

# Концептуализация данных наукометрических исследований в научных электронных библиотеках

© И.М. Зацман

Институт проблем информатики РАН  
im@a170.ipi.ac.ru

## Аннотация

Доклад посвящен проблеме концептуализации индикаторов сферы науки. Концептуализация индикаторов рассматривается как частный случай более общей проблемы обеспечения когнитивной интероперабельности в процессе совместной деятельности группы пользователей электронной библиотеки, то есть, обеспечение согласованного понимания ими информационных, программных и других объектов электронной библиотеки. Процессы концептуализации индикаторов анализируются в контексте совместной деятельности экспертов и ряда других групп пользователей, использующих электронные библиотеки для вычисления значений индикаторов и проведения наукометрических исследований. В докладе рассматриваются термины и семиотические основания постановки проблемы концептуализации наукометрических данных электронных библиотек и индикаторов сферы науки, а также предлагаемого подхода к ее решению.

## 1 Введение

Учитывая многозначность слова *концептуализация*, определим в явном виде значения двух словосочетаний: *концептуализация данных* и *концептуализация индикаторов*, которые являются ключевыми терминами для описания постановки *проблемы концептуализации данных электронных библиотек и индикаторов, вычисляемых на основе этих данных* (далее по тексту - проблема концептуализации).

Под *концептуализацией данных* понимается процесс формирования концептов, являющихся смысловым содержанием наукометрических данных и результатов их обработки, которые (концепты, данные и результаты обработки) используются экспертами в процессе анализа и оценивания результа-

тов научно-технической деятельности.

Под *концептуализацией индикаторов* понимается процесс формирования концептов, являющихся смысловым содержанием **новых видов** индикаторов и других количественных оценок.

Предполагается, что в процессе проведения анализа научно-технической деятельности возникает потребность в новых видах индикаторов, которые создаются и означиваются в рамках итеративного процесса их построения и вычисления значений на основе данных электронных библиотек.

Процесс концептуализации каждого нового индикатора сопровождается, как правило, разработкой алгоритма вычисления его значений и выбором названия, которое является вербальным обозначением формируемого концепта, а также может сопровождаться выбором табличных, образных или вербально-образных форм представления значений этого индикатора (например, в виде графика или диаграммы).

Таким образом, в докладе слово *концептуализация* обозначает не результат концептуального моделирования, а итерационный процесс формирования и изменения во времени смысловых значений индикаторов сферы науки. Описание процессов формирования и изменения смыслового содержания индикаторов влечет за собой необходимость рассмотрения ряда вопросов, связанных с семиотикой и лингвистикой.

Проблема концептуализации данных электронных библиотек, в том числе, частные задачи концептуализации данных наукометрических исследований и результатов их обработки являются относительно новыми. В частности, если искать на сайте журнала D-Lib Magazine ([www.dlib.org](http://www.dlib.org)) работы, в которых хотя бы один раз упоминаются любые вопросы концептуализации, то в результате выполнения такого запроса можно найти только около десяти публикаций или аннотированных библиографических ссылок на публикации в других журналах.

В найденных работах говорится о разных объектах (предметах) и средствах концептуализации, в том числе, о концептуализации пространства электронной библиотеки [1], концептуализации функциональности электронной библиотеки (ЭБ), включающей описание предметной области и абстрактной модели данных и процессов ЭБ [2], онтологии как средства концептуализации компьютерных дан-

---

Труды 10-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL'2008, Дубна, Россия, 2008.

ных [3], концептуализации цифровых ресурсов [4] и концептуализации как процесса формирования нового взгляда на данные ЭБ [5].

Проблема формирования смыслового содержания индикаторов имеет отношение ко всем перечисленным вопросам концептуализации. Если же дополнительно необходимо учитывать еще и эволюцию семантики индикаторов, то появляются новые задачи концептуализации, касающиеся экспликации стадий эволюции семантики индикаторов во времени.

Кроме индикаторов, которым в докладе уделено основное внимание, существуют и другие категории показателей в сфере науки. В процессе описания проблемы концептуализации в общем случае различаются следующие пять категорий: *индикаторы*, для которых далее будет дано определение; *характеристики* ресурсов, используемых в процессе научно-технической деятельности (НТД), кадрового потенциала или инфраструктуры сферы науки; *нормативно заданные* финансовые и другие *параметры* в сфере науки (например, бюджет РАН); *критерии* принятия решений в процессе организации НТД; количественные и качественные *экспертные оценки* [6].

Основной целью доклада является описание семиотических оснований постановки проблемы концептуализации и предлагаемого подхода к ее решению. Возникает естественный вопрос: «Какое отношение имеют проблема в целом и отдельные задачи концептуализации индикаторов сферы науки к ЭБ?»

Речь идет не только о том, что в ЭБ могут храниться данные наукометрических исследований. Отличительной чертой предлагаемого подхода к решению проблемы в целом и отдельных ее задач является то, что этот подход основан на использо-

вании средств лингвистического обеспечения ЭБ, в том числе, классификационных систем, семантических словарей и тезаурусов.

В докладе предполагается, что значения индикаторов и других количественных оценок могут быть вычислены на основе наукометрических данных, полученных в результате обработки собранных информационных ресурсов ЭБ, многоаспектно характеризующих сферу науки.

Для иллюстрации различий между традиционными и новыми индикаторами, рассмотрим ретроспективные сведения о числе научных статей ученых России и ряда других стран, а также об общем числе статей, опубликованных в ведущих мировых журналах в 1994-2005 годах (см. табл. 1, сформированную на основе работ [7] и [8]).

Данные этой таблицы получены в результате обработки информационных ресурсов ЭБ Института научной информации США, которые представляют собой библиографические описания научных статей (включая пристатейные списки цитируемых работ), опубликованных в ведущих мировых журналах.

Числовые значения в ячейках 8 строки таблицы представляют собой удельные веса (в процентах) научных статей ученых России в общем числе статей, опубликованных в ведущих мировых журналах в период 1994-2005 годы. Иначе говоря, 8 строка таблицы содержит значения **традиционного** индикатора сферы науки «Удельный вес научных статей ученых России».

Если же поставить цель оценить распределение научных статей в зависимости от возраста их авторов, то такой индикатор будет **новым**. В настоящее время значения этого индикатора вычисляются только для отдельных научных коллективов, организаций или научных фондов, финансирующих научную деятельность.

Таблица 1

Country	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Worldwide	567 204	564 645	581 751	588 488	602 381	610 203	630 452	629 353	638378	661708	688615	709541
USA	199 769	193 337	193 153	189 752	190 395	188 004	192 743	190 592	190495	196427	202075	205320
Germany	36 901	37 645	39 213	41 415	42 953	42 963	43 509	42 674	42436	42228	43009	44145
China	<b>7 821</b>	9 061	10 526	12 172	13 781	15 715	<b>18 479</b>	<b>21 134</b>	<b>23269</b>	<b>28768</b>	<b>34846</b>	<b>41596</b>
India	<b>9 928</b>	9 370	9 753	9 618	9 945	10 190	10 276	10 801	11665	12461	13367	<b>14608</b>
Russia	<b>21 612</b>	18 603	18 553	18 133	17 166	17 145	<b>17 180</b>	<b>15 658</b>	<b>15847</b>	<b>15146</b>	<b>14921</b>	<b>14412</b>
Russia/ Worldwide	3,81%	3,29%	3,19%	3,08%	2,85%	2,81%	2,73%	2,49%	2,48%	2,29%	2,17%	2,03%

Прежде чем описать суть проблемы концептуализации, сформулируем три дефиниции.

*Индикатором сферы науки* будем называть один из трех следующих видов показателей в виде количественных оценок, вычисляемых на основе данных и информационных ресурсов ЭБ, собранных в процессе информационного мониторинга:

- указатель эффективности НТД, включая, как частный случай, эффективность научных и научно-технических программ;
- указатель степени достижения целевых (запланированных) результатов НТД;

- указатель инновационного потенциала направлений научных исследований и влияния результатов НТД на развитие сферы науки и других сфер или взаимосвязей с ними (здравоохранение, образование, инновационная система страны, экономика и т.д.).

Индикаторы применяются экспертами и другими пользователями ЭБ систем мониторинга в процессе анализа и оценки деятельности юридических и физических субъектов сферы науки, а также для оценки научно-технических программ и инновационного потенциала направлений научных исследований [9]. Предлагаемое определение термина «индикатор»

намеренно ограничено только тремя перечисленными видами показателей, так как для других их видов в работе [6] было предложено использовать перечисленные ранее названия других категорий показателей: характеристики, параметры, критерии и экспертные оценки.

Важно отметить, что основной целью исследования являются не все возможные виды индикаторов, а только те из них, которые являются новыми и находятся в процессе разработки и «кристаллизации» их семантики. Процесс концептуализации касается, в первую очередь, именно новых индикаторов, смысловое содержание которых эволюционирует во времени намного быстрее по сравнению с традиционными индикаторами. При этом понимание смыслового содержания новых индикаторов иногда может быть несогласованным даже между экспертами. Иначе говоря, новый индикатор, значения которого вычисляются по некоторому итерационно изменяемому алгоритму, в процессе «кристаллизации» его семантики разными экспертами часто интерпретируется по-разному.

Для более четкого описания проблемы в целом и отдельных задач концептуализации определим еще два термина.

*Личностную концептуализацию индикатора* определим как процесс формирования концепта одним из пользователей ЭБ данных наукометрических исследований, являющегося его пониманием индикатора в рамках его системы знаний о предметной области мониторинга и наукометрии. Предполагается, что результаты процесса формирования концепта находят свое отражение в лингвистическом обеспечении ЭБ в соответствии с некоторым регламентом.

Процесс личностной концептуализации конкретного индикатора может сопровождаться выбором для него названия, которое будет авторским вербальным обозначением соответствующего сформированного личностного концепта, а также выбором табличных, образных или вербально-образных форм представления значений этого индикатора (например, в виде графика или диаграммы), что также отражается в лингвистическом обеспечении ЭБ.

*Коллективную концептуализацию индикатора* определим как процесс формирования концепта в процессе совместной деятельности группы пользователей ЭБ (минимум, два пользователя), являющегося согласованным между ними пониманием смысла индикатора. Иначе говоря, смысловое содержание индикатора должно быть согласовано в пределах группы пользователей ЭБ в процессе их совместной деятельности.

Предполагается, что результаты этого процесса согласования находят свое отражение в лингвистическом обеспечении ЭБ в соответствии с некоторым регламентом.

Процесс согласованной концептуализации конкретного индикатора, как правило, сопровождается выбором для него названия, которое является вер-

бальным обозначением соответствующего сформированного концепта, а также выбором табличных, образных или вербально-образных форм представления значений этого индикатора.

Приведенные дефиниции позволяют описать суть проблемы концептуализации, которая заключается в том, что эксперты и другие группы пользователей, работающие с одной и той же ЭБ, иногда по-разному (несогласованно) интерпретируют одни и те же данные наукометрических исследований, а также результаты их обработки с использованием одинаковых процедур. В этом случае, одной из актуальных задач проектирования средств лингвистического обеспечения ЭБ данных наукометрических исследований, в том числе, классификационных систем, семантических словарей и тезаурусов, является обеспечение согласованного понимания пользователями идентичных объектов ЭБ, которые используются при обработке наукометрических данных.

При таком понимании сути проблемы концептуализации она может рассматриваться как частный случай более общей проблемы обеспечения когнитивной интероперабельности пользователей ЭБ в процессе их совместной работы [10]. В соответствии с определением Р. Будденберга реализация когнитивной интероперабельности в процессе совместной работы предполагает обеспечение согласованного понимания пользователями идентичных информационных объектов (идентичной информации) [11].

Важно отметить, что различия в интерпретации информационных объектов, наукометрических данных и индикаторов могут служить причиной конфликтов среди экспертов, целью совместной деятельности которых является получение согласованных оценок научных и научно-технических программ. В работе [12] был приведен краткий обзор методики экспертной оценки научно-исследовательских программ США и тех нерешенных проблем, с которыми столкнулись эксперты в процессе применения этой методики.

Аналізу и обсуждению этих проблем был посвящен семинар по оценке федеральных научно-исследовательских программ США, состоявшийся 4-5 декабря 2003 года. В частности, в отчете по итогам этого семинара отмечалось, что между экспертами Административно-бюджетного управления при Президенте США и экспертами федеральных агентств США, финансирующих программы, достаточно часто возникают конфликты из-за различного понимания индикаторов эффективности реализации научно-исследовательских программ. Основной причиной упомянутых конфликтов нередко являются различия в понимании смысла индикаторов разными экспертами. В качестве превентивной меры (в целях предотвращения подобных конфликтов), в материалах семинара предлагается проводить семинары для экспертов [13].

Из краткого описания сути проблемы концептуализации уже видна та ключевая роль, которую

ЭБ играют в решении этой проблемы, так как результаты процессов концептуализации, в том числе, фиксация семантики индикаторов и ее эволюции во времени, предлагается отражать в лингвистическом обеспечении ЭБ данных наукометрических исследований, их классификационных системах, семантических словарях и тезаурусах.

## 2 Термины и семиотические основания постановки проблемы

В предыдущем разделе использовались термины «знания», «концепты», «информация (информационные ресурсы)» и «данные». Для более четкого описания проблемы концептуализации необходимо явно зафиксировать тот смысл каждого из этих терминов, который и будет использоваться при ее описании.

Начнем со слова *знания*, которое в докладе будем трактовать как результаты познавательной и креативной деятельности человека, носителем которых может быть только человек, и в которых могут быть выделены отдельные «кванты» знаний<sup>1)</sup>.

*Информацию* определим как формы эксплицитного и отчужденного от человека представления его знаний, предназначенные для передачи, непосредственного сенсорного восприятия и понимания их другими людьми.

В процессе описания проблемы концептуализации основное внимание будет уделяться тем «квантам» знаний, называемых концептами, которые являются элементарными единицами или сочетаниями элементарных единиц плана содержания, выражаемого в рамках некоторого естественного языка (в общем случае, в рамках той или иной знаковой системы [14]).

Приведенные определения терминов «знания» и «концепты» предполагают, что они являются результатом процесса членения знаний человека на «кванты», которые могут быть выражены (обозначены) в рамках некоторой знаковой системы. Процесс членения неразрывно связан с процессом выражения знаний человека в сенсорно воспринимаемой и отчужденной от человека форме, например, в виде текста на естественном языке, диаграммы или в виде геоизображения на языке карты.

Отметим, что в системе знаний человека могут быть выделены несколько планов содержания в зависимости от того языка или той знаковой системы, которыми он пользуется для представления своих знаний в отчужденной форме. Элементарные единицы плана содержания, имеющие значение в рамках некоторой знаковой системы, будем называть элементарными концептами. В докладе предполагается, что все эксперты и другие пользователи ЭБ данных наукометрических исследований используют один и тот же естественный язык, то есть ис-

пользуют одинаковые планы содержания и выражения в процессе формирования смысла индикаторов.

Здесь необходимо отметить разницу между элементарным концептом и значением слова, когда речь идет о естественном языке. В процессе описания проблемы концептуализации рассматривается только сигнификативный аспект значения. Экспрессивно-эмоциональные оценки и коннотации не рассматриваются, так как речь идет о представлении знаний в виде научно-технических текстов, графиков и диаграмм, а не в виде художественных произведений. Следовательно, термин «элементарный концепт» используется для обозначения только сигнификативной составляющей значения слова [15].

Термин *знаковая информация* определим как результаты процесса представления концептов человеком-генератором этих результатов в плане выражения сферы социальных коммуникаций пользователей ЭБ в любой отчужденной форме, которая является сенсорно воспринимаемой другими участниками коммуникаций. Отметим, что при таком определении термин «знаковая информация» имеет отношение только к формам представления концептов, а введенный ранее термин «информация» - к формам представления любых «квантов» знаний, включая концепты.

Выделение в системе знаний человека нескольких планов содержания позволяет учитывать различия в членении знаний экспертов в разных естественных языках и других знаковых системах [16]. В дальнейшем учет одновременного использования разных планов содержания может понадобиться при решении проблемы концептуализации в процессе совместной работы пользователей ЭБ, говорящих на разных языках.

Термин «коды» определим как компьютерные эквиваленты двоичных цифр (или их последовательностей), которые могут представлять собой намагниченность или ее отсутствие, наличие электрического тока или его отсутствие, способность к отражению света или ее отсутствие в цифровой среде [17]. Двоичные цифры «0» и «1», о которых говорится в определении термина «коды» принадлежат сфере социальных коммуникаций ЭБ, а их компьютерные эквиваленты - цифровой среде.

При описании проблемы концептуализации будем выделять среди всех возможных кодов цифровой среды ЭБ три следующие категории:

- коды, соотнесенные с концептами знаний пользователей ЭБ (*первая категория*),
- коды, соотнесенные с эксплицитными и отчужденными от пользователей ЭБ формами представления концептов в плане выражения сферы социальных коммуникаций ЭБ (*вторая категория*),
- коды, соотнесенные с информационными объектами, алгоритмами, программами и другими категориями денотатов ЭБ (*третья категория*).

<sup>1)</sup> «Кванты» знаний являются переводом английского словосочетания «knowledge parts».

Например, для кодирования смысла индикаторов будут использоваться коды первой категории, для кодирования названий индикаторов - коды второй категории, а для кодирования программ вычисления значений индикаторов - коды третьей категории.

Также определим еще одну (нулевую) категорию кодов, которую будем называть «цифровыми данными» и к которой будем относить все коды цифровой среды, не относящиеся к трем выше определенным категориям. Таким образом, к цифровым данным будем относить результаты любых измерений, полученных с помощью цифровых компьютерных и других технических систем (устройств), и результаты вычислений, не являющихся формой представления знаний человека.

Формы представления цифровых данных в сфере социальных коммуникаций ЭБ в сенсорно воспринимаемом виде будем называть данными. Отметим, что данные как отображение в сфере социальных коммуникаций результатов цифровых измерений и вычислений могут сопровождаться метаданными, являющиеся одним из видов информации.

Перечислим рассмотренные термины, разделив их на три части в зависимости от их природы (ментальная, социальная и цифровая):

- знания, концепты и элементарные концепты (*ментальная сфера пользователей ЭБ*),
- информация, данные и знаковая информация (*сфера социальных коммуникаций пользователей ЭБ*),
- цифровые данные и три категории кодов (*цифровая среда ЭБ*).

Отметим, что определение термина «концепт» существенно зависит от дефиниции семиотического термина «знаковая система», так как любой концепт определяется в рамках той или иной знаковой системы. Однако по определению знака две его стороны (форма как материально выраженная его составляющая и значение как его идеальная составляющая), будучи поставлены в отношении постоянной связи, опосредованной сознанием, составляют устойчивое единство, которое посредством сенсорно воспринимаемой формы знака репрезентирует конвенционально приданное ему значение [18]. При этом, знаковая система определяется как совокупность знаков, объединенных между собой различными видами связей [14].

Таким образом, из определения знака следует, что концепты являются и конвенциональными, и стабильными, то есть, чтобы определить личностные и коллективные концепты пользователей ЭБ, необходимо предварительно определить понятия авторского и коллективного знаков.

*Авторский знак* отличается от традиционного семиотического знака тем, что две стороны авторского знака – форма и значение знака – могут находиться в отношении временной связи, опосредованной сознанием одного пользователя ЭБ, составлять нестабильное единство, которое посредством сенсорно воспринимаемой формы знака этим пользователем может репрезентировать персонально при-

данное ему значение в течение некоторого периода времени.

*Коллективный знак* отличается от авторского тем, что две стороны коллективного знака могут находиться в отношении временной связи, опосредованной сознанием каждого из пользователей-участников совместной деятельности, то есть используют и согласованно понимают коллективные знаки, как минимум, два пользователя ЭБ.

Авторские и коллективные знаки не возникают и тем более не функционируют отдельно. В своей совокупности с традиционными знаками они образуют единую знаковую систему, на основе которой проектируется лингвистическое обеспечение ЭБ данных наукометрических исследований. Иначе говоря, авторские и коллективные знаки означаются в совокупности с традиционными знаками, являющимися составными элементами естественного языка или другой знаковой системы<sup>2)</sup>.

Используя определения авторского и коллективного знаков, определим элементарный личностный концепт как значение авторского знака, а личностный концепт как значение выражения на естественном языке, содержащего хотя бы один авторский знак, либо как новое значение выражения на естественном языке (например, дефиниция нового индикатора) без авторских знаков, новый смысл которого определен автором в явном виде.

Определим элементарный коллективный концепт как значение коллективного знака, а коллективный концепт как значение выражения на естественном языке, содержащего хотя бы один коллективный знак, либо как новое значение выражения на естественном языке без коллективных знаков, новый смысл которого определен в явном виде и согласованно понимается как минимум, двумя пользователями ЭБ.

Сопоставляя приведенные определения с дефинициями из предыдущего раздела, можно сказать, что в результате личностной концептуализации индикаторов получают личностные концепты и авторские знаки, а в результате согласованной концептуализации индикаторов – коллективные концепты и знаки.

В заключение этого раздела отметим важное положение: из приведенных определений терминов следует, что соотносительность знаковой информации ЭБ, например, библиографических описаний научных статей, и обозначаемых этой информацией концептов, позволяющих определить источник публикации этих статей, по определению является конвенциональной для всех пользователей ЭБ, понимающих язык (языки) представления этой информации.

Однако данные и значения новых индикаторов, полученные в результате обработки этой информации, не всегда и не всеми пользователями ЭБ могут интерпретироваться согласованно. Поэтому, в на-

<sup>2)</sup> Приведенные определения авторского и коллективного знаков являются развитием положений из работы [18].

звании доклада говорится именно о *концептуализации данных*, которые, в отличие от информации, в рамках рассмотренной системы терминов по определению не являются результатом процесса конвенционального представления знаний пользователей ЭБ.

### 3 Проблема концептуализации и виды обеспечения электронной библиотеки

В процессе разработки, вычисления значений и применения индикаторов сферы науки принимают участие разные группы пользователей ЭБ данных наукометрических исследований [19]:

- лица, принимающие решения,
- организационные и программные менеджеры,
- эксперты по областям знаний,
- специалисты по мониторингу,
- лингвисты и IT-специалисты, участвующие в разработке ЭБ.

Как отмечалось ранее, проблема концептуализации заключается в том, что пользователи, работающие с одной и той же ЭБ, иногда по-разному интерпретируют значения индикаторов, полученные в результате обработки одних и тех же ее цифровых данных и кодов ЭБ с использованием одинаковых процедур. Естественно, что различия в понимании индикаторов являются серьезным препятствием в проведении анализа и получении согласованных экспертных оценок результатов НТД.

Поэтому, уже в процессе создания ЭБ данных наукометрических исследований необходимо предусматривать разработку средств обеспечения согласованного понимания индикаторов в пределах как отдельных групп, так и всего сообщества пользователей этой ЭБ. Разработка средств формирования согласованного понимания является одной из ключевых задач проблемы концептуализации, от решения которой во многом зависят результаты наукометрических исследований. При этом, речь не идет обязательно о двухэтапном процессе: сначала личностная концептуализация, а затем коллективная. Целью разработки этих средств является формирование с самого начала совместной деятельности именно согласованного понимания данных наукометрических исследований и новых индикаторов.

В интересах разработки средств обеспечения согласованного понимания данных и индикаторов необходимо определить причины возникновения проблем, связанных с непониманием и/или различным пониманием разными пользователями ЭБ смысла новых индикаторов. Перечислим основные причины, используя работу [19].

Во-первых, названия значительного числа индикаторов, используемых в наукометрических исследованиях, не обладают той степенью конвенциональности, которая позволила бы определить значения этих индикаторов в процессе диалога пользователей ЭБ. Даже наличие дефиниции не всегда помо-

гает согласованному пониманию смысла индикаторов.

Во-вторых, существуют индикаторы, зависящие от нескольких параметров, изменение которых существенно изменяет их численные значения (например, значения индикаторов цитирования зависят от глубины ретроспективы используемого массива научных статей) и/или от выбора варианта используемого алгоритма вычисления, что может иногда изменить и смысл индикатора (например, учет самоцитирования и ссылок соавторов меняет значения и смысл индексов цитирования авторов).

В-третьих, численные значения индикаторов могут зависеть от числа, структуры и наполнения записей в используемых нормативных файлах (например, значения индексов цитирования зависят от конкретного списка журналов, используемых в процессе мониторинга и оценки), а также от текущего состояния семантического словаря ЭБ, включающего дефиниции индикаторов, ссылки на алгоритмы их вычисления, схемы информационных ресурсов ЭБ и используемых нормативных списков.

Приведенный перечень причин, не претендующий на полноту, иллюстрирует необходимость информирования пользователей ЭБ, проводящих наукометрические исследования, о зависимости смысла и прагматических аспектов использования индикаторов от разных видов обеспечения ЭБ. Например, пользователям ЭБ может понадобиться следующая информация о видах обеспечения ЭБ:

- реализованные алгоритмы вычисления индикаторов (алгоритмическое обеспечение),
- глубина ретроспективы используемого массива научных статей (информационное обеспечение),
- изменение наполнения семантического словаря (лингвистическое обеспечение),
- действующие нормативные акты и документы в сфере мониторинга и оценивания НТД (нормативное обеспечение).

Приведем пример, когда появление новых нормативных актов привело к существенному изменению прагматических аспектов использования удельного веса научных статей ученых России в общем числе статей, опубликованных в ведущих мировых журналах (см. табл. 1). В Стратегии развития науки и инноваций в РФ на период до 2015 года в качестве индикатора оценки степени достижения целевых результатов реализации этой Стратегии предложено использовать удельный вес научных статей ученых России [20]. В Стратегии указаны следующие целевые (запланированные) значения этого индикатора: 2004 - 2,7%, 2005 - 2,9%, 2006 - 3,1%, 2007 - 3,2%; 2008 - 3,4%, 2009 - 3,5%, 2010 - 4,0%.

Как видно из табл. 1 известные в настоящее время фактические значения этого индикатора меньше целевых: 2004 - 2,17% и 2005 - 2,03% [8]. Важно отметить, что до утверждения Стратегии удельный вес научных статей позиционировался только как одна из характеристик сферы науки, а после ее ут-

верждения он дополнительно приобрел роль индикатора степени достижения целевых результатов Стратегии.

С онтологической точки зрения удельный вес научных статей является вершиной двух разных отношений. В первом случае эта вершина входит в отношение «тип-тип» (удельный вес научных статей - это характеристика сферы науки), а во втором случае она входит в отношение «тип-роль» (удельный вес научных статей - это индикатор оценки степени достижения целевых результатов). Различие заключается в том, что удельный вес научных статей является характеристикой сферы науки до утверждения Стратегии, в течение ее срока действия и после его окончания, а его использование в качестве индикатора степени достижения целевых результатов определено только в течение срока действия Стратегии [21].

Вернемся к проблеме концептуализации и попытаемся сформулировать один из возможных методологических подходов к ее решению, а также определить, как и в какой форме описать и визуализировать влияние вариантов реализации разных видов обеспечения ЭБ на формирование согласованного понимания индикаторов и прагматические аспекты их использования. При этом желательно, чтобы выбранная форма описания была понятна пользователям ЭБ. Кроме того, одни и те же информационные ресурсы сферы науки, например библиографические описания научных статей, могут использоваться в ЭБ разных систем мониторинга. В таком случае, каким же образом можно сопоставлять значения и смысловое содержание индикаторов, вычисляемых с помощью информационных ресурсов ЭБ разных систем мониторинга, чтобы можно было взаимно верифицировать результаты наукометрических исследований, полученные разными коллективами?

Предлагаемый подход к решению проблемы концептуализации существенно использует идеи Г.П. Щедровицкого и Н.Г. Алексева, «связанные с противопоставлением двух планов, или аспектов изучения мышления – плана образов (или знаний) и плана "процессов" (или деятельности) ...». В работе [22] сформулирована задача «операционно-деятельностного анализа понятий и знаний, позволяющего, исходя из формы какого-либо сложившегося понятия или знания, сводить его к системе операций и действий, порождающих содержание этого понятия или знания (выделено мною. – И.З.)» [23].

Предлагаемый в докладе подход имеет два основных отличия. Во-первых, в нем учитывается то, что для индикаторов нередко отсутствуют конвенциональные формы представления их смысла, то есть, общепринятые названия индикаторов. Во-вторых, в нем учитывается то, что смысловые значения новых индикаторов не являются устоявшимися и общепринятыми. И, тем не менее, идея Г.П. Щедровицкого и Н.Г. Алексева о системе операций и действий, порождающих содержание понятий, вероятно, может быть использована в качестве ме-

тодологической основы решения проблемы концептуализации.

Общепринятое название и четкий смысл у нового индикатора могут отсутствовать, но если разработано соответствующее лингвистическое обеспечение ЭБ данных наукометрических исследований, то ее процессы, содержание и реализация видов обеспечения со временем помогут сформировать согласованное понимание этого индикатора в процессе совместной деятельности экспертов и других групп пользователей ЭБ.

Важно отметить, что, во-первых, процессы формирования согласованного понимания отчасти являются управляемыми; во-вторых, обеспечивая формирование согласованного понимания индикаторов, ЭБ выполняют важную и социально значимую функцию фиксации новых знаний в предметной области наукометрии.

Одной из форм описания влияния содержания и реализации видов обеспечения ЭБ на формирование согласованного понимания индикаторов могут быть полидоменные модели видов обеспечения ЭБ, фиксирующие стадии и результаты процесса концептуализации индикаторов. Их описанию посвящена работа [24].

Эти модели, которые позволяют описывать влияние разных видов обеспечения ЭБ на значения, смысл индикаторов и прагматические аспекты их использования, включают следующие компоненты [25]:

- 1) информационный компонент в форме набора схем информационных ресурсов ЭБ;
- 2) математический компонент в виде множества функций, используемых для вычисления значений индикаторов;
- 3) алгоритмический компонент, необходимый для определения значений параметров функций на основе информационных ресурсов ЭБ;
- 4) информационно-математический компонент, определяющий связи между параметрами функций и теми информационными ресурсами, которые необходимы для определения значений параметров;
- 5) лексико-семантический компонент, включающий используемые в ЭБ классификаторы, семантические словари и тезаурусы;
- 6) аналитический компонент в форме набора критериев принятия решений, построенных на основе индикаторов, методики построения критериев и правил их применения пользователями ЭБ;
- 7) нормативно-правовой компонент модели.

В полидоменной модели для каждого индикатора предлагается фиксировать в роли его денотата следующие составляющие: 1) схемы соответствующих ему информационных ресурсов, 2) функции для вычисления его значений, 3) алгоритмы для определения параметров функций на основе информационных ресурсов ЭБ, 4) связи между схемами, функциями, алгоритмами, параметрами функций и теми информационными ресурсами, которые необ-

ходимы для определения значений этих параметров [25].

#### 4 Пример индикатора

В начале доклада была приведена таблица, которая содержит значения традиционного индикатора сферы науки «Удельный вес научных статей ученых

Проценты (%)

России». Был также упомянут новый индикатор, основная цель разработки которого заключается в том, чтобы оценить распределение научных статей в зависимости от возраста их авторов. Это даст возможность оценивать публикационную активность каждой возрастной группы ученых.

В настоящее время разработка нового индикатора ведется в рамках создания Информационной системы обеспечения реализации и мониторинга Программы фундаментальных научных исследований РАН. Идея разработки индикатора для оценивания распределение научных статей в зависимости от возраста их авторов возникла в процессе разработки средств вычисления традиционного индикатора «Возрастное распределение сотрудников научного коллектива (подразделения, организации)» (см. рис. 1).

На этом рисунке приведены значения индикатора в виде графика процентного распределения сотрудников одного из научных коллективов ИПИ РАН по 14 возрастным группам (20 – 24, 25 – 29 и далее до группы 85 – 89 лет).

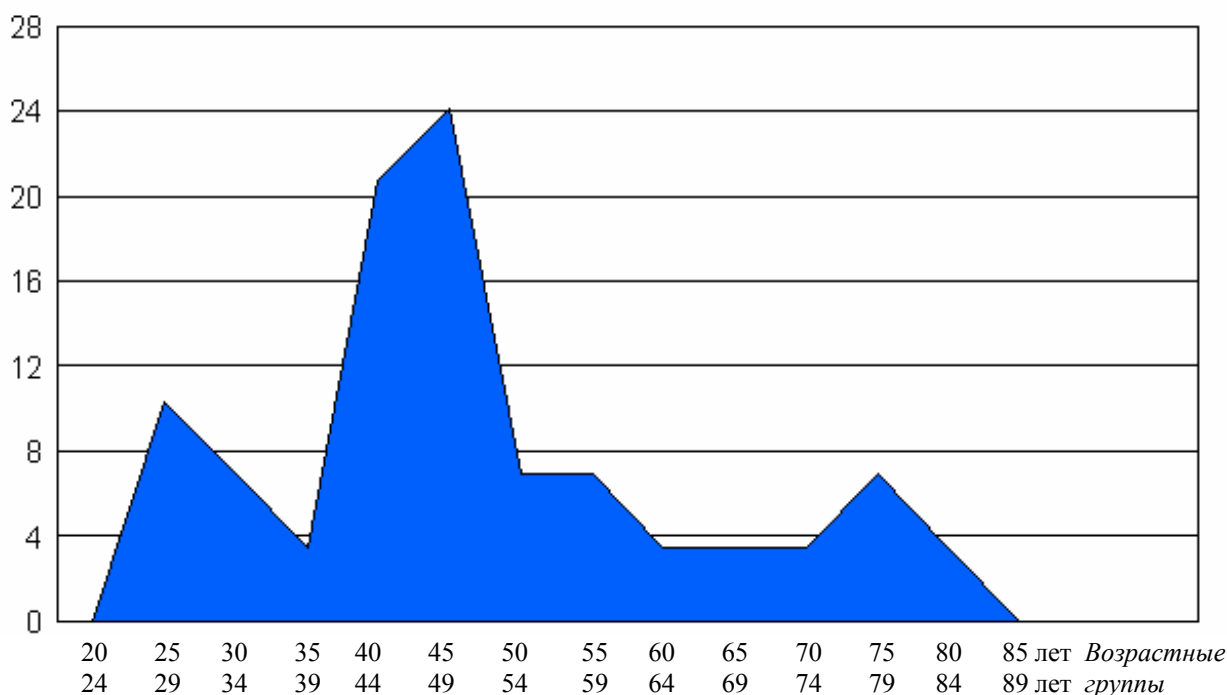


Рис. 1. График возрастного распределения сотрудников научного коллектива

Для каждой группы сотрудников вычисляется ее доля в общей численности научного коллектива. На рисунке видно, что для этого коллектива наибольшую долю имеет возрастная группа 45 – 49 лет (24%). Далее по убыванию следуют следующие пять возрастных групп: 40 – 44 года (20,7%), 25 – 29 лет (10,3%), 50 – 54 года, 55 – 59 лет и 75 – 79 лет (по 6,9%). В настоящее время имеется возможность построения графиков возрастного распределения с шагом от одного года и более<sup>3)</sup>.

Для определения публикационной активности возрастных групп, сейчас разрабатываются средства вычисления значений этого нового индикатора. Кроме этого, разрабатываются средства вычисления индикаторов патентной активности и участия в научных мероприятиях также с определением доли каждой возрастной группы.

В презентации доклада будут описаны основные стадии разработки индикатора публикационной активности возрастных групп, будет продемонстрирована эволюция семантики этого индикатора и тех факторов, которые повлияли на изменение его смыслового содержания.

#### 5 Заключение

Процесс концептуализации индикаторов охватывает следующие сущности: концепты, информация (информационные ресурсы), данные, три категории кодов и цифровые данные. В предлагаемой системе терминов эти сущности в соответствии с их природой, зафиксированной в дефинициях терминов, отнесены к трем разным сферам (средам) и двум разным источникам их происхождения – человек и цифровая техническая система. По смыслу перечисленные термины не пересекаются (имеют изолированные семантические поля в рамках рас-

<sup>3)</sup> Программы вычисления значений этого индикатора разработаны В.В. Косариком.



смотренной системы терминов), что позволяет более четко описывать процессы личностной и коллективной концептуализации через взаимодействие соответствующих сущностей. Изолированность семантических полей исключает родовидовые и синонимические отношения между этими терминами, что необходимо для более четкого описания проблемы концептуализации.

Предлагаемый подход к решению этой проблемы основан на идеи Г.П. Щедровицкого Н.Г. Алексеева о системе операций, порождающих содержание понятий. В качестве основного средства решения проблемы предложено использовать полидоменные модели видов обеспечения ЭБ, фиксирующие стадии и результаты процесса концептуализации индикаторов [25].

Основное назначение полидоменной модели как интегрального описания процесса концептуализации индикаторов заключается в том, что она делает «прозрачными» для экспертов и других пользователей ЭБ структурные, процедурные и содержательные элементы разных видов обеспечения ЭБ, которые вовлечены в процесс вычисления значений любого индикатора.

Предложенный подход к решению проблемы концептуализации индикаторов опробуется в настоящее время в процессе создания Информационной системы обеспечения реализации и мониторинга Программы фундаментальных научных исследований государственных академий наук на 2008 - 2012 годы.

Программа утверждена распоряжением Правительства РФ от 27 февраля 2008 г. № 233-р. В ней определены восемь индикаторов эффективности ее реализации. Для каждого индикатора заданы пять целевых значений для каждого из 2008 - 2012 годов. При этом предполагается, что число индикаторов, методика их определения и их целевые значения могут изменяться (корректироваться и уточняться). Кроме индикаторов эффективности, в утвержденной Программе заданы параметры финансирования на 2008 - 2012 годы для каждого направления научных исследований. Например, для РАН параметры финансирования заданы для 88 направлений на период 2008 - 2012 годы.

Разработка полидоменной модели для концептуализации индикаторов и других показателей эффективности реализации Программы на 2008 - 2012 годы начата в ИПИ РАН с нормативного (совместно с ЦЭМИ РАН), информационного и лексико-семантического видов обеспечения ЭБ. Последний компонент включает семантический словарь и ряд классификаторов.

В заключение отметим, что полидоменные модели могут использоваться не только для описания процессов концептуализации индикаторов. Они могут быть применены для описания всей функциональности ЭБ, используемых в наукометрических исследованиях, и системы мониторинга в целом.

## Литература

- [1] Candela L., Castelli D., Pagano P. et al. Setting the Foundations of Digital Libraries: The DELOS Manifesto // D-Lib Magazine, March/April 2007, Vol. 13, No 3/4.
- [2] Lavoie B., Henry G., Dempsey L. A Service Framework for Libraries // D-Lib Magazine, July/August 2006, Vol. 12, No 7/8.
- [3] Shvaiko P., Euzenat J. In Brief: Ontology Matching // D-Lib Magazine, December 2005, Vol. 11, No 12.
- [4] Mankita I., Meltzer E. A Handful of Things: Calisphere's Themed Collections from the California Digital Library // D-Lib Magazine, May 2006, Vol. 12, No 5.
- [5] Dillon A. Spatial-Semantics: How Users Derive Shape from Information Space // Journal of the American Society for Information Science, 2000, Vol. 51, No 6, pp. 521-528 (D-Lib Magazine опубликовал аннотацию этой статьи по адресу: [www.dlib.org/dlib/march00/03JASIS\\_TOC.html](http://www.dlib.org/dlib/march00/03JASIS_TOC.html)).
- [6] Зацман И.М. Терминологический анализ нормативно-правового обеспечения создания систем мониторинга в сфере науки // Экономическая наука современной России. – 2005, №4. – С. 114–129.
- [7] National Science Board, Science and Engineering Indicators – 2006. Two volumes. Arlington, VA: National Science Foundation, 2006.
- [8] National Science Board, Science and Engineering Indicators – 2008. Two volumes. Arlington, VA: National Science Foundation, 2008.
- [9] Зацман И.М., Курчавова О.А., Галина И.В. Информационные ресурсы и индикаторы для оценки инновационного потенциала направлений научных исследований // Системы и средства информатики. Вып. 18. – М. : Наука, 2008.
- [10] Зацман И.М., Курчавова О.А. Термины для описания процессов представления научно-технических знаний в цифровой среде // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной Международной конференции «Диалог». Вып. 7(14). – М. : РГГУ, 2008. – С. 164–170.
- [11] Buddenberg R. Toward an Interoperability Reference Model - [http://web1.nps.navy.mil/~budden/lecture.notes/interop\\_RM.html](http://web1.nps.navy.mil/~budden/lecture.notes/interop_RM.html) (состояние страницы на 12.02.2008).
- [12] Зацман И.М., Веревкин Г.Ф. Информационный мониторинг сферы науки в задачах программно-целевого управления // Системы и средства информатики. Вып. 16. – М. : Наука, 2006. – С. 164-189.
- [13] Planning for Performance and Evaluating Results of Public R&D Programs: Meeting the OMB PART Challenge (Workshop Report). – Washington: The Washington Research Evaluation Network, 2004.
- [14] Соломоник А. Синтаксис в знаковых системах. – Минск : МЕТ, 2007.

- [15] Гак В.Г. Лексическое значение слова // Языкознание : Большой энцикл. слов. – М. : Большая рос. энцикл., 1998. – С. 261–263.
- [16] Vossen P. (ed.) EuroWordNet General Document (Version 3) (<http://www.illc.uva.nl/EuroWordNet/docs/GeneralDoc>).
- [17] McArthur D. Information, its forms and functions: The elements of semiology. – Lewinton: The Edwin Mellen Press, Ltd., 1997.
- [18] Уфимцева А.А. Знак языковой // Языкознание : Большой энцикл. словарь. – М. : Большая рос. Энцикл., 1998. – С. 167.
- [19] Зацман И.М., Кожунова О.С. Семантический словарь системы информационного мониторинга в сфере науки: задачи и функции // Системы и средства информатики. Вып. 17. – М. : Наука, 2007. – С. 124–141.
- [20] Стратегия развития науки и инноваций в РФ на период до 2015 года. Утверждена Межведомственной комиссией по научно-инновационной политике, протокол от 15 февраля 2006 г. № 1 (<http://mon.gov.ru/science-politic/conception/strateg.zip>).
- [21] Лукашевич Н.В. Описание понятий-ролей в лингвистических и онтологических ресурсах // Тезисы постерных докладов 9-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2007. – Переславль-Залесский : Изд-во «Ун-т города Переславля», 2007. – С. 81–89.
- [22] Щедровицкий Г.П., Алексеев Н.Г. О возможных путях исследования мышления как деятельности // Докл. АПН РСФСР. 1957. № 3; 1958. № 1, 4; 1959. № 1, 2, 4; 1960. № 2, 4, 5, 6; 1961. № 4, 5; 1962. № 2-6.
- [23] Пископел А.А. К творческой биографии Г.П. Щедровицкого (1929–1994) // Щедровицкий Г.П. Избр. труды. – М. : Шк. культ. полит., 1995. – С. XIII–XXXVII.
- [24] Зацман И.М. Полидоменные модели в системах оценки инновационного потенциала и результативности научных исследований // Труды международной конференции Диалог-2006 «Компьютерная лингвистика и интеллектуальные технологии». – М. : Изд-во РГГУ, 2006. – С. 178–183.
- [25] Зацман И.М. Полидоменные модели электронных библиотек систем мониторинга сферы науки // Труды 8-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2006. – Ярославль : Ярославский государственный ун-т, 2006. – С. 75–81.

## Conceptualization of Scientometric Data in Scientific Digital Libraries

I.M. Zatsman

The problem of science indicator conceptualization is considered. Indicator conceptualization processes are analyzed in terms of teamwork of digital library users carrying out scientometric research. A number of new terms, the semiotic foundations for a *science indicator conceptualization problem*, and semiotic approach to its solution are discussed. The science indicator conceptualization problem is formulated as a special case of the cognitive interoperability of digital library users.

Definitions are formulated for new terms “personal indicator conceptualization” and “coordinated indicator conceptualization”. As it follows from these definitions, digital libraries play a key role in indicator conceptualization processes. The key role implies that fixing of indicator significance and an evolution of indicator semantics for the time is reflected in a digital library lingware, classification systems, semantic dictionaries and thesauruses.