572	460	354	1116	81853	

Опираясь на данные таблицы 1, отображающей в количественном выражении взаимодействие между областями знания, мы можем графически изобразить информационные потоки в пределах анализируемого уровня (рис. 3). Так для класса УДК 5 отмечается высокая потребность обмена информацией с приклад-

ными науками и классом 0, куда входят, в том числе, компьютерные науки. Очевидна нарастающая потребность в экономической информации из класса 3. Эта потребность за анализируемый период времени превысила по количеству аспектных ссылок постоянный поток историко-географической информации из класса 9.

К Л А С С Ы	5	
	темы	аспекты
	0	1844 6361
	1	121 757
	2	6 10
	3	506 1366
	5	23465 23465
	6	8454 5526
	7	47 55
	8	137 20
	9	336 672

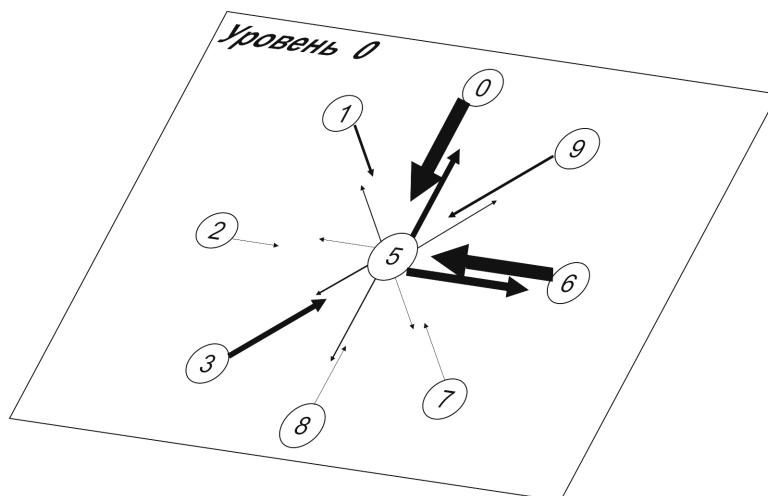


Рис. 3 Взаимодействие областей знания, соответствующих главным классам УДК с классом 5 Естественные науки и Математика. Стрелками показаны информационные потоки между областями знаний

Дальнейшая детализация представлений о динамике информационных процессов происходит на иерархическом уровне 1, где области знания конкретизируются и класс 5 распадается

на десять самостоятельных разделов, отвечающих комплементарным естественнонаучным дисциплинам. Общее распределение индексов систематизации показано на рис. 4.

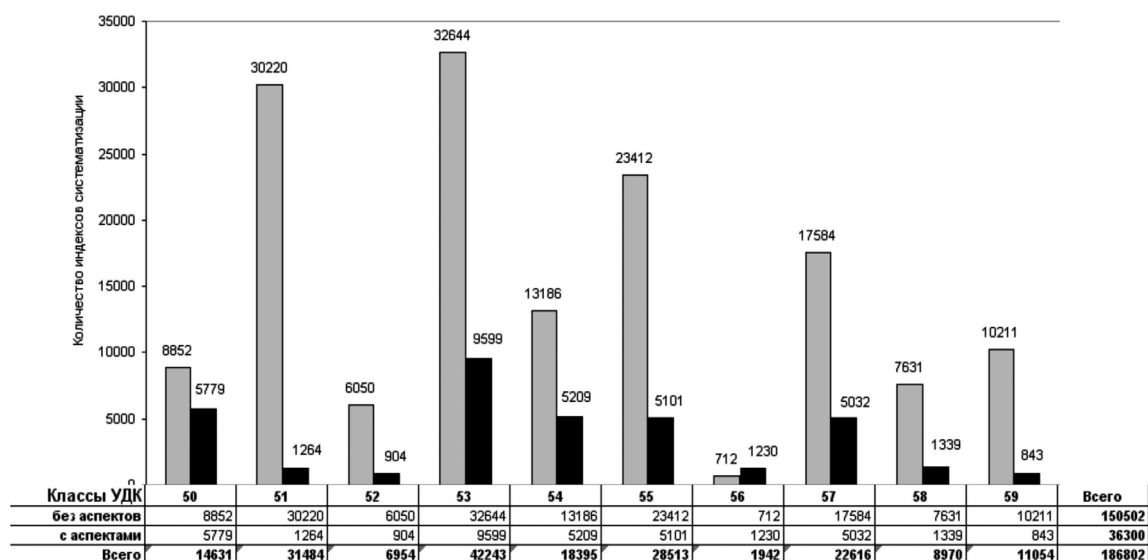


Рис. 4 Распределение индексов систематизации по десяти областям знания, соответствующим первому уровню класса УДК 5 Математика. Естественные науки. (Тёмные столбцы — аспектные индексы; светлое — безаспектные; цифры — абсолютные значения количества индексов).

По сравнению с рис. 2 распределение на иерархическом уровне 1 существенно выровнялось и все классы, кроме класса 50, стали отвечать конкретным естественнонаучным дисциплинам в привычной нотации. На этом фоне выделилась <56 Палеонтология>, систематизация которой содержит почти в два раза больше составных индексов против безаспектных. При этом количество ссылок в главной диаго-

нали таблицы 2 равно 1, что свидетельствует о практически полном отсутствии интереса к внутренним проблемам. Наблюдаемый эффект характеризует идеальную междисциплинарность: служебную дисциплину, занятую проблемами смежных наук. По тематике своей прямой деятельности она привлекает все науки, кроме Астрономии и Физики, а в своих ссылках опирается на Науку о Земле и Биологию.

Таблица 2

Матрица перекрёстных аспектных ссылок первого уровня. Выделена главная диагональ и данные по классу УДК 53 Физика. (Цифры — абсолютное количество аспектных ссылок)

		Количество аспектных ссылок по классам										
Темы документов по классам	Классы	50	51	52	53	54	55	56	57	58	59	Всего
	50	1014	27	78	45	411	223	2	436	450	418	3104
	51	2	436	7	287	5	12	0	6	0	0	755
	52	37	36	134	32	4	210	1	10	3	0	467
	53	0	2570	8	3407	297	10	0	5	1	1	6299
	54	115	173	8	351	3078	189	0	97	7	1	4019
	55	279	149	294	205	492	2005	26	109	16	1	3576
	56	6	173	0	0	3	819	1	316	92	31	1441
	57	95	186	46	356	663	123	21	620	537	466	3113
	58	28	6	19	4	46	29	0	297	95	5	529
	59	15	2	4	6	5	15	1	205	13	60	326
Всего		1591	3758	598	4693	5004	3635	52	2101	1214	983	23629

Аспектный анализ представляет собой анализ распределения количества аспектных ссылок по классам заданного иерархического уровня и последующее рассмотрение матрицы перекрёстных аспектных ссылок на том же уровне иерархии. Полнота картины связей наук определяется с одной стороны совокупностью ссылок отдельной науки на своё окружение, а с другой стороны ссылками соседних наук на исследуемую область знания или дисциплину. Очевидно, что если для классов, расположенных на одном иерархическом уровне, такого рода информация заключена в одной таблице перекрёстных ссылок, то для соседних междууровневых взаимодействий их потребуется уже две (рис.5). Если матрицу, представляющую связи сверху-вниз считать прямой, то вторая, представляющая связи снизу-вверх будет считаться обратной. Причём, если матрица на одном уровне всегда квадратная, то для междууровневых сравнений она может быть прямоугольной, так как число классов на других уровнях может различаться. Из схемы на рис.5 следует, что выполнение аспектного ана-

лиза возможно только в последовательности предопределённой иерархической структурой используемой классификации, и сопоставления 0 уровня с 2 уровнем — невозможны без учёта данных уровня 1.

Обсуждая методическую сторону использования аспектного анализа [11], авторы сознательно не касались вопросов интерпретации приведённого материала поскольку такого рода обсуждение представляет собой самостоятельное и глубокое исследование. Вместе с тем, представленный фактический материал позволяет даже при беглом взгляде получить далеко идущие выводы относительно связей и закономерностей взаимодействия дисциплин в естественнонаучной сфере. Аспектный анализ может служить основой оценки междисциплинарности и действенным инструментом выявления связи между науками. Привлечение данных индексирования, заключённых в библиотечных каталогах, использующих универсальные иерархические классификации, позволяет полнее использовать накопленный по-

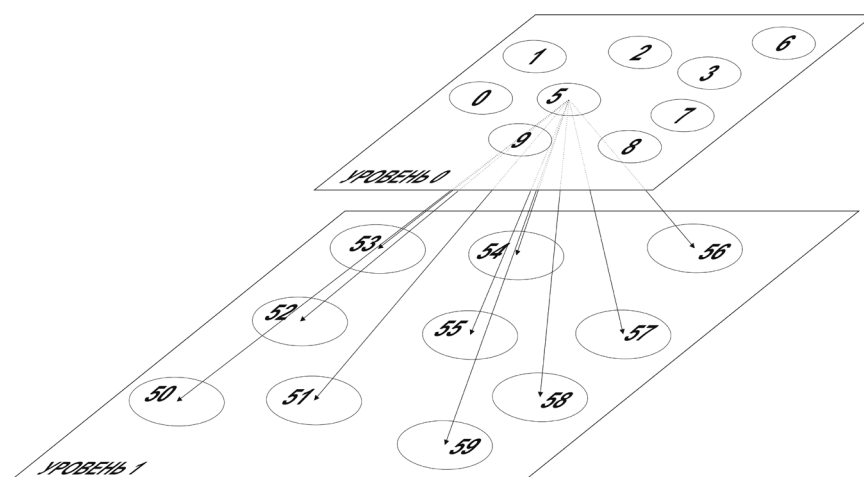


Рис. 5 Схема организации данных при проведении аспектного анализа для класса <5 Естественные науки и Математика> между уровнями иерархии

тенциал систематизации документных фондов библиотек.

ЛИТЕРАТУРА

1. Hicks D., Wouters P., Waltman L., Rijcke S. de, Rafols I. Bibliometrics: The Leiden Manifesto for Research Metrics // Nature. Vol. 520(7548). — 2015. — pp. 429–431. DOI:10.1038/520429a.

2. Леонова Ю.В., Федотов А.М. Извлечение знаний и фактов из текстов диссертаций и авторефератов для изучения связей научных сообществ / Труды 15-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL-2013, Ярославль, Россия, 14-17 октября 2013 г. с.32-41.

3. Девяткин Д.А., Швеиц А.В., Тихомиров И.А. Выявление направлений исследований и научных коллективов на основе анализа полнотекстовых коллекций научных публикаций // Материалы второго международного профессионального форума «Книга. Культура. Образование. Инновации». («Крым-2016»). — 2016. — с.302-306.

4. Панин С.Б. Современные наукометрические системы «WOS» и «SCOPUS»: издательские проблемы и новые ориентиры для Российской вузовской науки / Гуманитарные исследования Центральной России № 3 (12), 2019. — с.51-65.

5. Leydesdorff L. Diversity and interdisciplinarity: how can one distinguish and recombine disparity, variety, and balance? / Scientometrics (2018) 116:2113–2121; URL: [https://doi.org/10.1007/s11192-018-2810-y\(0123456789\(\),-volV\)\(0123456789\(\),-volV\)](https://doi.org/10.1007/s11192-018-2810-y(0123456789(),-volV)(0123456789(),-volV)).

6. Leydesdorff L., Wagner C.S., Bornmann L. Diversity measurement: Steps towards the measurement of interdisciplinarity? / Journal of Informetrics (2019. — т.13). — №3. — pp. 904-905.

7. Белоозеров В.Н., Астахова Т.С., Антошкова О.А., Смирнова О.В., Сукиасян Э.Р. Методическое пособие по Универсальной Десятичной Классификации. — М.: ВИНТИ, 2007. — 151 С.

8. Сукиасян Э. Р. Школа индексирования: практическое пособие / Э. Р. Сукиасян. — М.: Издательство «Либерея». — 2006. — 172 С.

9. Белоозеров В.Н. Структура и функции УДК // Научно-техническая информация, Сер.1. — 2006. — №10. — С. 5-8.

10. СИБИБ ГОСТ Р 7.0.66-2010 (ИСО 5963:1985) Индексирование документов. Общие требования к координатному индексированию = ISO 5963:1985 Documentation — Methods for examining documents, determining their subjects and selecting indexing terms (MOD). / Издание официальное / М.: ИПК Издательство стандартов. — 2011. — 16 С.

11. Сысоев А.Н., Белоозеров В.Н. Аспектный анализ классификационных индексов документов библиотечного фонда // Научно-техническая информация, Сер. 1. — 2020. — №8. — (в печати).

REFERENCES

1. Hicks D., Wouters P., Waltman L., Rijcke S. de, Rafols I. Bibliometrics: The Leiden Manifesto

for Research Metrics. // *Nature*. Vol. 520(7548). — 2015. — pp. 429–431. DOI:10.1038/520429a.

2. Leonova YU.V., Fedotov A.M. Izvlechenie znaniy i faktov iz tekstov dissertacij i avtoreferatov dlya izucheniya svyazey nauchnyh soobshchestv / *Trudy 15-j Vserossijskoj nauchnoj konferencii «Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kollekcii»* — RCDL-2013, YAroslavl', Rossiya, 14-17 oktyabrya 2013 g. s. 32-41.

3. Devyatkin D.A., SHvec A.V., Tihomirov I.A. Vyyavlenie napravlenij issledovaniy i nauchnyh kolektivov na osnove analiza polnotekstovykh kollekciy nauchnykh publikacij // *Materialy vtorogo mezhdunarodnogo professional'nogo foruma "Kniga. Kul'tura. Obrazovanie. Innovacii"*. («Krym-2016»). — 2016. — с.302-306.

4. Panin S.B. Sovremennyye naukometricheskie sistemy «WOS» i «SCOPUS»: izdatel'skie problemy i novye orientiry dlya Rossijskoj vuzovskoj nauki / *Gumanitarnye issledovaniya Central'noj Rossii* № 3 (12), 2019. — s.51-65.

5. Leydesdorff L. Diversity and interdisciplinarity: how can one distinguish and recombine disparity, variety, and balance? / *Scientometrics* (2018) 116:2113–2121; URL: [https://doi.org/10.1007/s11192-018-2810-y\(0123456789\(\),-volV\)\(0123456789\(\),-volV\)](https://doi.org/10.1007/s11192-018-2810-y(0123456789(),-volV)(0123456789(),-volV)).

6. Leydesdorff L., Wagner C.S., Bornmann L. Diversity measurement: Steps towards the measurement of interdisciplinarity? / *Journal of Informetrics* (2019. — t. 13). — № 3. — pp. 904-905.

7. Beloozerov V.N., Astahova T.S., Antoshkova O.A., Smirnova O.V., Sukiasyan E.R. Metodicheskoe posobie po Universal'noj Desyatichoj Klassifikacii. — M.: VINITI, 2007. — 151 S.

8. Sukiasyan E. R. SHkola indeksirovaniya: prakticheskoe posobie / E. R. Sukiasyan. — M.: Izdatel'stvo «Libereya». — 2006. — 172 S.

9. Beloozerov V.N. Struktura i funkcii UDK // *Nauchno-tekhnicheskaya informaciya*, Ser.1. — 2006. — № 10. — S. 5-8.

10. SIBID GOST R 7.0.66-2010 (ISO 5963:1985) Indeksirovanie dokumentov. Obshchie trebovaniya k koordinatnomu indeksirovaniyu = ISO 5963:1985. Documentation — Methods for examining documents, determining their subjects and selecting indexing terms (MOD) / *Izдание oficial'noe* / M.: IPK Izdatel'stvo standartov. — 2011. — 16 S.

11. Sysoev A. N., Beloozerov V. N. Aspektnyj analiz klassifikacionnykh indeksov dokumentov bibliotecnogo fonda // *Nauchno-tekhnicheskaya informaciya*, Ser. 1. — 2020. — № 8. — (v pechati).