



## О перспективах использования самого крупного в России массива научной информации – базы данных ВИНТИ РАН

Ю. Н. Щуко

*Всероссийский институт научной и технической информации РАН, г. Москва, Россия*

**Резюме:** Политематическая реферативная база данных (БД) – основной информационный продукт ВИНТИ РАН – формируется в результате аналитико-синтетической переработки входного потока отечественной и зарубежной научно-технической литературы, интеграции и систематизации разнородных информационных ресурсов. БД ВИНТИ РАН характеризуют политематичность, подокументная индексация, общая структура элементов описания разных видов документов как единого массива данных. Являясь аналитическим информационным ресурсом, БД представляет перспективы в решении ряда задач, в том числе анализ структуры отечественной и мировой науки, выявление наиболее актуальных или теряющих свою актуальность научных направлений, отслеживание генезиса конкретных научных идей и истории их развития, определение продуктивности работы исследователей, анализ структуры научного сообщества. Важнейшим приоритетом в деятельности ВИНТИ является подготовка материалов прогнозно-аналитического и обзорного характера, в том числе на основе своей БД.

**Ключевые слова:** реферативная база данных; аналитико-синтетическая переработка; политематичность; классификатор; рубрикатор; ГРНТИ; система взаимосвязанных классификаторов; точка роста; мониторинг

**Для цитирования:** Щуко Ю. Н. О перспективах использования самого крупного в России массива научной информации – базы данных ВИНТИ РАН // Научное издание международного уровня – 2019: стратегия и тактика управления и развития: материалы 8-й Международ. науч.-практ. конф., Москва, 23–26 апреля 2019 г. Екатеринбург: Изд-во Урал. ун-та, 2019. С. 127–130. DOI: 10.24069/konf-23-26-04-2019.19.

## On the prospects for using the largest array of scientific information in Russia – the VINITI RAS database

Yu. N. Shuko

*Russian Institute for Scientific and Technical Information, Moscow, Russia*

**Abstract:** The main information product of VINITI RAS is a polythematic database (DB), which is formed as a result of analytical and synthetic processing of the input stream of domestic and foreign scientific and technical literature, integration and systematization of heterogeneous information resources. The features of the DB of the VINITI RAS are: wide thematic coverage, document indexation, the general structure of describing different types of documents as a single data set. As well as being an analytical information resource, the database has prospects in solving a number of tasks, including analyzing the structure of domestic and world science, identifying the most relevant or losing their relevance scientific areas, tracking the genesis of specific scientific ideas and their development history, determining the productivity of researchers, analyzing structures of the scientific community, etc. The most important priority in the VINITI activity is the preparation of forecast-analytical reviews, including on the basis of its database.

**Keywords:** database; analytical and synthetic processing; wide thematic coverage; classifier; rubricator; GRNTI; system of interrelated classifiers; growth point; monitoring

**For citation:** Shuko Yu. N. On the prospects for using the largest array of scientific information in Russia – the VINITI RAS database. In: *World-Class Scientific Publication – 2019: Strategy and Tactics of Management and Development: Proc. 8<sup>th</sup> Int. Sci. & Pract. Conf., Moscow, April 23–26, 2019*. Ekaterinburg: Ural University Press, 2019, pp. 127–130. DOI: 10.24069/konf-23-26-04-2019.19.

Основным информационным продуктом ВИНТИ РАН является политематическая реферативная БД, которая включает 28 тематических фрагментов по различным отраслям знаний, 217 проблемно-ориентированных разделов БД, в том числе по приоритетным направлениям науки, технологий и техники. БД ВИНТИ генерируется с 1981 г., содержит свыше 36 млн аналитических записей (метаданные и рефераты) с ежегодным пополнением 700 тыс. документов.

БД ВИНТИ формируется на основе результатов аналитико-синтетической переработки входного потока отечественной и зарубежной научно-технической литературы (НТЛ), интеграции разнородных информационных ресурсов, тематической и предметной систематизации НТЛ. Входной поток включает около 7000 периодических и сериальных изданий и 12000 изданий книжного типа ежегодно, а также российские патентные документы из БД ФИПС и зарубежные патенты из БД Questel, труды конференций, депонированные работы, монографии, стандарты и нормативы, справочники, терминологические словари, авторефераты диссертаций.

Входной поток Scopus – свыше 23 тыс. наименований периодических изданий. Входной поток WoS – свыше 12 тыс. наименований.

По своим функциональным и технологическим возможностям, а также по спектру представляемой информации БД ВИНТИ РАН сопоставима с такими крупными банками научно-технической информации как Scopus, Web of Science, Dialog и др. Отличительной особенностью БД ВИНТИ РАН являются ее политематичность (широкий охват отраслей знаний), подокументная индексация (кодами классификационных систем ГРНТИ, УДК, рубрикатора ВИНТИ, ключевыми словами), а также общая структура элементов описаний для разных видов документов (не только для журнальных статей), обеспечивающая представление всех фрагментов БД и всех видов документов (от статей до монографий) как единого массива данных – всего свыше 150 полей в зависимости от вида документа.

Политематическая БД ВИНТИ РАН по объему, видам и странам издания первоисточников, форме представления реферативно-аналитической информации, предметной и классификационной индексации документов является глобальным информационно-аналитическим ресурсом по естественным, техническим и точным наукам.

Широко распространено мнение, что в эпоху Интернета в нем можно найти всю необходимую информацию, что агрегаторы научной и технической информации больше не нужны. В Интернете почти все поисковые машины (вслед за Гуглом) выдают сайты с запрошенной информацией по частоте посещения сайтов. Другими словами, Интернет в первых десятках выдает сайты, которые ищутся всеми пользователями. Поскольку ученые ищут, как правило, совсем не то, что массовый читатель, нужная ему информация может выдаваться в конце многотысячного перечня сайтов. Для иллюстрации высказанных положений мы провели простой опыт. В несколько поисковых машин (*Google*, Яндекс, Рамблер, *Yahoo!*, *Bing*) был введен неспецифический запрос: «Последние достижения нанотехнологий в органической химии». Среди просмотренных страниц (2–3 первых десятков сайтов) встречались материалы пресс-релизов, статьи из популярных журналов, материалы выставок, сообщения рекламного характера.

Для поиска в политематической базе данных ВИНТИ запрос был преобразован следующим образом: «органическая химия\$нанотехнологии\$достижения». В качестве обязательных ограничений были выбраны: страна публикации, язык оригинала статьи, хронологические рамки и вид первоисточника и тематика БД (химия). Было получено около 10 тыс. библиографических описаний с рефератами на русском языке, ранжированных по годам опубликования статей, которые можно отсортировать по названиям журналов. Ясно, что такой поиск более комфортен и эффективен для научного работника, чем бессистемный просмотр сайтов Интернета.

Тематические и проблемно-ориентированные БД, формируемые на основе БД ВИНТИ, выполняют информационно-поисковые и науковедческие функции, нейтрализуя действие закона рассеяния статей определенной тематики (закона Брэдфорда) по всему массиву научных изданий. Ведь общеизвестно, что до 70% всей информации ученый получает из журнальных статей. По закону рассеяния Брэдфорда из журналов, публикующих статьи по определенной теме, можно получить лишь треть от всех статей по этой теме. Остальные рассеяны по смежным и многоотраслевым журналам.

Цифровые БД, каковыми являются документально-реферативные базы данных ВИНТИ, собирают эти рассеянные статьи по узким темам-рубрикам под интересы конкретных пользователей, для которых это единственная возможность обеспечить охват и полноту поиска информации. Базы данных институтов научной информации избавляют ученого от потери времени на поиски в библиотеках, поскольку их производители берут на себя ответственность за отбор релевантных статей из фиксированного списка журналов.

В БД содержатся рефераты и авторские резюме каждой статьи на языке пользователя. Разумеется, большинство высокоцитируемых в мире журналов выходит на английском языке, а хороший ученый должен им владеть. Однако если мы хотим иметь российскую науку, то в ней должна быть своя терминология и свое видение любой актуальной проблемы. Это и дают рефераты на русском языке и классификаторы-рубрикаторы БД, составляемые квалифицированными отечественными специалистами.

Поскольку речь идет об аналитическом информационном ресурсе столь значительного объема, использование его для дальнейшей постобработки методами интеллектуального анализа данных и наукометрическими методами представляется весьма перспективным в решении ряда задач, в числе которых:

- анализ структуры отечественной и мировой науки;
- определение тенденций и процессов, происходящих в мировой и региональной науке;
- выявление наиболее актуальных или, напротив, теряющих свою актуальность научных направлений;
- отслеживание генезиса конкретных научных идей и истории их развития;
- определение продуктивности работы исследователей и эффективности материальных затрат в конкретной научной области;
- анализ структуры научного сообщества и изучение науки как социального организма.

В свою очередь внедрение в процесс аналитико-синтетической переработки первоисточников научно-технической информации (НТИ) и генерации БД ВИНТИ, разработанной Системы взаимосвязанных классификаций НТИ, ядром которой является Государственный рубрикатор научно-технической информации (ГРНТИ), позволит серьезнейшим образом подойти к решению проблемы создания онтологии единой информационной среды и в целом к созданию единого информационного пространства знаний.

Современное развитие направления производства информационно-аналитических продуктов как составляющей научно-информационной системы нового уровня невозможно без прогнозно-аналитической оценки научных исследований. В этой связи подготовка информационных продуктов прогнозно-аналитического и обзорного характера, в том числе на основе БД ВИНТИ, является на сегодняшний день в деятельности ВИНТИ РАН важнейшим приоритетом. Использование результатов этой деятельности создает реальную основу для:

- определения тенденций и процессов в научно-технической сфере;
- выявления точек роста, наиболее актуальных и/или стагнирующих научных направлений;
- мониторинга структуры (программ) отечественного научно-промышленного комплекса.

#### Список литературы

1. Шуко Ю.Н. Краткая история, текущее положение и некоторые аспекты развития ВИНТИ РАН. *Научно-техническая информация. Сер. 1.* 2018;(9):1–6.

#### Информация об авторе

**Шуко Юлия Николаевна** – кандидат географических наук, ВРИО директора ВИНТИ РАН, г. Москва, Россия; e-mail: dir@viniti.ru.

#### Information about the author

**Yulia N. Shuko** – Candidate of Sciences (Geography), Acting Director of VINITI RAS, Moscow, Russia; e-mail: dir@viniti.ru.