

ДОКУМЕНТАЛЬНЫЕ ИСТОЧНИКИ ИНФОРМАЦИИ

УДК 519.237:((051):66)

Б. Л. Мильман, Ю. А. Гаврилова

НАУЧНЫЕ НОВОСТИ В БИЗНЕС-ОРИЕНТИРОВАННЫХ ПЕРИОДИЧЕСКИХ ИЗДАНИЯХ КАК ИСТОЧНИК ИНФОРМАЦИИ О ПРИКЛАДНЫХ И СТРАТЕГИЧЕСКИХ ИССЛЕДОВАНИЯХ И НАУЧНОЙ ПОЛИТИКЕ

Исследованы научные новости в периодических изданиях, включенные в виде экстрактов в бизнес-ориентированную базу данных Chemical Industry Notes. Проведен содержательный анализ экстрактов; изучены соответствующие индексированные термины, в том числе с использованием метода совместной встречаемости слов. Показано, что научные новости хорошо воспроизводят основные предметные и национальные пропорции в прикладных, промышленных, стратегических исследованиях и разработках. Документы этого типа позволяют построить прямые карты научной политики, организации и планирования исследований и отразить структуру этих видов деятельности. Охарактеризованы другие особенности научных новостей как индикатора прикладных и стратегических исследований и научной политики.

Потенциальные возможности применения наукометрических методов в научной и технологической политике и управлении исследованиями и разработками (ИР) определяются, во-первых, развитием самих этих методов и, во-вторых, разнообразием баз данных (источников информации, типов документов), используемых для количественного анализа. Среди такого рода методов, позволяющих описывать и моделировать научную и технологическую деятельность, прежде всего выделяются методы цитирования и совместной встречаемости слов (СВС). Выбор того или иного аналитического инструмента далеко не в последнюю очередь определяется и видом используемых документов (баз данных).

Первым в практику науковедения и научной политики введен анализ цитирования, достигший высокого уровня развития к середине 80-х гг. [1—3]. Основное достоинство этого метода заключается в том, что он позволяет отчетливо выделить фронты фундаментальных исследований в естественных науках в виде кластеров высокоцитируемых и высокосоцитируемых работ и текущих публикаций, цитирующих эти работы. Специальные методики анализа дают возможность определить фундаментальные исследования переднего края и в прикладных, технологических областях [4]. Сами кластеры цитирования могут быть исходной точкой для построения лексических и смешанных наукометрических моделей [5—7]. Метод цитирования можно использовать при обработке различных типов документов, имеющих ссылки (статьи, книги, отчеты, возможно, описания изобретений к патентам), но он подходит в основном для базы данных SCI со всеми ее достоинствами и недостатками.

Метод СВС стал интенсивно развиваться во второй половине 80-х гг. усилиями исследователей парижского

Центра социологии инноваций Высшей горной школы [8—15], других групп в Париже [15—18], Киле [9, 15, 19, 20], Лейдене [21, 22] и Амстердаме [23—25]. До этого техника СВС («первого поколения») была, скорее, лишь приложением к социологии перевода [26] и оказалась малоприменимой для экспериментов по «картированию науки для планирования исследований» [27]. В большинстве отмеченных работ с использованием метода СВС «второго поколения», начиная с публикаций [8, 15, 16], были сформированы кластеры ключевых слов (или слов в заголовках статей), наиболее часто встречающихся совместно в одном и том же документе (его описании в базе данных). Для выделения кластеров максимизировано число слов в них. Такие кластеры, отвечающие дискретным предметным единицам науки, были названы темами. Силы связей между словами внутри кластеров и силы межкластерных связей фиксировались в виде показателей плотности и центральности. Стратегическое положение в сетях кластеров занимают темы с высоким показателем центральности. Временные зависимости этих и других показателей, таких, как состав кластеров, являются основой динамического анализа отдельных научных и технологических областей [8, 10—15, 17, 20, 22, 25].

Подобные лексические модели на основе СВС позволяют также провести сопоставительный анализ фундаментальных и прикладных исследований, а также разработок в одной и той же научно-технологической области [8, 13]. При этом сравниваются сети кластеров слов, построенные с использованием специально сформированных файлов различных релевантных документов. Так, технологические разработки моделируются кластерами слов, взятых из описаний изобретений к патентам [8, 12, 14, 25].

По-видимому, метод СВС лучше подходит для анализа прикладных и технологических работ, чем техника цитирования, так как эти работы цитируются сравнительно редко (см., например, [4, 28]) даже в описаниях изобретений к патентам [29]. То же самое можно утверждать и об изучении взаимодействия науки и технологии. Действительно, и статьи, и описания изобретений к патентам одинаково подходят для лексического анализа и не требуются существенных изменений в его методологии при переходе к другой группе документов. В то же время системы ссылок в научных статьях и описаниях изобретений к патентам существенно различаются. Например, доля высокоцитируемых патентов ниже, чем высокоцитируемых статей, а доля нецитируемых патентов соответственно выше [30].

Тот факт, что для лексического анализа подходят документы по существу любого вида обещает технике СВС

Таблица 1

Периодические издания, содержащие научные новости

Название	Страна издания	Число сообщений
Biotechnol. News	США	19
Genet. Eng. News	США	8
Chem. Eng. News	США	7
Chem. Mark. Rep.	США	6
Eur. Biotechnol. News	Франция	6
N. Y. Times (N. Y. Edit.)	США	5
Financ. Times (North Am. Ed)	США (Великобритания)	4
Jpn. Chem. Week	Япония	4
J. Commer.	США	3
Rubber World	США	3
Am. Paint. Coat. J.	США	2
Bio/Technology	США	2
Chem. Br.	Великобритания	2
Chem. Bus.	США	2
Chem. Weekly.	Индия	2
China Dly (North Am. Ed.)	США (Китай)	2
Congr. Rec.	США	2
Jpn. Econ. J.	Япония	2
Mod. Paint Coat	США	2
Mod. Plast.	США	2
Am. Met. Mark	США	1
Asian. Wall St. J.	Гонконг	1
Chem. Eng. (N. Y.)	США	1
Chem. Ind. (London)	Великобритания	1
Chem. Rundsch.	Швейцария	1
Chem. Week	США	1
China Bus. Trade	США	1
Coal (Chicago)	США	1
East West Technol.	США	1
Eur. Chem. News	Великобритания	1
Eur. Plast News	Великобритания	1
Farbe Lack	Германия	1
Fed. Regist.	США	1
Ned. Chem. Ind.	Нидерланды	1
Oil Gas J.	США	1
Paper (Tonbridge)	Великобритания	1
Wall St. J. (East. Ed.)	США	1
Всего: 37 названий	9(10) стран	96

широкое применение в науковедении. Так, сравнение лексических сетей для проектов исследований, представленных к финансированию, и итоговых отчетов финансируемых исследователей позволяет установить приоритеты в поддержке тех или иных направлений в рамках научной программы [8, 12]. (Исследовательские предложения в базах данных по контрактам были изучены и другими статистическими методами [31].)

Метод СВС для изучения научной политики и ее влияния на развитие науки широко применяется в работах французских исследователей [8, 12, 15, 17, 18], что свидетельствует о новой ступени развития количественных наукометрических методов [32]. Последнее относится и к возможностям прогнозирования динамики научных и технологических областей по соотношению центральности и плотности кластеров слов [14].

Нам представляется важным дальнейшее развитие метода СВС в двух указанных аспектах. Первый из них касается описания и моделирования прикладных и технологических исследований; эти возможности ограничены у конкурирующего метода — анализа социализированных. Во-вторых, нам кажется необходимым расширение круга анализируемых документов (баз данных) как предпосылки увеличения возможностей техники СВС для получения данных о различных сторонах научной и сопутствующей деятельности. Прежде всего имеется в виду анализ научной и технологической политики, управления и планирования ИР* и их связи с развитием собственно науки и технологии.

Преследуя эти цели, в настоящей работе мы впервые изучили научные новости (events) в бизнес-ориентированных периодических изданиях в области химии и

Таблица 2

Темы научных новостей

Тема	Число сообщений*
Результаты новых исследований	35
Открытие, строительство или планирование нового исследовательского центра (лаборатории)	13
Финансирование планируемых исследований (проектов)	12
Программы и проекты исследований	8
Соглашение о совместных исследованиях	6
Обзор исследований в определенной области в той или иной стране (регионе)	5
Финансирование состоявшихся исследований	4
Состоявшаяся или планируемая научная конференция	4
Коммерциализация исследований	3
Применение химических продуктов	3
Приоритеты и перспективы исследований	3
Разрешение на проведение исследований (испытаний)	2
Изменения в законодательстве, регулирующем научно-техническую политику	2
Три других	3x1
Всего	103

* Сумма этих показателей превышает число новостей в выборке (96), поскольку нескольким сообщениям приписаны две темы.

* Далее все эти сферы подразумеваются, даже если употребляется только термин *научная политика*.

Наиболее часто встречающиеся индексирующие термины

Таблица 3

Ранг	Слово	Частота	Классификационная группа
1	US	339	III
2	drug	173	I
3	treatment	169	
4	center	144	IV
5	use	135	
6	biotechnology	125	I
7	chemical	120	
8	company	115	IV
9	Japan	103	III
10	Laboratory(-lum)	88	IV
11	virus	80	I
12	technology(-ies)	78	
13	development	75	
14	plant	73	I
15-16	gene	69	I
	cancer	69	I
17-18	protein	63	I
	UK	63	III
19	science	62	
20	AIDS*	60	I
21-22	genetic (-ally)	54	II
	superconductor	54	II
	(-ing.)		
23	engineering**	51	I
24	disease	45	
25-26	cell	43	I
25-26	W. Germany	43	III
27	nuclear	41	II
28-29	material	40	II
	program	40	IV
30	fusion***	39	II
31-33	bill	36	IV
	expenditure	36	IV
	polymer	36	II
34	oil	35	II
35	Introduced****	34	IV
36-37	environment	32	II
	fund (-Ing)	32	IV
38	vaccine	31	I
39	plastic	29	II
40-41	DNA	28	I
	Industry (-ial)	28	
42-43	France	27	III
	waste	27	II
44	growth (-Ing)	26	I
45-48	antibody	25	I
	composite	25	II
	prevention	25	
	update	25	
49	new	24	
50-53	metal	23	II
	overview	23	
	profile	23	
	project	23	IV
54-55	enzyme	22	I
	product	22	
56-57	coating	21	II
	effect	21	
58-60	energy	20	II
	paper	20	II
	world	20	III
61-64	ceramic	19	II
	synthesis (-tic)	19	I
	control	19	II
	ozone	19	II
65-71	bacteria	18	I
	EC	18	III

Продолжение табл. 3

Ранг	Слово	Частота	Классификационная группа
	India	18	III
	investment	18	IV
	meeting	18	
	production	18	
	budget	18	IV
72-77	agreement	17	
	biology (-ical)	17	
	molecular (-e)	17	
	coal	17	II
	resistant (-ce)	17	
	system	17	
78-80	cosmetic	16	II
	facility	16	IV
	study	16	
81-85	high	15	
	inhibitor	15	
	source	15	
	China	15	III
	temperature	15	

* В основном в сочетании со словом *virus*.

** В основном в сочетании со словом *genetic*.

*** В основном в сочетании со словом *nuclear*.

**** В сочетании со словом *bill*.

смежных наук и технологий; провели содержательный и лексический анализ, в том числе анализ СВС, этих сообщений. Работа в определенной степени продолжает выяснение возможностей библиометрических методов применительно к торгово-экономической (trade) литературе [28].

МЕТОДОЛОГИЯ

Для исследования выбраны научные новости, представленные в виде «экстрактов» в бизнес-ориентированной базе данных Chemical Industry Notes (CIN), которая выпускается Chemical Abstracts Service. Нами использован печатный вариант этой базы данных за 1989 г. Это издание предназначено прежде всего тем, кто принимает решения и ответственен за управление, инвестиции, маркетинг и производство в химической промышленности», как извещает вводная часть каждого (еженедельного) выпуска CIN. Экстракты распределены по разделам: Продукция, Цены, Продажи, Установки, оборудование, Продукты и процессы, Деятельность корпораций, Деятельность правительства, Персонал. Указатель ключевых слов и названий компаний приводится в каждом выпуске и выходит отдельно один раз в год.

Поиск научных новостей, встречающихся в нескольких разделах, проведен нами по ключевому слову *исследование (research)* в годовом указателе ключевых слов. Найдено 1176 таких сообщений, что составляет 2,3% общего количества экстрактов. Экстракты включают часть оригинального текста новостей (от нескольких процентов до нескольких десятков процентов текста): как правило, начальную часть текста, описывающую «главное событие».

Для подробного изучения содержания и источников этих новостей сформирована систематическая выборка в количестве 96 экстрактов. Их просмотр показал, что они опубликованы в 37 периодических изданиях (табл. 1), издаваемых главным образом в США. Содерж-

жание новостей, каким оно представляется авторам, отражено в табл. 2.

Каждый экстракт индексирован несколькими терминами: в среднем на одно сообщение приходится 4,8 слова, включая слово *исследование*. Такой же, как правило, набор ключевых слов служит заголовком экстрактов новостей в отдельных выпусках журнала. Эти индексирующие термины и использованы нами для лексического и предметного анализа новостей. В табл. 3 представлены наиболее часто встречающиеся (не менее 15 раз) слова, их частоты и принадлежность к той или иной предметной области (группе понятий), причем в основном классифицированы только те научные термины, которые более или менее определенно связаны с конкретными областями ИР. Классификационные группы охарактеризованы в табл. 4.

Таблица 4

Классификация ключевых слов		
№ группы	Предметная область	Количество слов
I	Биомедицина, биохимия, биотехнология, фармация	19
II	Химия, химическая технология, техника, окружающая среда	19
III	Страны, регионы	9
IV	Финансирование и организация исследований	12
Всего		59

По ключевым словам, указанным в табл. 3, проведено сопоставление тем научных новостей с тематикой научных статей, представленных в РЖ Chemical Abstracts за этот же 1989 г. (табл. 5). Кроме того, для сравнения было изучено мнение экспертов, входивших в состав специалистов комитета в США в 80-х гг. Этот комитет, возглавлявшийся Дж. Пиментелом, сформулировал приоритетные области ИР в химии, направленных на решение общественных и экономических

вопросов. Результаты работы комитета изложены в соответствующем докладе [34]. Здесь мы воспользовались более поздней версией этой работы [35] и сопоставили как заголовки ее разделов, так и их содержание с ключевыми словами в CIN.

Число новостей, относящихся к ИР в различных странах, сравнивали с другими национальными (научными

Таблица 5

Сопоставление CIN с СА

Раздел СА	Относительное число рефератов (по данным [33])	Соответствующая группа слов в CIN	Относительное число экстрактов
Биохимия	100	I	100
Прикладная химия и химическая технология	46	II*	40
Макромолекулы	15		16

* Термины этой группы разделены на две подгруппы соответственно разделам СА

и экономическими) показателями. Эти новости выбирали по присутствию названий стран (с любой частотой) среди индексов. Соответствующие ранговые коэффициенты корреляции Кендалла для различных рядов показателей приведены в табл. 6.

Темы научных новостей и связи между ними кроме того определены на основе метода СВС. В качестве показателя силы связи между словами использована частота их совместного присутствия в группах индексов в одних и тех же новостях, отнесенная к квадратному корню из произведения отдельных частот этих слов (формула косинуса). Были построены карты СВС двух типов. Во-первых, использованы лишь научные и технические термины (слова категорий I и II в табл. 3). Такого рода предметная карта научных новостей показана на рис. 1. Во-вторых, построена карта, в которой учтены слова групп III и IV и другие термины, которые связаны с ними (рис. 2).

Таблица 6

Коэффициенты ранговой корреляции Кендалла различных национальных информационных и экономических показателей*

Показатели	Число		Общее число сообщений в CIN	Затраты на ИР [33]	Объем продаж химической продукции [36-38]
	рефератов в СА [33]	научных новостей в CIN			
Число рефератов СА		0,7±0,22	0,59±0,27	0,79±0,27	0,87±0,22
Число научных новостей в CIN			0,72±0,17	0,72±0,20	0,76±0,22
Общее число сообщений в CIN				0,59±0,22	0,73±0,23
Затраты на ИР					0,83±0,20
Объем продаж химической продукции					

* Ряды колеблются от 10 до 32 сопоставимых показателей, подавляющее большинство которых относится к 1989 г.

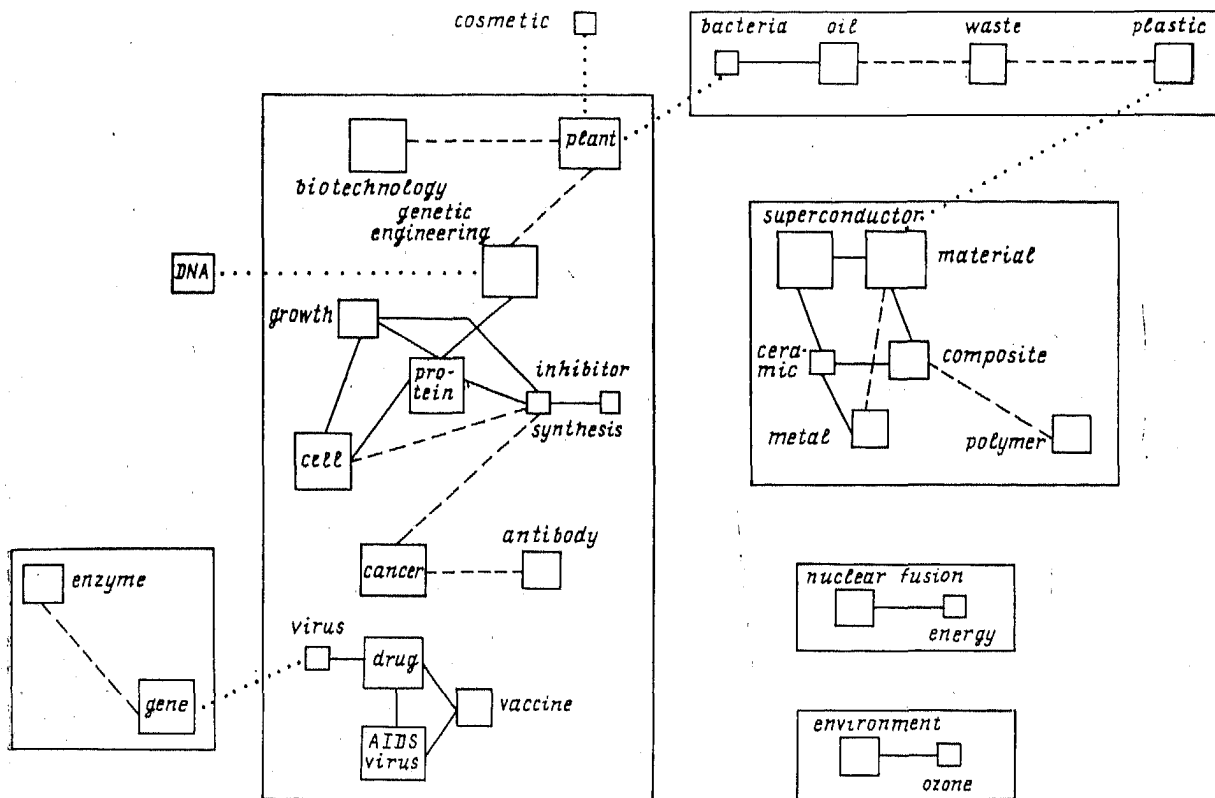


Рис. 1. Карта совместной встречаемости научных терминов. Кластеры образованы за счет сильных (сплошные линии) и средних (пунктир) связей. Связи этого вида составляют по 20% от общего количества связей всех слов. Более слабые связи (точки) показаны только между кластерами или между кластерами с отдельными терминами. Размер квадратов приблизительно соответствует частоте встречаемости терминов, отдельные термины не показаны

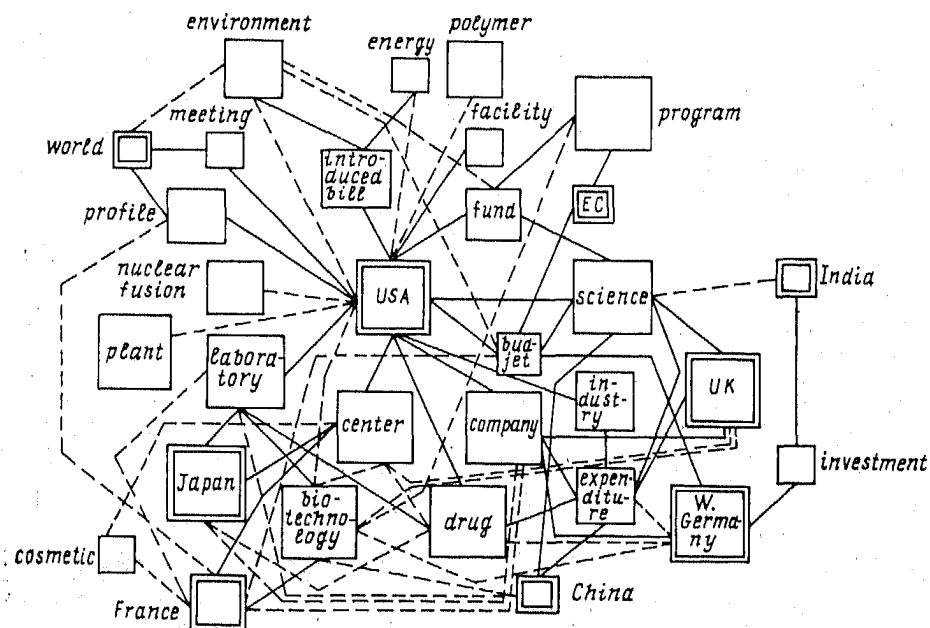


Рис. 2. Карта терминов, относящихся к научной политике и управлению ИР. Оставлены слова группы III и IV и другие термины, если они связаны с первыми, за исключением слов, имеющих здесь достаточно общее значение (например «Chemical»). Показаны только сильные (сплошные линии) и средние (пунктир) связи, составляющие по 20% от общего количества связей всех слов. Размер квадратов приблизительно соответствует частоте встречаемости терминов.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Темы новостей

Обсуждение этого вопроса целесообразно начать с собственных оценок выборки новостей. Данные табл. 2 показывают, что более 1/3 новостей сообщают о новых исследованиях (назовем новости этого типа «предметными»). Это — аналоги обычных научных статей или описаний изобретений к патентам. Часть сообщений имеет оперативный характер и, по-видимому, опережает соответствующие научные публикации. Как это следует из характера базы данных и типа соответствующих периодических изданий, обсуждаемые исследования должны быть, как правило, прикладными. Анализ выборки экстрактов подтверждает это, но показывает, тем не менее, что некоторые исследования являются фундаментальными, хотя они и выполнены в прикладных или потенциально прикладных областях. Такого рода исследования с более отдаленным сроком применения результатов можно отнести к стратегическим. Лексика значительной части новостей показывает, что они отражают исследования в таких областях, которые традиционно считаются стратегическими: биотехнология, полупроводники, композиционные материалы (см. соответствующие термины в табл. 3 и на рис. 1). Таким образом, научные новости в базе данных CIN выступают как индикаторы прикладных и стратегических исследований.

Высокая частота слова *компания* (см. табл. 3) предполагает, что значительная доля ИР выполнена или будет выполнена в промышленных компаниях. Действительно, анализ выборки экстрактов показывает, что на долю исследований в фирмах приходится около 40% новостей. Остальные сообщения распределяются по секторам следующим образом: государственные и, по-видимому, неприбыльные организации — 19%, университеты — 12%, исследования в нескольких секторах и неопределенные ситуации — 30%. Подобные пропорции с преобладанием промышленности типичны как раз для прикладных исследований [33].

Приблизительно половина новостей, хотя и может быть связана с той или иной научной дисциплиной (тематикой исследований), относится к организации науки, научной политике, управлению ИР («научно-политические новости»). Так, сообщения о финансировании исследований (инвестициях, бюджете, др.) и новых исследовательских центрах (лабораториях) уступают по своему числу лишь чисто научным новостям (см. табл. 2), и соответствующие термины группы IV входят в число наиболее часто встречающихся (см. табл. 3).

Области исследований в научных новостях

Первая из двух групп научных и технических терминов относится к биохимии (в СА так называется раздел, охватывающий всю тематику в этой области), биомедицине, генной инженерии, биотехнологии, фармации. Вторая группа терминов — к химической технологии, прикладной химии, экологии, энергетике, материаловедению и др. В СА им соответствуют разделы «Прикладная химия и химическая технология» и «Макромолекулы». Необходимо отметить, что слова групп I и II, как правило, представляют собой макро- или мезотермины. Им соответствуют крупные разделы науки и технологии (например, «Биотехнология» или «укрупненные» понятия: *белок (protein)*, *полимер (polymer)* и т. д.).

Данные табл. 5 показывают, что имеется устойчивая количественная связь между числом рефератов статей в

аналогичных разделах СА и числом релевантных экстрактов в CIN. Таким образом, научные новости хорошо воспроизводят макропропорции* в прикладных и потенциально прикладных исследованиях, выполненных в мире в целом. Отметим здесь же, что в CIN имеется сравнительно немного экстрактов, отвечающих фундаментальным исследованиям в области химии, и поэтому разделы СА «Физическая химия» и «Органическая химия» отсутствуют в табл. 5.

Темы исследований, представленных в CIN, также оказываются значимыми, если их сравнивать с мнениями экспертов о приоритетных направлениях ИР в химии [35]. Итак, содержание главы «Химия и удовлетворение потребностей человека» и всех 7-и крупных разделов в главе «Химия и состояние окружающей среды», где освещаются все прикладные вопросы химии, биомедицины и биотехнологии, полностью или частично отражено в научных новостях. Если сравнить на мезоуровне, которому соответствуют подразделы [35], то наш анализ показывает, что для 22 из 47 менее крупных тем найдены аналоги в научных новостях. Это, конечно, не полное, но все-таки достаточно хорошее соответствие. Действительно, здесь, как и в случае с СА, сравнение проводилось при неполноте сопоставимой терминологии и всего лишь для однолетнего массива научных новостей. Корреляция последних с мнениями экспертов, а также с потоком научных публикаций (СА) на мезо- и микроуровнях — в случае более узких направлений — требует дальнейших исследований с использованием больших массивов новостей и, по-видимому, дополнительного их индексирования (см. далее).

География научных новостей

В CIN отражены результаты или планы исследований в приблизительно 30 странах, если судить по наличию их названий в индексах. Лишь семь стран (см. табл. 3) представлены 15-ю и более сообщениями, и поэтому установленные нами количественные корреляции числа научных новостей с другими индикаторами статистически более значимы именно для таких стран-лидеров.

Данный показатель хорошо коррелирует (см. табл. 6) с научным библиометрическим (число рефератов статей в СА) и научным экономическим (национальные затраты на все ИР), а также с чисто экономическим показателем (продажи химической продукции). Несколько хуже, но, тем не менее, удовлетворительно коррелиция с информационным экономическим показателем — полным числом сообщений в CIN. Этот сложный по смыслу показатель в целом дает менее сильные парные связи с другими показателями (и в отличие от них), хотя в пределах погрешности совпадает с остальными коэффициентами корреляции.

Первая из отмеченных корреляций еще раз подтверждает хорошее количественное соответствие между научными новостями и потоком традиционных документов (статьи), охваченным СА. Наблюдается и хорошее соответствие этих двух библиометрических и двух экономических показателей — затрат на ИР и объемов продаж (см. табл. 6). Это подтверждает действенность моделей вход-выход в науковедении и научной политике (см. [39, 40]) и возможность использования в них показателя «число научных новостей» (см. далее).

* Попытки установить связи между более узкими направлениями, которым соответствуют отдельные индексированные термины в CIN и подразделы СА, не привели к достаточно сильным корреляциям. Причина этого в определенной степени в том, что индексированные термины CIN и слова в заголовках подразделов и индексированные термины СА не совпадают. Кроме того, массив научных новостей, возможно, в данном случае недостаточно представительен. Действительно, одна научная новость приходится на более чем на 400 документов СА.

Примечательно, что хорошую корреляцию обеспечивают оба экономических показателя (также тесно связанных между собой), хотя смысл связи очевиден только в случае первого из них (ассигнование на ИР), отражающего вводимые ресурсы. Очевидно, процент отчислений от продаж химической продукции на ИР примерно одинаков в различных странах, что и обеспечивает приблизительно такое же соотношение между национальными объемами продаж, как и в случае общих затрат на ИР. Известны подобные сильные корреляции между ВВП и ассигнованиями на науку [40]. Удовлетворительно коррелируют между собой ВВП и число статей в этих странах в SCI [41]. Наш подсчет по данным [41] для 30 стран (лидеров по числу статей в 1984 г.) дает ранговый коэффициент корреляции последних двух показателей $0,63 \pm 0,20$, что близко к другим его значениям в табл. 6.

Научные новости как смешанный индикатор

Рассматриваемые сообщения не только гетерогенны в тематическом плане, но и в отличие от других обсуждаемых в предыдущем разделе показателей не занимают определенного положения в бинарной классификации вход—выход. Если экономические показатели являются входными, число рефератов в СА — выходным, то число новостей (экстрактов в CIN) представляет собой смешанный показатель, указывающий и на входные ресурсы, и на выходные результаты ИР.

Информация о новых исследовательских центрах (лабораториях) указывает на новые (перемещенные) научные кадры, на новое оборудование, приборы и т. д., которые будут использоваться в ИР. О соответствующих и других материальных затратах свидетельствуют сообщения о финансировании исследований. Программы и планы ИР, а также и новые центры можно рассматривать как некоторые организационные единицы входа в науку, связанные также с ассигнованиями и человеческими ресурсами. В целом число этих и некоторых других сообщений о входе в исследования (другими словами, научно-политических новостей) сопоставимо с числом сообщений об уже состоявшихся исследованиях (их выход — предметные новости). Таким образом, научные новости, точнее различные их массы, можно использовать, в том числе совместно, и как индикаторы входа, и как индикаторы выхода в соответствующих моделях науки.

Признаками входных новостей являются ключевые слова группы IV. Подборка экстрактов показывает, что предметные новости в их большей части (строка 1 в табл. 2) индексируются лишь словами групп I и II, и это основной признак выходных экстрактов. Наличие названий стран (регионов) среди индексов не является прямым признаком отнесения новостей к входным или выходным, хотя в большинстве случаев реализуется первая ситуация. Например, в сообщениях о новых научных центрах практически во всех случаях присутствуют эти термины. Те же «географические» экстракты, которые относятся к выходным, можно выбрать по другим признакам (ключевым словам). Так, термин *обзор* (*overview* или *survey*) соответствует обзорам исследований, проведенных в тех или иных странах.

Однозначное разделение экстрактов на две обсуждаемые группы требует, по-видимому, более строгого и расширенного индексирования научных новостей. Например, среди индексов нами обнаружено отсутствие соответствующих слов группы III в отдельных случаях, когда речь идет об ассигнованиях на науку.

Карта предметных новостей

Карты, построенные с использованием химических, биологических и технических терминов, аналогичны картам науки, созданным при использовании СВС на основе научных публикаций [8—13, 15, 17—21]. Если же проводить сравнение с картами социотирования, то, учитывая укрупненность индексируемых терминов в нашем случае, им должны соответствовать карты уровня С3 или С4 [2]. Исходя из этого, мы рассматриваем сеть слов аналогично сети кластеров (тем), а число связей с другими словами — как показатель центральности и соответственно стратегической значимости данной проблемы.

Центральное место на нашей карте предметных новостей (см. рис. 1) занимает большой кластер (14 слов или словосочетаний), включающий большинство часто встречающихся биотерминов. В нем выделяется нижняя часть, которая имеет сильные внутренние связи и могла бы быть самостоятельным кластером при другом алгоритме кластеризации, например ограничивающем его размер пятью словами. Этот подкластер охватывает исследования главным образом в области фармации и биомедицины. Сам термин *лекарство* (*drug*) занимает центральное положение. Это вполне соответствует тому, что в химической промышленности наибольшая доля ассигнований на ИР связана с лекарствами [33].

Средняя часть биокластера охватывает тесно переплетенные между собой различные исследования в пересекающихся областях: биотехнологии, генной инженерии, биохимии. Центральное положение занимает мезотермин *белок* как объект или цель таких исследований. Макротермины *генная инженерия* (*genetic engineering*) и особенно *биотехнология* (*biotechnology*) занимают более периферийную позицию в верхней части кластера. Это положение «биотехнологии» связано с техническими причинами, а именно особенностью индексирования: это слово (в отличие от *лекарства*) относительно редко упоминается в предметных новостях в связках с другими мезотерминами, которые входят в состав биокластера. В сущности, центральное место биотехнологии здесь определяется более узкими терминами (тем же *белком*), относящимися к родовому понятию *биотехнология* (см. рис. 1). Наоборот, макротермин *биотехнология* (как и *лекарство*) сравнительно часто присутствует среди индексов научно-политических новостей и на соответствующей карте (рис. 2) оказывается одним из узловых (см. далее).

Другой достаточно крупный кластер на карте научных новостей относится к междисциплинарной теме «материалы». Как и в рассмотренном биокластере, в нем наблюдается сравнительно высокая плотность связей между словами, что указывает на высокую степень взаимодействия между проблемами, на сформировавшуюся широкую область исследований.

Остальные связи между терминами групп I и II достаточно редки и слабы, за исключением, может быть, нескольких тем исследований *нефть* (*oil*), *пластики* (*plastic*), связанных с проблемой *отходов* (*waste*) (см. рис. 1). Связи между терминами группы I почти в три раза слабее в среднем, чем у биотерминов (группа II). Это наблюдение с некоторой степенью приближения указывает, что различные исследования в области прикладной химии, химической технологии не отличаются согласованностью.

Связи между кластерами (или между кластерами и словами, не вошедшими в кластер) также оказываются редкими и слабыми. Центральное место в этой рыхлой лексической сети занимает опять-таки биокластер, что еще раз демонстрирует стратегическую важность биотехнологии и родственных областей ИР в политике

наиболее развитых государств и их фирм. Подобная биотехнологизация химической технологии (промышленности) в данном случае выглядит как вторжение информации об ИР в биообластях в чисто химическую информацию и дискриминация последней. Действительно, химические проблемы в традиционном их определении сравнительно редки на предметной карте (см. рис. 1), и требуются специальные приемы, чтобы сделать их более представительными. (Этому могло бы способствовать, например, искусственное отделение «химических» новостей и уменьшение порога частоты анализируемых слов.) Аналогичную проблему в свое время решал Г. Смолл, ограничивая экспансию биомедицины и биохимии на общенаучных картах социтивирования [1, 2].

Научно-политическая карта

Сеть слов, построенная с использованием как основы терминов групп III и IV (см. рис. 2), является гораздо более плотной, чем рассмотренная выше сеть научных и технических терминов. Кластеры можно выделить в этом случае, используя пороговые значения силы связи или ограничивая кластеры по размеру. Однако эти образования будут включать различные по смыслу термины, т. е. окажутся гетерогенными, и их будет довольно трудно интерпретировать. Поэтому целесообразнее обсудить «научно-политическую» сеть в целом, отмечая отдельные наиболее важные ее узлы.

Главное стратегическое место в этой сети занимает аббревиатура США (USA), имеющие 18 сильных или средних по силе связей с другими терминами. Часть последних идентична тем, с которыми связаны и другие развитые страны (Япония, Западная Германия, Великобритания, Франция): со словами *центр (center)*, *лаборатория (laboratory)*, *затраты (expenditure)* и др. Остальные же термины встречаются только в связи с США. К ним относятся несколько предметных областей ИР — *энергия (energy)*, *полимеры (polymer)* и др. Важный факт — присутствие в окружении аббревиатуры США термина *закон (introduced bill)*, демонстрирующего что карты этого типа показывают роль законодательства в научно-технической политике развитых стран. Тот же факт, что по данной карте нельзя увидеть указанные области исследований в других странах помимо США, объясняется, возможно, особенностью базы данных CIN, имеющей отчетливый уклон в сторону американских изданий (см. табл. 1) и, следовательно, соответствующей информации.

Стратегические положения на этой карте Японии, Франции, Западной Германии и Великобритании сопоставимы. Каждая из этих «стран» имеет 5–8 связей, причем каждая из них с любой другой имеет 3–4 аналогичных связи (с одним и тем же словом). Общие узлы сети для всех названных четырех стран — *биотехнология* и *компания*, что еще раз указывает на основную здесь стратегическую область ИР и основной их сектор. Китай (4 связи) и особенно Индия (2 связи) занимают на карте более периферийное положение.

Заслуживают внимания еще несколько деталей этой карты в связи с возможным ее применением. Присутствующее на ней слово *наука (science)* отражает сообщения о национальных ассигнованиях на научные исследования; с этим согласуется контекст данного термина — названия государств и финансовые термины (*фонд (fund)*, *бюджет (budget)*, *затраты*). Научная активность на региональном уровне представлена на карте *программами (program)* ЕС в области *биотехнологии*. Глобальные черты науки прослеживаются здесь через термин *мир (world)*.

Среди тематических областей ИР центральное место принадлежит, как указывалось, *биотехнологии* (9 свя-

зей — второй по величине показатель после США). Менее центральны — *лекарства* и *окружающая среда (environment)*. Тот факт, что данная карта демонстрирует эти области, доказывает пригодность научных новостей в CIN как индикатора, стратегических ИР. С другой стороны, отсутствие или периферийное положение здесь других важных тем (энергетика, материалы) может быть следствием многих факторов, в том числе и того, что эти темы на самом деле стратегически более периферийны.

Отдельно следует сказать о том, что на характер данных карт также влияет, по-видимому, специфическое индексирование в CIN и общий «эффект индексиатора». Во-первых, число индексов для каждого экстракта невелико и, на наш взгляд, недостаточно для детального описания каждого из них. Этот факт отмечен ранее для базы данных СА [8]. Кроме того, обнаружены и такие случаи, когда сообщения о результатах ИР не индексируется словом *исследование*. Во-вторых, индексирующие стремятся присваивать документам индексы более общего характера и в меньшей степени специфические для конкретных исследований отдельных авторов (см. [19]). В-третьих, база данных CIN — бизнес-ориентированная, и эффект индексиатора может также проявляться в отборе индексирующих терминов. Таким образом, карта, показанная на рис. 2, отражает многие важные, но далеко не все события в мире научной политики, планирования и управления ИР. (Аналогичный вывод в той или иной степени относится и к предметным картам.) По-видимому, для более детального и глубокого изучения научной политики с использованием метода СВС необходимо специальное расширенное индексирование научных новостей с предварительным созданием релевантного тезауруса. В этом случае многие полезные данные, например, связи между различными секторами ИР, о чем есть свидетельства в экстрактах новостей, будут отражены и на картах.

ЗАКЛЮЧЕНИЕ

В настоящей работе изучены научные новости в бизнес-ориентированных периодических изданиях, собранные в виде экстрактов в базе данных CIN, проведен содержательный анализ этих экстрактов и лексический анализ соответствующих индексирующих терминов. Таким образом, темы новостей определены по содержанию экстрактов, наиболее часто встречающимся индексирующим терминам и группировкам взаимосвязанных терминов, в том числе на картах СВС.

Научные новости описывают различные виды деятельности. Большинство новостей представляет собой сообщения (чаще всего оперативные) либо о новых результатах научных исследований, либо о различных событиях (сводках данных), касающихся сферы научной политики, планирования и управления ИР — новых лабораториях, ассигнованиях на науку, программах ИР и др. Сообщения первого типа (предметные новости) можно считать своеобразными отчетами об ИР и отнести к выходным показателям науки. Новости второго типа описывают входные ресурсы науки.

Тематически научные новости представлятельно отражают те области ИР, которые в последние годы традиционно рассматриваются как приоритетные в научной и технологической политике наиболее развитых стран и промышленных компаний химического и смежного профиля: биотехнологию, генную инженерию, фармацевтику, материалы (полупроводниковые, композиционные), защиту окружающей среды. Число новостей, суммированных по крупным областям ИР или отдельным странам, хорошо воспроизводит эти же пропорции в мировом потоке научных публикаций. Число «национальных» новостей коррелирует с затратами на ИР и объе-

мом продаж химической продукции в этих странах.

Карта СВС, построенная с использованием наиболее часто встречающихся естественнонаучных и технических терминов, содержит центральный большой кластер тесно взаимосвязанных исследований в области биотехнологии, фармации, биохимии. Наблюдается и другая относительно большая и плотная область связанных исследований — материалы. Такого рода научная карта напоминает другие карты науки, основанные на научных статьях, но отличается от них тем, что выделяет прикладные, промышленные и стратегические исследования.

Другие распространенные индексующие термины позволяют создать прямую карту научной политики. На ней представлены основные «актеры» (наиболее развитые страны), национальные науки в целом, макрообласти ИР, организационные научные структуры (лаборатории, центры), промышленность (компания), финансовые и материальные ресурсы и др. Эта карта оказывается сложной (гетерогенной по смыслу, высокоплотной) конструкцией, интегрирующей в основном макрособытия и макротенденции. Наблюдается перекокс в сторону освещения научной политики США в сравнении с другими странами — из-за преобладания соответствующей информации. Использование таких карт для анализа научной политики представляется перспективным, но требует совершенствования техники индексирования и технологии создания карт.

Научные новости в целом можно оценить как самостоятельный источник информации о науке и научной политике, не совпадающий с другими типами документов (например, статьями, описаниями изобретений к патентам), также отражающими результаты ИР. Возможность использования научных новостей данного типа в наукометрии определяется оригинальностью и полезностью полученных с их помощью данных, которые относятся к прикладным, промышленным, стратегическим исследованиям в «чистом» виде, научной политике, организации и планированию ИР.

На основе этих документов можно также проводить прогнозирование ИР, анализируя динамику карт СВС или изучая входные новости, которые указывают на возможные научные результаты в будущем.

СПИСОК ЛИТЕРАТУРЫ

1. Small H., Sweeney E. Clustering the Science Citation Index using co-citation. I. A comparison of methods // *Scientometrics*.— 1985.— Vol. 7.— P. 391.
2. Small H., Sweeney E., Greenlee E. Clustering the Science Citation Index using co-citation. II. Mapping science // *Scientometrics*.— 1985.— Vol. 8.— P. 321.
3. Franklin J. J., Johnston R. Co-citation bibliometric modeling as a tool for S and T policy and R and D management: Issues, applications and developments // *Handbook of Quantitative Studies of Science and Technology* / A. F. J. Van Raan (Ed.).— Amsterdam: Elsevier Science Publishers B. V. (North-Holland), 1988.— P. 325.
4. Milman B. L., Gavrilova Yu. A. Analysis of citation and co-citation in chemical engineering // *Scientometrics*.— 1993.— Vol. 27.— P. 53.
5. Браам Р. Р., Моед Н. Ф., Ван Раан А. Ф. Mapping of science by combined co-citation and word analysis. I. Structural aspects // *J. of the American Society for Information Science*.— 1991.— Vol. 42.— P. 233.
6. Браам Р. Р., Моед Н. Ф., Ван Раан А. Ф. J. Mapping of science by combined co-citation and word analysis. II. Dynamical aspects // *J. of the American Society for Information Science*.— 1991.— № 42.— P. 252.
7. Milman B. L. Individual co-citation clusters as nuclei of complete and dynamic informetric models of scientific and technological areas // *Scientometrics* (submitted).
8. Callon M., Courtial J.-P., Laville F. Co-word Analysis: a Tool for the Evaluation of Public Research Policy // *CSI Report*.— 1989.— March.
9. Courtial J.-P., Law J. A co-word study of artificial intelligence // *Social Studies of Science*.— 1989.— Vol. 19.— P. 301.
10. Courtial J.-P., Michelet B. A mathematical model of development in research field // *Scientometrics*.— 1990.— Vol. 19.— P. 127.
11. Le Marc M., Courtial J.-P., Drozda Senkowska E., Petard J.-P., Py Y. The dynamics of research in the psychology of work from 1973 to 1987: From the study of companies to the study of professions // *Scientometrics*.— 1991.— Vol. 21.— P. 69.
12. Courtial J.-P., Callon M. Indicators for the identification of strategic themes within a research programme // *Scientometrics*.— 1991.— Vol. 21.— P. 447.
13. Callon M., Courtial J.-P., Laville F. Co-word analysis as a tool for describing the network of interactions between basic and technological research. The case of polymer chemistry // *Scientometrics*.— 1991.— Vol. 22.— P. 155.
14. Courtial J.-P., Calon M., Sigogneau A. The use of patent titles for identifying the topics of invention and forecasting trends // *Scientometrics*.— 1993.— Vol. 26.— P. 231.
15. Law J., Bauin S., Courtial J.-P., Whittaker J. Policy and the mapping of scientific change. A co-word analysis of research into environmental acidification // *Scientometrics*.— 1988.— Vol. 14.— P. 251.
16. Turner W. A., Charton G., Laville F., Michelet B. Packing information for peer review: new co-word analysis techniques // *Handbook of Quantitative Studies of Science and Technology* / A. F. J. Van Raan (Ed.). Amsterdam: Elsevier Science Publishers B. V. (North-Holland), 1988.— P. 291.
17. Bauin S., Michelet B., Schweighofer M. G., Vermeulin P. Using bibliometrics in strategic analysis: «understanding chemical reactions» at the CNRS // *Scientometrics*.— 1991.— Vol. 22.— P. 113.
18. Turner W. A., Rojonan F. Evaluating input/output relationships in a regional research network using co-word analysis // *Scientometrics*.— 1991.— Vol. 22.— P. 139.
19. Whittaker J. Creativity and Conformity in Science: Titles, Keywords and Co-word Analysis // *Social Studies of Science*.— 1989.— Vol. 19.— P. 473.
20. Law J., Whittaker J. Mapping acidification research: a test of the co-word method // *Scientometrics*.— 1992.— Vol. 23.— P. 417.
21. Braam R. R., Moed H. F., Van Raan A. F. J. Comparison and combination of co-citation and co-word clustering // *Science and Technology Indicators: Their Use in Science Policy and Their Role in Science Studies* / A. F. J. Van Raan, A. J. Nederhof, and H. F. Moed (Eds.).— Leiden: DSWO-Press, 1989.— P. 307.
22. Peters H. P. F., Van Raan A. F. J. Co-word based science maps of chemical engineering.: Research Report CW TS—91—03 / Centre for Science and Technology Studies.— Leiden, 1991, April.

23. Leydesdorf L. Words and co-words as indicators of intellectual organization // *Research Policy*.— 1989.— Vol. 18.— P. 209.
24. Leydesdorf L. In search of epistemic networks // *Social Studies of Science*.— 1992.— Vol. 21.— P. 75.
25. Leydesdorf L. A validation study of «LEXI-MAPPE» // *Scientometrics*.— 1992.— Vol. 25.— P. 295.
26. Mapping the Dynamics of Science and Technology / M. Callon, J. Law, A. Rip (Eds).— L.: Macmillan, 1986.
27. Healey P., Rothman H., Hoch P. K. An experiment in science mapping for research planning // *Research Policy*.— 1986.— Vol. 15.— P. 233.
28. Van Steijn F., Rip A. The role of trade literature in the communication system // *Scientometrics*. 1988.— Vol. 13.— P. 81.
29. Van Vianen B. G., Moed H. F., Van Raan A. F. J. An exploration of the science base of recent technology // *Research Policy*.— 1990.— Vol. 19.— P. 61.
30. Narin F., Olivastro D. Technology indicators based on patents and patent citations // *Handbook of Quantitative Studies of Science and Technology / A. F. J. Van RAAN (Ed.)*.— Amsterdam: Elsevier Science Publishers B. V. (North-Holland), 1988.— P. 465.
31. Mauguin P. Using a contracts database for evaluating the dynamics of a technological program: The case of «Non-Nuclear Energy» program // *Scientometrics*.— 1991.— Vol. 22.— P. 207.
32. Rip A. Mapping of science: possibilities and limitations // *Handbook of Quantitative Studies of Science and Technology / A. F. J. Van Raan (Ed.)*.— Amsterdam: Elsevier Science Publishers B. V. (North-Holland), 1988.— P. 253.
33. Facts and figures for chemical Rand D. // *Chemical and Engineering News*.— 1991.— Vol. 69, № 33.— P. 32.
34. Pimentel G. Opportunities in Chemistry.— Washington: National Academic Press, 1985.
35. Пиментел Дж., Куррод Дж. Возможности химии сегодня и завтра.— М.: Мир, 1992.
36. Facts and figures for the chemical industry // *Chemical and Engineering News*.— 1990.— Vol. 68, № 25.— P. 34.
37. World chemical outlook // *Chemical and Engineering News*.— 1990.— Vol. 68, № 50.— P. 24.
38. Дедов А. Г., Сафонова И. Л., Никулина Е. П. и др. Основные направления химизации экономики капиталистических стран.— М.: Химия, 1988.
39. King J. A review of bibliometric and other science indicator and their role in research evaluation // *J. of Information Science*.— 1987.— Vol. 13.— P. 261.
40. Яблонский А. И. Математические модели в исследовании науки.— М.: Наука, 1983.— С. 151.
41. Frame J. Modelling national technological capacity with patent indicators // *Scientometrics*.— 1991.— Vol. 22.— P. 327.

Материал поступил в редакцию 10.03.94.