

## Использование ключевых слов в информационно-поисковой системе по проектам в области радиоэкологии

*Предлагается включить ряд дескрипторов и словосочетаний в список ключевых слов баз данных ВИНТИ по экологии и охране окружающей среды для повышения эффективности поиска информации по радиоэкологической тематике. Представлена методика дополнения списка ключевых слов радиоэкологической направленности и проанализировано их использование в информационно-поисковой системе по проектам в области радиоэкологии.*

*Дается обоснование целесообразности расширения перечня ключевых слов, приводятся примеры проектов в области радиоэкологии.*

### ВВЕДЕНИЕ

Разрозненность тематически многоаспектной информации по экологическим исследованиям и разработкам, проводимым научными и производственными организациями в различных предметных областях, ведет к дублированию, нечеткой координации исследований и нерациональному использованию финансовых средств. Для эффективного поиска и минимизации степени дублирования планируемых экологических проектов, целесообразно построение специализированных информационных систем, содержащих информацию о проектах, таких, например, как созданная в лаборатории геоинформатики ИГЕМ РАН система “ЭкоПро” [1], которая содержит сведения о проектах в области экологии (около 1300 наименований информационных документов) и предусматривает различные режимы поиска проектов (по теме, по автору, по исполнителю, по ключевым словам, по территориальной привязке).

В связи с повышением значимости радиоэкологических проблем в России становятся более актуальными вопросы информационного обеспечения и в этой сфере. В плане развития информационно-поисковой системы “ЭкоПро” создается специализированная информационная система по проектам в области радиоэкологии, радиогеоэкологии и близких к ним отраслям природоохранной тематики [2] с целью повышения эффективности поиска, оценки степени дублирования проектов, а также их тематического анализа (по динамике, территории, исполнителям, финансированию). Радиоэкология, являясь частью экологии, содержит междисциплинарные аспекты, и радиоэкологические исследования касаются различных видов природоохранной деятельности. Поэтому необходимо создание более детальных поисковых предписаний и развитых информационно-поисковых языков. В формируемой радиоэкологической базе данных, включающей материалы информационно-поисковой системы “ЭкоПро” (<http://www.ecoprojects.ru>), справочники экологических проблем России [3, 4]

(<http://www.igem.ru/igem/eco/index.html>), содержащих сведения об экологических проектах Российской академии наук, базы данных Международного научно-технического центра (МНТЦ) (<http://www.tech-db.ru/istc/db/projects.nsf/>), исследований лаборатории радиогеоэкологии и радиогеоэкологии ИГЕМ РАН и других источников, на данный момент насчитывается более 300 описаний проектов радиоэкологической и радиационной направленности охраны окружающей среды (Табл. 1, графа 3).

Таблица 1

**Анализируемый объем документов по проектам в области радиоэкологии**

Источник	кол. док. в БД	исследуемые проекты	не определены ключевыми словами	% к исследуемым (п.4/п.3×100%)	определены ключевыми словами (п.3-п.4)	сомнительные по тематике проекты	всего проектов (п.3+п.7)
РАН (справочники экологических проблем России РАН)	ок. 4000	144	22	15,28	122	29	173
БД «Эко Про»	1600	25	1	4,00	24	0	25
МНТЦ (раздел окружающая среда)	неск. тыс.	136	11	8,09	125	18	154
Интернет		12	1	8,33	11		12
Итого по базе данных:		317	35	11,04	282	47	364

На основе данных по проектам, а также дополнительно по материалам публикаций в области радиоэкологии, отобранных в базе данных по геологии, природопользованию и недропользованию Всероссийского научно-исследовательского института

**Количественный анализ поиска информации радиэкологической направленности  
с использованием ключевых слов списка (с учетом поиска в базе данных по радиэкологическим проектам)**

№ столбца	Ключевые слова списка в алфавитном порядке	Проекты в области радиэкологии			Yandex		Rambler		Aport		среднее по поисковым системам	
		общее количество слов в системе	% к общему количеству проектов	% к общему количеству слов	количество найденных сайтов	% к общему количеству	количество найденных сайтов	% к общему количеству	количество найденных документов	% к общему количеству	среднее арифметическое	% к общему среднему
1	АТОМНАЯ ПРОМЫШЛЕННОСТЬ	1	0,32	0,30	1296	3,00	18466	3,70	4125	4,45	1377,72	4,30
2	АТОМНАЯ ЭНЕРГЕТИКА	11	3,47	3,31	997	2,31	16350	3,28	2489	2,68	831,65	2,60
3	АТОМНЫЕ ПОДВОДНЫЕ ЛОДКИ	3	0,95	0,90	884	2,05	10603	2,13	2364	2,55	789,56	2,47
4	АТОМНЫЕ УСТАНОВКИ	1	0,32	0,30	1299	3,01	18268	3,66	2580	2,78	862,15	2,69
5	АЭС	13	4,10	3,92	1896	4,39	22688	4,55	1074	1,16	359,90	1,12
6	ЕСТЕСТВЕННЫЕ РАДИОНУКЛИДЫ	1	0,32	0,30	1179	2,73	3837	0,77	1237	1,33	413,03	1,29
7	ЖИДКИЕ РАДИОАКТИВНЫЕ ОТХОДЫ	14	4,42	4,22	899	2,08	3668	0,74	1232	1,33	411,35	1,28
8	ИОНИЗИРУЮЩЕЕ ИЗЛУЧЕНИЕ	7	2,21	2,11	1183	2,74	9089	1,82	4210	4,54	1405,45	4,39
9	РАДИАЦИОННАЯ БЕЗОПАСНОСТЬ	2	0,63	0,60	1054	2,44	13767	2,76	3231	3,49	1079,08	3,37
10	РАДИАЦИОННАЯ ОБСТАНОВКА	3	0,95	0,90	1027	2,38	7769	1,56	2867	3,09	957,22	2,99
11	РАДИАЦИОННАЯ ОПАСНОСТЬ	4	1,26	1,20	1299	3,01	7998	1,60	2749	2,97	817,86	2,87
12	РАДИАЦИОННАЯ ЭКОЛОГИЯ	1	0,32	0,30	1093	2,53	6815	1,37	2421	2,61	808,33	2,52
13	РАДИОАКТИВНО ЗАГРЯЗНЕННЫЕ ТЕРРИТОРИИ	1	0,32	0,30	1163	2,69	3916	0,79	1475	1,59	492,46	1,54
14	РАДИОАКТИВНОЕ ЗАГРЯЗНЕНИЕ	20	6,31	6,02	1123	2,60	9265	1,86	3694	3,98	1233,28	3,85
15	РАДИОАКТИВНОСТЬ	16	5,05	4,82	1327	3,07	9321	1,87	2398	2,59	800,82	2,50
16	РАДИОАКТИВНЫЕ ЖИДКИЕ ОТХОДЫ	0	0,00	0,00	914	2,12	3666	0,74	1231	1,33	411,02	1,28
17	РАДИОАКТИВНЫЕ ОТХОДЫ	76	23,97	22,89	975	2,26	12608	2,53	2778	3,00	927,84	2,90
18	РАДИОИЗОТОПЫ	2	0,63	0,60	1347	3,12	1483	0,30	304	0,33	1044,67	3,26
19	РАДИОЛОГИЯ	3	0,95	0,90	971	2,25	3765	0,76	1689	1,82	563,86	1,76
20	РАДИОНУКЛИДЫ	66	20,82	19,88	1327	3,07	8227	1,65	2139	2,31	714,32	2,23
21	РАДИОЭКОЛОГИЯ	19	5,99	5,72	976	2,26	1393	0,28	864	0,93	288,40	0,90
22	РАДОН	2	0,63	0,60	1356	3,14	6032	1,21	2457	2,65	820,29	2,56
23	ЧАЭС	1	0,32	0,30	912	2,11	5456	1,09	1792	1,93	598,34	1,87
24	ЧЕРНОБЫЛЬСКАЯ АЭС	10	3,15	3,01	934	2,16	11900	2,39	3005	3,24	1003,54	3,13
25	ЯДЕРНАЯ БЕЗОПАСНОСТЬ	1	0,32	0,30	838	1,94	23916	4,80	2654	2,86	887,22	2,77
26	ЯДЕРНАЯ ЭНЕРГЕТИКА	3	0,95	0,90	828	1,92	13485	2,70	2424	2,61	809,77	2,53
27	ЯДЕРНАЯ ЭНЕРГИЯ	0	0,00	0,00	1116	2,58	22239	4,46	2924	3,15	977,20	3,05
28	ЯДЕРНОЕ ОРУЖИЕ	0	0,00	0,00	1426	3,30	25161	5,05	2292	2,47	766,51	2,39
29	ЯДЕРНОЕ ТОПЛИВО	22	6,94	6,63	821	1,90	13401	2,69	2233	2,41	746,03	2,33
30	ЯДЕРНО-РАДИАЦИОННЫЕ ПРОИЗВОДСТВА	2	0,63	0,60	774	1,79	1298	0,26	287	0,31	95,86	0,30
31	ЯДЕРНЫЕ АВАРИИ	2	0,63	0,60	1112	2,58	12084	2,42	3008	3,24	1004,56	3,14
32	ЯДЕРНЫЕ ВЗРЫВЫ	6	1,89	1,81	1302	3,02	21072	4,23	3547	3,83	1185,02	3,70
33	ЯДЕРНЫЕ ИСПЫТАНИЯ	2	0,63	0,60	998	2,31	17263	3,46	3182	3,43	1062,96	3,32
34	ЯДЕРНЫЕ МАТЕРИАЛЫ	6	1,89	1,81	969	2,24	29084	5,83	1810	1,95	605,93	1,89
35	ЯДЕРНЫЕ ПОЛИГОНЫ 1	0	0,00	0,00	1060	2,46	9097	1,82	3391	3,66	1132,16	3,54
36	ЯДЕРНЫЕ ПРОБЛЕМЫ	1	0,32	0,30	596	1,38	32883	6,59	3074	3,32	1027,97	3,21
37	ЯДЕРНЫЕ РЕАКТОРЫ --- 7	5	1,58	1,51	1089	2,52	13822	2,77	2205	2,38	736,72	2,30
38	ЯДЕРНЫЕ ТЕХНОЛОГИИ	0	0,00	0,00	898	2,08	25420	5,10	2117	2,28	708,13	2,21
39	ЯДЕРНЫЕ УСТАНОВКИ	0	0,00	0,00	1074	2,49	19628	3,94	2477	2,67	827,87	2,59
40	ЯДЕРНЫЙ ТОПЛИВНЫЙ ЦИКЛ	5	1,58	1,51	844	1,95	2467	0,49	981	1,06	327,52	1,02
	<b>ВСЕГО</b>	<b>332</b>	<b>104,73</b>	<b>100,00</b>	<b>43176</b>	<b>100,00</b>	<b>498670</b>	<b>100,00</b>	<b>92707</b>	<b>100,00</b>	<b>32013,57</b>	<b>100,00</b>

Приложение к Табл. 2

Источник:	max	min	размах	мода	медiana	среднее (арифмети- ческое)	среднее взвешенное	среднее линейное отклонение
<b>радиэкологические проекты</b>	76,00	0,00	76,00	1,00	2,50	8,3	37,57	9,20
<b>Yandex</b>	1896,00	596,00	1300,00	1299	1057,00	1079,4	1127,95	175,19
<b>Rambler</b>	32883,00	1298,00	31585,00	-	11251,50	12466,75	1347,43	6755,98
<b>Aport</b>	4210,00	287,00	3923,00	-	2424,00	2377,103	2694,13	721,26
<b>среднее ариф. по поисковым системам Internet</b>	1405,45	95,86	1309,60		815,03	5332,38	5386,44	228,84
средние величины по поисковым системам Internet			12269,33			5307,751		

экономики минерального сырья и недропользования (ВИЭМС МПР<sup>1</sup>), был проанализирован полный список ключевых слов [5]. Особое внимание уделено ряду ключевых слов, отражающих радиоэкологическую направленность (Табл. 2, графа 1), предусмотренных для поиска информации в общем массиве данных по проблемам окружающей среды и природопользованию, включающем информацию по радиационной экологии. Отобранные слова являются основными признаками радиоэкологической информации (например, словосочетание *радиационная экология* и слова, обозначающие предмет ее изучения), отражающими явления и источники искусственных ионизирующих излучений.

С целью расширения тематики и увеличения глубины поиска радиоэкологической информации, повышения релевантности выдачи данных по запросам пользователей, предлагается пополнить список ключевых слов по радиоэкологической тематике.

## ОСНОВНЫЕ КРИТЕРИИ И МЕТОДИКА АНАЛИЗА ПОИСКА ПО КЛЮЧЕВЫМ СЛОВАМ И ДОПОЛНЕНИЕ СПИСКА КЛЮЧЕВЫХ СЛОВ

Разумеется, радиационная экология включает множество специфических терминов, касающихся объекта изучения, условий, компонентов окружающей среды, в которых он находится, методов, приборов и материалов, а также результатов работ, определяющих оценку релевантности информации по тому или иному документу в данной области.

Соблюдение пропорций между детализацией и укрупнением различных тематических аспектов — важная задача при формировании информационно-поисковых языков. Вопрос о детализации возникает и при поиске информации по ключевым словам.

Правила по составлению информационно-поискового тезауруса описаны в ГОСТ 7.25-2001 [6] и включают следующие пункты:

1) определение тематического охвата (в данном случае это экология и охрана окружающей среды, массив данных, включающий сведения по радиоэкологическим проектам);

2) сбор массива лексических единиц из первоначального массива документов (массив имеющихся лексических единиц представляет собой список ключевых слов, формирование дополнительных слов описано ниже);

3) формирование словника;

4) построение словарных статей и указателей (ссылки, связи, ассоциативные, иерархические и пр., между лексическими единицами);

5) оформление;

6) регистрация.

В данной статье рассмотрены первые три и частично четвертый пункты.

Таким образом, при построении словаря ключевых слов, а также при его корректировке и дополнении следует принимать во внимание:

• Наличие в радиоэкологических документах дескрипторов в форме специфических терминов, сложившихся словосочетаний, а также, отдельно, сокращений, формирующих поисковый образ (общедисциплинарные; касающиеся окружающей среды и ее компонентов; природных, технологических, химических, физических, аналитических и пр. процессов и явлений; радиоактивных веществ и материалов, в том числе использующихся при захоронении радиоактивных отходов; их состояния, свойств и характеристик, методов исследования и анализа в радиоэкологии; приборов и аппаратов исследования ионизирующего излучения и пр.) [5].

Словосочетания, согласно стандартам, формируются в случае, когда [6]:

а) хотя бы один компонент сочетания не употребляется в составе других словосочетаний или употребляется в другом смысле;

б) для данного словосочетания существуют полные синонимы в тезаурусе;

с) словосочетание является устойчивым;

д) отдельные слова имеют слишком широкое значение;

е) для данного слова в тезаурусе имеется общепринятая аббревиатура;

ф) разбиение словосочетания приводит к потере важных для поиска связей.

Например, словосочетание *радиационная обстановка* соответствует пунктам а) и с); *атомная электростанция* — е); *ядерно-радиационные производства, радиоактивно-загрязненные территории* б) и ф); *ионизирующее излучение 2* и т. д.

• Структуру, тематическую точность, специфику и эффективность поиска документов по радиационной экологии в общем массиве данных (в данном случае на основе количественного анализа). При альтернативном выборе ключевого слова следует ориентироваться на [6]:

а) полноту выражения значения;

б) соответствие стандартам и рекомендациям научно-технической литературы;

с) краткость и понятность;

д) при прочих равных условиях на наибольшую частоту использования в документах и запросах.

• Использование булевой логики при формировании поискового запроса при помощи ключевых слов списка.

В булевой алгебре логически ключевое словосочетание рассматривается как одно слово, то есть  $K1 \cap K2$  и при условии соседства в определенном порядке, поэтому ключевое словосочетание должно иметь терминологически устоявшийся (общепринятый) смысл.

Отдельные ключевые слова создают условия более гибкого, чем по словосочетанию поиска, в то же время ключевое словосочетание обеспечивает семантически точный (детальный) поиск информации с меньшими, чем при отдельных ключевых словах, нерелевантными запросами выдачи данных, например, поиск по словам *отходы* и *радиоактивность*, *радиоактивные отходы* или *радионуклиды*, *естественные радионуклиды* и др.

Поисковая операция  $I \cap$ :

<sup>1</sup> Автоматизированная информационная система научно-технической информации (БД АСНТИ) по геологии, природопользованию и недропользованию, № регистрации 2317 9 июня 1997 г., содержит материалы ВИЭМС, ВИНТИ, ВНИЦ (707 516 записей).

$K1 \cap K2 \cap K3$  — найденные документы должны содержать и  $K1$ , и  $K2$ , и  $K3$  (не обязательно рядом в тексте).

Поисковая операция ИЛИ  $\cup$ :

$K1 \cup K2 \cup K3$  — будут найдены документы, содержащие ключевые слова  $K1$ ,  $K2$ ,  $K3$ , (хотя бы одно слово).

• Параметры и особенности поисковой системы и поискового интерфейса системы.

Весь список ключевых слов касается общих и специфических для различных предметных областей (в том числе радиационной) слов и сочетаний. Однако следует отметить, что, за исключением общих терминов и словосочетаний, анализируемый массив документов можно подразделить на две составляющие: радиационную и экологическую (охрана окружающей среды).

Исходя из описанного выше, при дополнении списка ключевых слов баз данных ВИНТИ для поиска в массиве данных по экологии, охране окружающей среды и природопользованию, на основе анализа информационных документов по проектам в области радиационной экологии можно выделить следующие критерии для отбора слов или словосочетаний, которыми дополняется список.

1. Наличие документов радиэкологической тематики, содержащих слова-дескрипторы, не включенные в список, при отсутствии которых в системе релевантная информация по радиационной экологии не будет найдена (поиск возможен только при наличии данного слова в списке).

2. Наличие слов или сочетаний, являющихся терминами или устоявшимися сочетаниями, наиболее точно отражающими поисковый запрос по большому количеству документов. (Поиск возможен при использовании других слов списка, определяющих природоохранную и/или экологическую тематику, но результат по радиэкологической или радиационной тематике будет содержать много информационного “шума”).

3. Наличие слов или сочетаний, не включенных или частично включенных в список, уточняющих поисковый запрос по широкому тематическому профилю радиэкологических исследований. (Результаты поиска по имеющимся словам радиационной тематики содержат большой массив разноплановых документов, новая лексическая единица уточнит поиск).

Для окончательного внесения дополнительных ключевых слов радиэкологической тематики, отобранных по описанным выше критериям, в список слов по массиву данных по охране окружающей среды необходимы: проверка наличия каждого отдельного слова в имеющемся списке (для внесения словосочетаний); анализ поиска информации по сочетанию или дескриптору.

В специализированную информационную систему по радиэкологическим проектам, требующую более детального поиска, дополнительные ключевые слова могут вноситься в прямом порядке, поскольку система предполагает корректировку всего словаря в соответствии с заданной тематикой содержащейся в ней информации (согласно наличию слов в документах по радиэкологическим проектам).

Таким образом, дополнения для повышения эффективности поиска по радиэкологии будут вноситься, исходя из слов, определяющих радиационную тематику, с использованием данных по проектам с учетом тематического анализа, выделения дескрипторов и количественных характеристик поиска по словам, как в системе по радиэкологическим проектам, так и в Интернет.

## АНАЛИЗ ИСПОЛЬЗОВАНИЯ КЛЮЧЕВЫХ СЛОВ В МАССИВЕ ДАННЫХ ПО РАДИАЦИОННОЙ ЭКОЛОГИИ

Отбор слов для использования их в качестве ключевых может содержать в себе элементы всех трех критериев. В большинстве случаев ключевые слова для специализированной системы не подходят для использования в массиве данных по экологии и природопользованию. Ниже будут рассмотрены примеры таких слов, а чаще — сочетаний.

Рассмотрим пример. Поиск по ключевому слову *радионуклиды*, более всего отражающему радиэкологическую тематику и являющемуся вместе с использованием слов, отражающих компоненты и виды окружающей среды, основным признаком радиэкологической тематики и близких областей знания. Данное слово встречается в названиях более половины документов, а еще чаще в аннотациях, и может иметь различную семантическую нагрузку, а содержание запроса может отражать разнообразные материалы по исследованиям, проводимым в рамках проектов.

В радиэкологии термин *радионуклиды* чаще всего ассоциируется с исследованиями по механизму миграции (переносу, накоплению, распределению, трансформации) радионуклидов в биоте, экосистеме, почвенном покрове, геологической среде и т. п.

Результатами поиска по запросу “*радионуклиды и миграция*” (эти слова есть в списке) будут, например, такие проекты: **Миграция искусственных радионуклидов и тяжелых металлов по пищевым цепям экосистемы<sup>2</sup> Азовского моря** [4], (проект № 6.57.3)<sup>3</sup>. Или, допустим, результатом по запросу “*радионуклиды и миграция и моделирование*” будет проект: **Разработка аналитических и численных моделей подземной миграции радионуклидов от техногенных источников загрязнения** [4] (проект № 7.36.5). Можно сузить поиск благодаря добавлению в поисковое предписание других слов.

Выбор экологической информации по таким ключевым словам списка, как *экосистема, окружающая среда, геологическая среда, биота*, может содержать информацию по исследованию миграции радионуклидов в этих средах вместе с исследованиями других направлений природоохранной тематики. В то же время тематика миграции радионуклидов включает вопросы анализа экосистем, оценки состояния окружающей и геологической среды и биоты.

<sup>2</sup>В примерах названий документов слова, имеющиеся в списке ключевых слов баз данных ВИНТИ выделены жирным шрифтом и вместе со словом-примером составляют поисковый образ документа.

<sup>3</sup>В скобках обозначен номер проекта, под которым он значится в указанном источнике.

Результаты поиска по поисковому запросу “радионуклиды или окружающая среда”, кроме сведений по миграции радионуклидов, могут отражать различные аспекты радиоэкологических исследований, например:

Общие исследования: *Радионуклиды в среде и биоте Западной Арктики* [4], (проект № 7.33.8)

Исследования по консервации радионуклидов в геологических формациях, радионуклидному загрязнению. Например, *Теоретические основы и принципы геоконсервации радионуклидов по моделям формирования месторождений урановых руд на хемосорбционных геохимических барьерах с использованием катионообменных бентонитовых глин*, [4], (проект № 7.26.8).

Радионуклидное загрязнение: *Анализ полей радионуклидного загрязнения в различных ландшафтах с использованием геоинформационной технологии* [3] (проект № 9.31.1)

В названиях проектов предполагаемой информационной системы по радиоэкологическим исследованиям, слово *радионуклиды* встречается 66 раз (Табл. 2, графа 2, номер слова 20), четверть из которых связана с поведением радионуклидов в разных средах в сочетаниях со словами: *миграция, поведение, перемещение, а также модель переноса, моделирование переноса и распределения, моделирование потока, механизм переноса и трансформации, распределение*, которые также можно отнести к миграции. Миграция — это перемещение, передвижение, перераспределение (от лат. *migrō* — перемещаю, переселяюсь). Более близкое к миграции радионуклидов словосочетание *миграция элементов* (относящееся к геохимии), означает перемещение и перераспределение химических элементов в земной коре и на её поверхности (термин введён А. Е. Ферсманом в 1923 г.). Однако и слово *миграция* и слово *радионуклиды* имеются в списке ключевых слов, к тому же с ними могут быть сформированы другие сочетания. В поисковых системах Интернет более целесообразно использовать именно отдельные слова. Но, если поисковый интерфейс системы и предписание позволяют вводить только два слова в рамках поиска по ключевым словам (как в системе “ЭкоПро”), то введение в качестве ключевого словосочетания *миграция радионуклидов*, (хотя оба слова имеются в списке), увеличит возможности более детального поиска радиоэкологических данных по данному тематическому разделу по природным средам с использованием математической логики Дж. Буля.

В большей части проектов, входящих в создаваемую радиоэкологическую информационную систему, а также в научных публикациях по радиоэкологической тематике, отобранных из баз данных ВИЭМС, кроме непосредственно слова *радиационная экология*, использующегося в общем смысле (допустим, *Радиационная экология Сибири*), приводятся и такие ключевые словосочетания, как *радиоэкологическая обстановка, радиационная безопасность* и др. Имеют место и такие словосочетания, как *радиоэкологический (радиационный) мониторинг, радиоэкологическое состояние, радиоэкологическая ситуация, радиоэкологический (радиационный) риск, радиоэкологические карты*

(картирование). Например: *Моделирование возможных экологических последствий от объектов радиационного риска в Европейской Арктике* [1] или *Разработка методики районирования и составление радиоэкологической карты России масштаба 1:2 500 000* [1].

Все приведенные словосочетания могут формироваться с помощью булевой логики с использованием полного списка ключевых слов. Особое место занимает слово *мониторинг*. Такие словосочетания, как *радиоэкологические карты* (в названии одного проекта системы) и *радиационный риск* (в названии одного проекта) встречаются относительно редко. Слово *мониторинг* в названиях проектов формируемой информационной системы встречается 13 раз (5%), в восьми из которых применяется устоявшееся словосочетание *радиоэкологический мониторинг* (может формироваться с использованием имеющихся в списке слов: *радиоэкология* и *мониторинг*). Например, в статьях базы данных ВИЭМС около 12% публикаций, посвященных радиоэкологическому мониторингу.

Предлагается отдельно выделить в качестве ключевого слова и внести в список общепринятое и устоявшееся словосочетание *радиационный мониторинг*, поскольку в настоящее время данные исследования приобретают особую важность в связи с созданием единого экологического мониторинга. В поисковых системах Интернет результатов поиска по *радиационному мониторингу* больше, чем по радиоэкологическому, и в три раза больше количества документов о мониторинге в массиве поиска по словам *радиоэкология, радиоактивность*. Предположительно, количество данных и поисков в этой предметной области будет расти в связи с развитием комплексного экологического мониторинга.

Для информационно-поисковой системы по радиоэкологическим проектам достаточно слова *мониторинг*, которое, как и в случае со словом *радионуклиды*, отражает большой объем исследований по радиоэкологическому и радиационному мониторингу территорий, почвенного покрова, наземных вод, экосистем или окружающей среды в целом по региону.

Примеры по словосочетаниям:

**радиоэкологический мониторинг:**

**региона**, *Обобщение и анализ экспериментальной информации по материалам радиоэкологического мониторинга Московского региона* [7];

**экосистемы реки**, *Радиоэкологический мониторинг экосистемы реки Иртыш в районе Тобольской биостанции ИПЭЭ РАН Радиоэкологический мониторинг наземных и водных экосистем в Брянской и Челябинской областях* [3], (проект № 7.25.25),

*Радиоэкологический и биохимический мониторинг наземных и водных экосистем в Сибири и Европе* [4], (проект № 6.40.17);

**прибрежных вод**, *Радиоэкологический мониторинг Мурманского мелководья и побережья Баренцева моря* [4], (проект № 6.57.4);

**флоры и фауны**. *Радиоэкологический мониторинг объектов флоры и фауны на территории Российской Федерации, загрязненной радионуклидами в результате аварии на ЧАЭС* [3], (проект № 7.25.24);

**радиационный мониторинг:**

*Разработка принципов и методов радиационного мониторинга почвенного покрова Дагестана* [3], (проект № 7.36.10);

*Научно-методическое сопровождение экспедиционных и аналитических работ, проводимых в рамках радиационного мониторинга территорий, загрязненных в результате Чернобыльской и других радиационных аварий, а также ядерных взрывов, проведенных ранее* [4], (проект № 7.38.16);

*Развитие системы радиационного мониторинга окружающей среды Баренцрегиона*  
**мониторинг — хранилища радиоактивных отходов:**

*Мониторинг на объектах хранения РАО, отходов промышленных предприятий и эксплуатации атомно-энергетических комплексов* [4], (проект № 7.35.7);

*Основные закономерности распределения радиоактивных элементов в почвенно-элювиальных покровах типовых ландшафтов юга Западной Сибири и радиогеохимический мониторинг хвостозащитных жидких и твердых отходов сибирских предприятий Минатома России (Новосибирский завод химического концентратов, Сибирский химический комбинат, Красноярский электрохимический завод, Ангарский электролизно-химический комбинат)*, [4], (проект № 7.26.3).

В приведенных и в последующих примерах можно проследить использование слов в сочетаниях, как со словом *радиоэкология* (радиоэкологический), имеющемуся в списке (Табл. 2, номер слова: 21), так и со словом *радиация* (радиационный). Например, *радиоэкологический риск* и *радиационный риск* и др.

Слово *радиация* отражено в списке ключевых слов только в виде ключевых словосочетаний *радиационная безопасность*, *радиационная обстановка* и др. (Табл. 2, номера слов: 9–12). В связи с этим возникает вопрос о введении ключевого слова *радиация*. Однако результаты поиска отдельно по данному слову, предположительно, могут отражать большой объем исследований (Рисунок, номер слова: 26д) по солнечной, ультрафиолетовой, атмосферной, ионизирующей радиации, радиационному контролю, а также дозам радиации, воздействию радиации на организм, радиационному фону, детекторам радиации, радиационной дозиметрии. Поэтому дополнение списка отдельными словосочетаниями представляется более целесообразным, хотя, возможно, выделение слова отдельно (с примечанием в связи с широким диапазоном его использования во всевозможных, иногда редко встречающихся и не вполне устоявшихся сочетаниях путем булевой логики (*радиационный климат* и пр.)).

Почти все названия проектов радиоэкологической тематики, в которых отсутствуют слова списка, содержат сочетания со словом *радиация* (Табл. 3).

Подобные исследования, в названиях которых есть данные сочетания (см. Табл. 3, номера слов: 1–10, 19–23), направлены на радиационный контроль, а также изучение воздействия радиации на организм человека, животных, экосистемы, и более касаются радиобиологии, допустим: *Комплексный анализ адаптационных реакций организма*

*человека в условиях хронического радиационного воздействия* [3], (проект № 3.16.8) (радиационное воздействие, иначе — *радиация* и *воздействие*). Часть таких исследований можно найти при помощи ключевого слова *радиационная безопасность*, в понятие которого входит комплекс мер и мероприятий при работе с применением радиоактивных веществ и других источников ионизирующих излучений, обеспечивающий снижение суммарной дозы от всех видов ионизирующего излучения до предельно допустимой дозы. Для поиска подобной информации, кроме возможностей использования булевой логики из комбинаций слов, представленных в списке, (например, *радиоактивность*, *естественные радионуклиды*, *ионизирующее излучение* и *воздействие* и *контроль* и *организм человека*, и *растения*) необходимо ввести такие слова, как *детектор излучения*, *дозы облучения*, *радиационная дозиметрия*, *дозы радиации* и др. (Табл. 3.) использующиеся в названиях проектов (без применения других слов списка). Например:

*дозы радиации*, *Биологическое действие малых доз радиации* [4], (№ проекта 2.16.1);

*радиационная дозиметрия*, *Радиационная дозиметрия Нижненовгородской области* (БД МНТЦ <http://www.tech-db.ru/istc/db/projects.nsf/>, проект № 0824);

*радиоактивный фон*, *радиационный фон*, *Природный радиоактивный фон. Его значение для биосферы Земли. Роль вторичного биогеогенного излучения* [4], (проект № 6.23.17), *Вклад β-излучения в дозу облучения растений в зоне с повышенным естественным радиационным фоном* [10];  
*уф-радиация*, *Разработка моделей и анализа процессов, связанных с экологической безопасностью, в том числе разработка программы автоматизированного расчета ультрафиолетовой радиации системы атмосфера—океан на суперЭВМ* [3], (проект № 13.11.8);

*детектор излучения неохлаждаемые детекторы ионизирующего излучения на базе диоксида ртути*, [3], (проект № 10.18.2);

В списке ключевых слов нет слов *дозы*, *защита*, *фон*, *облучение*, *ультрафиолетовое (уф) излучение* для формирования подобных словосочетаний. К тому же некоторые из приведенных словосочетаний относятся к устойчивым понятиям. Например, *радиоактивный фон* (естественный радиационный фон), — ионизирующие излучения, источниками которых являются космические лучи и естественно распределённые в природе радионуклиды.

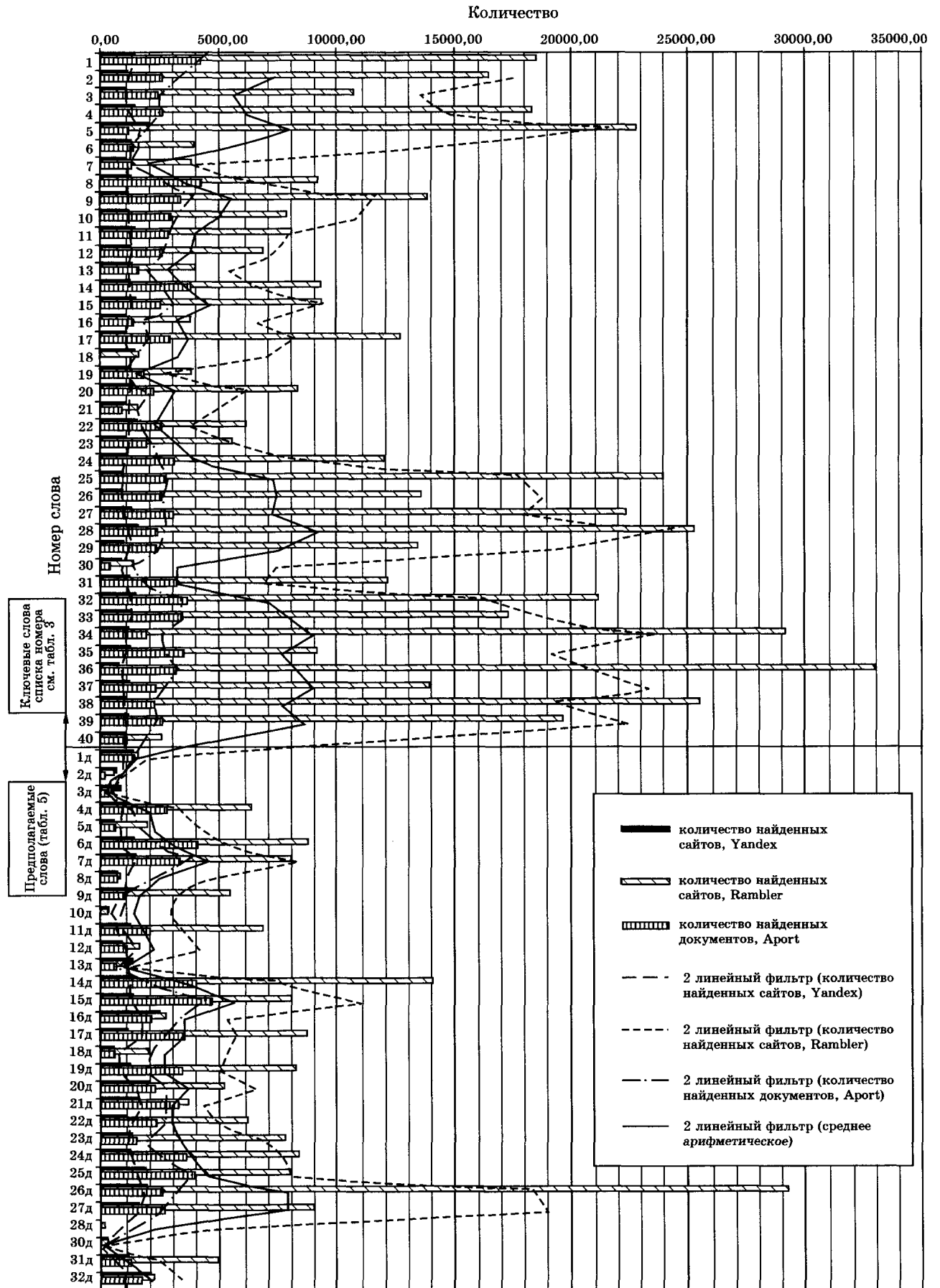
В названиях проектов формируемой информационной системы данные словосочетания встречаются достаточно мало (один — три документа на каждое словосочетание) и поиск по ним будет более точным.

Из всех рассмотренных слов (Табл. 3, номера слов: 1–10, 19–23) рекомендуется включить в список ключевых следующие наиболее подходящие согласно результатам поиска.

Пример поиска слов на рисунке — номера слов 13д–26д, 31д.

• *Детекторы радиации*, а для списка ключевых слов баз данных ВИНТИ более целесообразно введение слова *детекторы*, так как поиск с использованием булевой логики будет более эффективным, (например, результат поиска в Интернет с использованием словосочетания меньше, чем по отдельным словам).

Диаграмма количественного анализа ключевых слов списка по поисковым системам Internet с учетом добавленных слов



**Дескрипторы в названиях радиоэкологических проектов,  
не использующих ключевые слова списка\***

№	дескрипторы	Источники:					% к общ. (распреде- ление слов в 11,67% проектов)	% к общему количеству проектов
		РАН	ЭкоПро	МНТЦ	Internet	Итого:		
1	атмосферная радиация	1		1		2	5,71	0,63
2	BAO	1		1		2	5,71	0,63
3	воздействие малых доз радиации	2				2	5,71	0,63
4	воздействие радиации, радиационное воздействие	1		2		3	5,71	0,63
5	воздействие радиационных факторов	1				1	2,86	0,32
6	детекторы радиации			1		1	2,86	0,32
7	дозы радиации	1		1		2	5,71	0,63
8	малые дозы радиации	1				1	2,86	0,32
11	радиационные аварии	2				2	2,86	0,32
12	радиационные инциденты	1				1	5,71	0,63
13	радиационные катастрофы	2				2	2,86	0,32
14	радиационные поражения			1		1	5,71	0,63
15	радиационные последствия	1				1	2,86	0,32
16	радиационные ситуации	1				1	2,86	0,32
17	радиационный мониторинг	1		2	1	4	2,86	0,32
18	радиационный риск	1	1,0	1		3	11,43	1,26
19	радиация	1				1	8,57	0,95
20	радиогеохимия	1				1	2,86	0,32
21	техногенное облучение			1		1	2,86	0,32
22	ультрафиолетовая радиация	1				1	2,86	0,32
23	фон радиации, радиационный фон	1		1		2	2,86	0,32
24	Итого слов:	21	1,0	12	12	35	100,00	11,04
25	количество проектов (см. табл. 1 п. 4)	21	1,0	12	1	35	100,00	11,04

\*Жирным выделены слова, имеющиеся в списке ключевых слов ВИНТИ

• *Радиоактивный фон* — более используемое словосочетание, чем *радиационный фон* (Табл. 4, номера слов: 24д, 27д), являющееся термином.

• *Дозы радиации* (Табл. 4, номер слова: 7д) — наиболее тематически точно подходящая альтернатива введения слова *радиация*, поскольку возможно использование булевой логики в сочетаниях с такими словами, как *воздействие*, *человек* и др.

Работы радиоэкологической тематики, с предупреждением и ликвидацией последствий аварий, чаще всего посвящены аварии на Чернобыльской атомной станции (ЧАЭС), поиск которых возможен с использованием ключевых слов, имеющих в списке: ЧАЭС и других слов (Табл. 2. номера слов: 23, 24, 31). Слово ЧАЭС фигурирует в названиях семи проектов и в большей половине аннотаций проектов, а также публикаций. Однако, кроме Чернобыльской катастрофы, имеются исследования по другим авариям и радиационным инцидентам (например, на производственном объединении «Маяк» в 1957 г.). Из всей информации по авариям в поисковых системах Интернет, на Чернобыльскую атомную электростанцию приходится около 20%, на ПО «Маяк» — 14% информационных документов. В названиях проектов такие исследования чаще отражены словами:

**радиационные катастрофы**, *Радиационные катастрофы: историко-методологический анализ причин возникновения, развития и ликвидации последствий* [3], (проект № 14.4.7).

**радиационные аварии**, *Разработка информационной системы оперативного реагирования на радиационные аварии для Кольского по-*

*лустровая стран Северного Калотта* [3], (№ проекта 9.45.23).

Для части проектов данные словосочетания являются определяющими (Табл. 3, номера слов: 11, 12, 13). По результатам поиска в Интернет, все найденные документы по словам *аварии* и *катастрофы*, отражают около 1/5 от радиационной тематики (по результатам сочетаний с различными словами списка: *радиоактивность*, *радиология*, *атомные электростанции* и др.). Хотя в списке ключевых слов баз данных ВИНТИ имеются слова *аварии*, *катастрофы*, ввиду важности и многоаспектности проблемы, предлагается их выделить отдельно.

Обширный круг вопросов охраны окружающей среды и радиоэкологии связан с транспортировкой, обезвреживанием, утилизацией и захоронением твердых и жидких радиоактивных отходов: исследования захоронения отходов ядерной энергетики, утилизации атомных подводных лодок (АПЛ) и др. Хотя данный тематический раздел включает много аспектов (долговременное и кратковременное захоронение, виды отходов, хранение, обезвреживание, транспортировка, утилизация: методы, способы, материалы), для поиска в объединенном массиве экологических и природоохранных данных в списке достаточно ключевых слов по данной тематике (Табл. 2, номера слов: 7, 16, 17, 29). Вместе с другими словами списка (чаще всего: *поллигоны*, *подземные хранилища*, *хвостозранилища*, *контейнер*, *бетон*), эти ключевые слова, согласно документам по радиоэкологическим проектам,



Список добавляемых ключевых слов радиационной тематики

№ слова п/п	Слова в алфавитном порядке	Среднее арифметическое по поисковым системам Internet	Добавлять в список ключевых слов		Примечание
			Радиоэкологические проекты	Список ключевых слов ВИНИТИ	
1д	атмосферная радиация	1362,33	✓	✓	Дескр., Тер,
2д	ВАО	356,67	✓	✓	Дескр., Тер., Слит. (аббрев.)
3д	высокорadioактивные отходы	385,00	✓		Тер, Дескр., Кол<
4д	детектор излучения	3368,00		✓	Дескр.,
5д	детектор радиации	918,00	✓	✓	Дескр.,
6д	дозы облучения	4650,33			Дескр., Ш (медицина)
7д	дозы радиации	4192,33	✓	✓	Дескр. Тер, Слит.
8д	искусственные радионуклиды	573,33	✓		Тер, Кол<
9д	малые дозы радиации	2568,67	✓		Тер, Дескр., Кол<
10д	низкорadioактивные отходы	132,67	✓		Тер, Кол<
11д	облучение населения	3272,00	✓		Дескр., Ш
12д	оружейный плутоний	1100,00	✓		Тер., Кол<
13д	радиационная дозиметрия	879,00	✓		Тер., Дескр.,
14д	радиационная защита	6396,67	✓		Дескр., Ш
15д	радиационное воздействие	4554,67	✓		Дескр.,
16д	радиационное поражение	2322,00	✓		Дескр.,
17д	радиационные аварии	4350,00	✓		Дескр., Кол<
18д	радиационные инциденты	928,33			Дескр., Кол<
19д	радиационные катастрофы	4195,67			Дескр.,
20д	радиационные ситуации	3058,00	✓		Дескр., Ш
21д	радиационные факторы	2720,67	✓		Тер, Дескр., Ш
22д	радиационный мониторинг	3098,00	✓	✓	Тер, Дескр.,
23д	радиационный риск	3411,67	✓		Тер, Дескр.,
24д	радиационный фон	4320,33			Тер, Дескр., Кол>
25д	радиационное воздействие	4526,00	✓		Тер, Дескр.,
26д	радиация	11166,00	✓		Тер, Кол>
27д	радиоактивный фон	4319,00	✓	✓	Дескр., Тер, Слит.
28д	радиогеоэкология	21,67			
29д	среднерadioактивные отходы	65,00	✓		Тер, Кол<
30д	твердые радиоактивные отходы	2358,00	✓		Тер, Кол<
31д	ультрафиолетовая радиация	1884,67	✓		Тер, Дескр., Кол<
	<b>Всего:</b>		<b>27</b>	<b>7</b>	

Примечание: Дескр. — слово является дескриптором в системе по радиоэкологическим проектам; Тер/ — слово является термином; Слит. — только в виде словосочетания компоненты не употребляются отдельно); Ш — поиск будет содержать много шума; Кол< — недостаточный объем поиска (< среднее — отклонение); Кол> — поисковый запрос отражает “много шума” (>> среднее + отклонение, неточность тематического профиля). Рекомендуемый интервал (среднее арифметическое, среднее взвешенное +- отклонение)

так или иначе присутствуют в названиях проектов рассматриваемой области. Для более целенаправленного поиска таких разноплановых исследований наиболее характерными словами являются: *высокорadioактивные отходы, жидкие радиоактивные отходы, твердые радиоактивные отходы, оружейный плутоний* (виды отходов см. Табл. 4, номера слов: 2д, 3д, 10д, 29д, 30д), *отработанное ядерное топливо* (практика показывает, что в качестве ключевого слова удобнее использовать просто *ядерное топливо*), а также *керамические матрицы, стеклование, полимерные материалы, изоляционные свойства пород* (которые также необходимо добавлять в систему по радиоэкологическим проектам по мере наполнения базы данных) и др.

Приведем примеры проектов:

**твердые радиоактивные отходы (ТРО), Реконструкция пункта захоронения твердых ра-**

**диоактивных отходов спецкомбината “Радон” [1]; высокорadioактивные отходы (ВАО), облученное ядерное топливо (ОЯТ), Оценка изменений изоляционных свойств скальных массивов при захоронении ВАО: проницаемости, напряженно-деформированного состояния, процессов фильтрации и массопереноса, физико-химических процессов, Геология окружающей среды (экологическая геология): эколого-геохимические исследования утилизации и захоронения ВА и ТО [3], (проект № 4.62.1), Миграция и концентрирование естественных и искусственных радионуклидов в условиях природных и техногенных ландшафтов. Геологические основы длительного хранения и подземного захоронения высокорadioактивных отходов и облученного ядерного топлива** (лаборатория радиогеологии и радиогеоэкологии ИГЕМ РАН).

Список ключевых слов баз данных ВИНТИ рекомендуется пополнить по крайней мере словосочетанием *высокорadioактивные отходы* или аббревиатурой *ВАО*, которые отражают наиболее важные проблемы размещения опасных радиоактивных отходов в подземном пространстве. Поиск по этим словосочетаниям будет более эффективным, чем по имеющимся в списке словам.

Последние примеры названий проектов, затрагивают геологические аспекты захоронения радиоактивных отходов, которые могут быть также связаны с радиационным контролем и изучением миграции радионуклидов в местах подземного захоронения. Геологические аспекты достаточно часто встречаются в радиоэкологических исследованиях разных направлений. В списке ключевых слов есть слова: *геоэкология*, *геохимия*, а также *геологическая среда*, *подземное пространство* и другие, которые в сочетаниях с помощью булевой логики со словами *радиология*, *радионуклиды*, *радиоактивные отходы* и т. д., могут использоваться для формирования запроса документов подобной тематики.

Для геологических аспектов впоследствии можно вводить такое ключевое слово, как *радиогеоэкология*, обозначающее целую отдельную тематическую область исследований, например:

**Радиогеоэкология, Радио-геоэкологические исследования таежно-мерзлотных ландшафтов Хиагдинского ураново-рудного поля** [4], (проект № 7.08.6).

Однако из-за малочисленности информации, приходящейся на данное слово (см. рисунок, номер слова: 28д), и возможности замены его другими дескрипторами, для списка ключевых слов баз данных ВИНТИ *радиогеоэкология*, в настоящий момент не подходит. Это слово войдет в систему по радиоэкологическим проектам в связи с важностью данного направления для исследований ИГЕМ.

Рассматриваемый список ключевых слов баз данных ВИНТИ содержит достаточно большое количество слов, характеризующих окружающую среду. Однако в информационно-поисковую систему по проектам в области радиоэкологии целесообразно также добавить слова из списка ключевых слов баз данных по геологии в части характеристики геологической среды, часто имеющей место в названиях радиогеоэкологических проектов, выполняемых ИГЕМ (например, *глины*, *скальные массивы* и пр.).

Все рассмотренные примеры можно отобразить в виде общей таблицы ключевых слов (Табл. 4), предлагаемых для расширения возможностей поиска информации в области радиоэкологии.

## ЗАКЛЮЧЕНИЕ

Тематический и логический анализ показал, что результаты поиска по словам в специализированной системе по радиоэкологическим проектам

и в больших массивах данных распределяются неравномерно со значительными структурными отличиями (Рис. 2). При дополнении списка ключевых слов, которое во многом определяется запросами пользователей, важное значение имеют не только терминологическая точность нового ключевого слова или сочетания, но и информационный поиск по нему в больших массивах данных.

Распределение результатов поиска по ключевым словам в базе данных по радиоэкологическим проектам отличен от данных поиска по большим информационно-поисковым системам (см. Табл. 2, графы 2–4 и 5–12, и приложение к Табл. 2), в которых наибольшее количество документов приходится на слова: *ядерные проблемы*, *ядерные материалы*, *ядерные технологии*, относящиеся более к тематике ядерной энергетики; из неядерного направления — *ионизирующее излучение*, *радиоактивные загрязнения*. Несмотря на различие в количестве документов (сайтов), выдаваемых поисковыми системами Интернет по одному и тому же слову (наибольшее значение у Rambler), различия в структуре поиска (доля выдаваемых документов по каждому слову относительно общего массива поиска), незначительны (как правило, первые документы, выдаваемые системами по запросу, одинаковы). Поэтому по данным Интернета более точно можно судить о характере поиска по тому или иному слову, нежели по информации о радиоэкологических проектах.

В радиоэкологических проектах, которые на данный момент входят в систему, доминирует тематика радиоактивных отходов, разноплановые изучения миграции и концентрации радионуклидов в различных средах. Это объясняется специализацией базы данных. Ключевые слова списка радиологической направленности встречаются в большинстве названий проектов радиоэкологической тематики 98% (Табл. 1, графа 6). В 40% названий проектов — по два, реже три, ключевых слова списка радиационной тематики.

С целью увеличения точности поиска по ключевым словам и более детального терминологического обеспечения информационной системы по радиоэкологическим проектам, список предлагается пополнить 27 словами (Табл. 4). Помимо добавляемых слов, явно радиационной направленности, по мере пополнения базы данных следует добавить дескрипторы, наиболее часто встречающиеся в названиях проектов по захоронению радиоактивных отходов.

Решение о добавлении ключевого слова в перечень сформировано (Табл. 4 графы: 3–6) на основании того, является ли это слово необходимым дескриптором, насколько внесение слова уточнит релевантность запроса, наличия в списке составляющих (для сочетаний), количественного критерия (по поиску в Интернете).

Таким образом, для включения в список ключевых слов баз данных ВИНТИ функционирования Государственной системы научно-технической информации, предлагается семь слов, отражающих

радиационную тематику. В результате этого предполагается повысить эффективность нахождения документов, соответствующих заданной тематике, в больших массивах данных по охране окружающей среды и природопользованию, в режиме поиска по ключевым словам.

#### СПИСОК ЛИТЕРАТУРЫ

1. Веселовский А. В., Микляев В. И., Плате А. Н. Экологические проекты в России.— М.: ИГЕМ РАН, 2000.— 396 с.

2. Кузьмина Д. А. Автоматизированная информационная система для оценки проектов в области радиоэкологии: Тез. докл. Годичной сессии Московского Отделения МО 2003.— М.: ИГЕМ РАН, 2003.— С. 67.

3. Хильчевская Р. И., Остроумова Н. К., Федорова Е. М., Белоусова А. П. Проблемы экологии: Тематический справочник РАН. Вып. 1.— М.: ИГЕМ РАН, 2000.— 425 с.

4. Хильчевская Р. И., Остроумова Н. К., Федорова Е. М., Белоусова А. П. Проблемы экологии: Тематический справочник РАН. Вып. 2.— М.: Рохос, 2003.— 704 с.

5. Чумакова Н. Ф., Потапов И. И., Лосев К. С. Список основных ключевых слов для координатного индексирования документов по проблемам охраны окружающей среды и экологии / НТП ВИНТИ.— М., 2004.— 32 с.

6. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления. Межгосударственный стандарт, ГОСТ 7.25-2001. ИПК.— М.: Изд-во стандартов, 2001.

7. Вербов В. В., Ивлиев М. В. Обобщение и анализ экспериментальной информации по материалам радиоэкологического мониторинга Московского региона за период 1993–1995 гг. // Тр. МосНПО “Радон”.— 1998.— № 5.— С. 91–92.

*Материал поступил в редакцию 16.02.05*