

КИРИЛЛОВА Ольга Владимировна – кандидат технических наук, заведующая Отделением информационно-библиотечного обслуживания Всероссийского института научной и технической информации (ВИНИТИ) РАН

МЕСТО ИНФОРМАЦИОННЫХ ПРОДУКТОВ ВИНТИ В СРЕДЕ ЭЛЕКТРОННЫХ РЕСУРСОВ ПО НАУКЕ И ТЕХНИКЕ

При решении вопросов о качественном информационном обеспечении российских ученых и специалистов научной и технической информацией неизбежно встает вопрос о составе информационных ресурсов, предназначенных для этой цели. Одним из важнейших блоков таких ресурсов являются отечественные и зарубежные реферативные базы данных (БД), имеющие большие объемы информации и позволяющие оперативно и полно удовлетворять запросы и потребности научно-технического сообщества. Речь идет о ресурсах, имеющих объемы, исчисляемые сотнями тысяч и миллионами документов. На подготовку российских ресурсов затрачиваются государственные средства, также как на покупку зарубежных. В связи с этим встают вопросы о целесообразности подготовки российских ресурсов, отражающих зарубежную информацию, если известно, что российские ресурсы отражают те же виды документов, которые включаются в зарубежные базы данных, доступные российскому пользователю. В то же время дискуссии о целесообразности подготовки российских ресурсов разворачиваются вокруг информационных продуктов, отражающих поток зарубежной опубликованной научно-технической литературы (в основном, периодических и продолжающихся изданий) [1]. Имея в виду, что кроме указанного опубликованного источника, существуют также неформальные обсуждения этого вопроса, обладая некоторой информацией о состоянии таких информационных ресурсов в мире и в России, а также сформулировав свое мнение на проблемы формирования и возможную реорганизацию главного российского ресурса, отражающего отечественную и зарубежную литературу, каким является БД ВИНТИ, автор взяла на себя смелость изложить свою точку зрения по этому поводу. В работе делается акцент именно на анализе ситуации, связанной с БД ВИНТИ, хотя исторически сложилось, что основным ресурсом ВИНТИ является Реферативный журнал, а БД считается его аналогом.

В настоящее время российские информационные ресурсы для обеспечения науки и технологий, обладающие большими массивами, можно пересчитать по пальцам. К ним относятся, прежде всего, БД федеральных (всероссийских) информационных центров, которые в период существования ГСНТИ имели статус всеоюзных центров и находились на 1-м уровне ее иерархической структуры. Им была предписана функция основных генераторов информационных ресурсов, формируемых по тематико-видовому принципу. К счастью для российского пользователя, эти центры и их ресурсы, хоть и с большим трудом и потерями, сохранились и получили дальнейшее развитие. На сегодняшний день мы имеем доступными через интернет: БД диссертаций и отчетов о НИР и ОКР объемом около 1,9 млн. записей ФГУП «ВНТИЦ»; организованные по нескольким поисковым массивам библиографические описания и полные тексты стандартов и других нормативно-технических документов России и зарубежных стран общим объемом примерно 400 тыс. документов ФГУП «Стандартинформ» (бывший ВНИИКИ); БД российских и зарубежных патентов общим объемом примерно 3 млн. документов ФГУП «ФИПС». По патентным документам недавно появилась новая БД ОАО ИНИЦ «Патент» - переведенные в электронный вид информационные бюллетени «Изобретения стран мира» (информация с 2002 г.). Кроме того, имеется федеральная БД ФГУП «ВИМИ» - около 5 млн. документов опубликованной и ведомственной информации. Все реферативные БД доступны через интернет на платной основе. Из зарубежных источников по патентным документам через консорциум НП НЭИКОН доступны по подписке системы Derwent открытые ресурсы Европейского и Американского патентных ведомств и других стран.

Основная информация об опубликованной научной и технической литературе представлена в БД трех всероссийских информационных центров:

- РКП, имеющая БД объемом 1,8 млн. описаний российских изданий и аналитическую роспись статей около 500 журналов;
- ИНИОН, подготавливающий библиографическую БД российской и зарубежной литературы по общественным и гуманитарным наукам (общий объем 3 млн. записей);
- ВИНТИ, генерирующий БД российской и зарубежной опубликованной литературы по точным, естественным, техническим наукам, медицине и экономике, имеющую на сегодняшний день объем 25 млн. реферативных записей.

Автор не имеет цели перечислять все электронные ресурсы, отражающие опубликованную литературу, например, сводные и распределенные электронные каталоги и системы (типа «МАРС»). Эти ресурсы имеют отличные от БД информационных центров принципы формирования и создаются на базе фондов библиотек, которые при существующем состоянии комплектования достаточно бедны, в значительной степени дублируются и имеют широкий профиль комплектования по тематике, видам документов и читательскому назначению.

Из новых российских электронных ресурсов, безусловно, значительным является «Научная электронная библиотека» (<http://elibrary.ru>), которая включает 9,2 млн. статей, 3,2 млн. пристатейных ссылок, оглавления, библиографические описания и авторские резюме 25,9 тыс. наименований журналов, из них 4 тыс. – российских, 3 БД – Web of Science, Zmath, ProQuest Dissertations. Однако важная зарубежная часть этого ресурса прервалась 2004 годом с прекращением формирования его Российским фондом фундаментальных исследований, а российская составляющая начала формироваться только последние 3 года. Этот ресурс имеет также широкую политематическую направленность.

В последние годы наряду с полнотекстовыми зарубежными научными электронными ресурсами российский информационный рынок насыщается крупными зарубежными базами данных, которые предлагают как консорциумы (РФФИ, НП «НЭИКОН») и российские представительства (например, системы STN International), так и российские менеджеры компаний-производителей (Elsevier, ProQuest-CSA и др.). В целом можно сказать, что при наличии потребности, финансовых средств и грамотного подхода, у российских пользователей в настоящее время практически не существует каких-либо серьезных ограничений в доступе к зарубежным базам данных. Наиболее крупными зарубежными производителями информационных ресурсов, отражающих опубликованные источники, являются Chemical Abstracts Service, США (более 200 БД), ProQuest LLC (включает в т.ч. более 100 БД Cambridge Scientific Abstracts), США (более 100 БД), Elsevier Publ., Нидерланды (около 100 БД), Thomson Corporation, США (более 42 БД и системы Dialog-DataStar, являющиеся агрегаторами 1200 БД различного типа) и некоторые другие. Перечисленные компании являются коммерческими. В то же время в США часть важнейших ресурсов подготавливается государственными информационными центрами, такими как Government Research Center, National Information Services Corporation, National Technical Information Service, National Agricultural Library, National Medical Library США (БД Agricola, Medline, Energy Science & Technology, NTIS и др.), которые распространяются как на платной, так и на бесплатной основе. Если говорить о приоритете стран по числу генерируемых ими реферативных БД и индексов, отражающих периодические издания, то по данным БД Ulrich's Periodicals Directory (<http://www.ulrichsweb.com>), из 1,3 тыс. реферативных изданий и индексов примерно 47 % подготавливаются в США, 17 % - в Великобритании, 4 % - в Нидерландах, 2 % - в Германии. 18 % в этом ряду занимает Россия числом выпусков РЖ/БД ВИНТИ. Ресурсы других стран незначительны.

Из этих ресурсов российскому пользователю доступны самые крупные БД, такие как Web of Science (Science Citation Index Expanded, для разных подписчиков России ретроспектива и объем доступных документов различны, 10 тыс. журналов обрабатываются ежегодно, пополнение примерно 1,2 млн. записей в год), SCOPUS (33 млн. зап., около 16 тыс. журналов, ежегодное пополнение более 1,3 млн. зап.), Chemical Abstracts (более 10 тыс. журналов, 28 млн. зап. с 1907 г.), INSPEC (около 10 млн. зап. с 1898 г., ок. 4 тыс. журналов), ProQuest-CSA, EBSCO Publishing (около 15 тыс. журналов).

Однако каково число подписчиков на эти БД? Итак: SCOPUS – 20 организаций; WoS – 6 крупных библиотек РАН с сетью своих библиотек, 6 организаций – через НЭИКОН, через НЭБ доступна ретроспектива WoS до 2004 г.; INSPEC и CSA – в пределах 10-15, EBSCO Publ. – 68 (около 30-ти – организации РАН). Chemical Abstracts доступна через систему STN International (230 БД, 350 млн. док.), центры доступа которой имеют некоторые региональные ЦНТИ Росинформресурса, МЦНТИ, 4 региональных центра РАН. Много ли это? Ответ понятен - очень мало! Это говорит о том, что российский пользователь в основной своей массе не очень стремится использовать зарубежные ресурсы, платить за них, часто незначительные (особенно в рамках консорциума), деньги, не хочет (и в значительной степени не может) читать информацию на английском языке. Безусловно, активность и подготовленность пользователей разная. Значительная часть академических кадров, занимающихся фундаментальными исследованиями, владеет профессиональным английским языком так же, как своим родным. Однако так нельзя сказать о других категориях специалистов и большей части молодых кадров. Опыт информационного обслуживания российских пользователей (студентов, аспирантов, ученых-отраслевиков, производственников) в Центре информационного обслуживания ВИНТИ, который имеет большую часть перечисленных зарубежных БД, также показывает, что уровень подготовленности к восприятию реферативной информации на английском языке очень низок. При опросе, из каких БД по выбору пользователь желает получить информацию, 90 % респондентов заявили, что хотят, прежде всего, получить информацию из БД ВИНТИ, пусть не полную, но с рефератами на русском языке.

Общая беда российской науки – ее качественный спад и потеря кадрового научного потенциала. В качестве формальных признаков этого можно назвать: слабый уровень российских публикаций, не позволяющий включать в обработку ВИНТИ сплошной, без отсева, российский поток; низкий импакт-фактор российских журналов в международной системе ISI Journal Citation Reports, в т.ч. их переводных версий; значительное сокращение с 2004-2005 гг. перечня российских журналов, выходящих на русском языке, обрабатываемых БД SCOPUS (при сохранении в системе их архива); низкое использование иностранной литературы в пристатейных списках, особенно по техническим и прикладным наукам.

В этой ситуации большое значение имеет развитие российских информационных ресурсов. Один из важных новых, развивающихся ресурсов – «Российский индекс научного цитирования» (РИНЦ), который создается на базе НЭБ и строится на основе анализа библиографических списков статей в российских журналах. Однако для быстрого формирования этого ресурса необходимо наличие информационного массива для быстрого и точного распознавания библиографических и конструирования гипертекстовых ссылок. В зарубежных системах цитирования WoS и SCOPUS для этой цели используются собственные ретроспективные

массивы с 1900 г. и с конца 60-х гг. XX в. В НЭБ такого массива нет. Имеющийся библиографический массив зарубежных периодических изданий бывшей НЭБ РФФИ ретроспективой 1997-2003 гг. недостаточен. Для этой цели можно было бы использовать БД ВИНТИ ретроспективой с 1981 г. Идеальным было бы преобразование в электронную форму всего информационного массива РЖ ВИНТИ с 1952 г., исчисляемого еще 30 млн. записей. Однако это, скорее, мечта, которой в нашей действительности вряд ли суждено сбыться.

11 стран (США (34 %), Великобритания (10 %), Германия (7 %), Канада (4 %), Франция, Италия, Австралия, Нидерланды, Япония, Китай, Испания) выпускают 76 % всех изданий, 75 % - научных, 84 % всех находящихся в интернете и 86 % научных, доступных в онлайн. Издания России в этом ряду составляют всего 1,3 %. Научные издания в мировом потоке составляют 64 тыс. периодических продолжающихся изданий ППИ (41 тыс. научных журналов). 59 тыс. ППИ доступны через интернет, в т.ч. 27,6 тыс. научных ППИ (23 тыс. электронных журналов). При этом во всех проанализированных (1,3 тыс.) информационных изданиях (БД, РЖ и индексах) отражается 70,5 тыс. ППИ (29 % от всего потока), из них 37 тыс. научных (62 % научных ППИ), 18,5 тыс. реферируемых научных изданий имеются в онлайн, основная их часть – журналы [2].

Эти и другие приведенные данные позволяют сделать выводы, что: а) мировой поток ППИ, в т.ч. научных, достаточно велик, чтобы быть полностью отраженным в существующих БД и индексах; б) основную часть информации дают издания небольшого числа стран, которые также преобладают в информационных продуктах; в) при сравнении объемов мирового потока и самых крупных баз данных нельзя сказать, что существует хоть одна БД, приближающаяся к отражению большей части мирового потока, точно так же, как не существует БД, полностью дублирующих друг друга. Последнее заключение подтверждается цифрами при сравнении перечней журналов, отражаемых в БД ВИНТИ, с другими крупными БД, например: пересечение с БД SCOPUS – 35 %, CA – 39 %, SCI и INSPEC – 38 %, Metadex – 44 % и т.д. Кроме того, наполнение БД ВИНТИ не дублирует ни одну зарубежную БД, даже если исключить ее российскую часть.

Рассуждая о том, должна ли БД ВИНТИ отражать только российский поток, необходимо учитывать, как количественно выглядит этот поток в настоящее время (не говоря снова об его качестве). По данным статистического справочника «Печать РФ в 2006» [3] в указанный год было издано: всего изданий книжного типа – 102 тыс. названий, в т.ч. научных – 19 тыс. (19 %), из них по естественнонаучной тематике – 3,5 тыс., сельскохозяйственной – 780, медицинской – 930; нормативно-производственных изданий – 1,8 тыс., учебных и методических изданий – 35 тыс. Журналы по всем тематическим направлениям и читательскому назначению составили 4,7 тыс. названий, 37,5 тыс. номеров (всего периодических и продолжающихся изданий 5,4 тыс. названий, 42,8 тыс. номеров). Журналы по естественным, техническим наукам, медицине, сельскому хозяйству, научно-популярные издания, универсального содержания, экономика охватили всего 2,7 тыс. названий, примерно 19 тыс. номеров, из них: наука – 1150 журналов, 6,2 тыс. номеров (1270 с продолжающимися, 6,7 тыс. номеров), экономика – 470 журналов, 4,6 тыс. номеров, универсального содержания – 170 журналов, 990 номеров. При условии, что среднее число статей в научно-технических журналах составляет приблизительно 10 на номер, примерный объем статей в год по указанным направлениям составляет примерно 190 тыс. (без отсева). Между тем, в БД ВИНТИ в 2007 г. было отражено 198,6 тыс. записей российских изданий всех видов. Вполне сопоставимые цифры, говорящие о том, что в БД ВИНТИ российский поток отражается в настоящее время достаточно полно. Безусловно, что работа по его пополнению постоянно продолжается.

В то же время по тематическим разделам российские издания распределяются достаточно неравномерно, и зависит это, прежде всего, от того, к какой категории наук относится раздел. Известно, что в БД ВИНТИ наиболее полно представлены фундаментальные науки: биология – 22,5 % (мировой поток по этому направлению априори больше всех других), физика и астрономия – 12 %, химия и химическая технология – 11 %, науки о Земле – 9 %, и далее автоматика и радиоэлектроника – 7,5 %, медицина – 7 %, машиностроение – 6,5 %, математика – 4 %, охрана окружающей среды и энергетика – по 3,8 %, транспорт и экономика промышленности – по 3 %, металлургия – 2,7 %, механика – 2,4 %, электротехника – 1,6 %, информатика и издательское дело – 0,9 %. Однако российские издания представлены в разделах в совершенно других пропорциях: экономика промышленности – 82 %, транспорт – 53 %, науки о Земле – 52 %, энергетика – 45 % и т.д. На последних 3-х позициях стоят биология (32 %), химия и химическая технология (25 %) и физика с астрономией – 18 %. Причины такого распределения, думаю, понятны. Основные из них – безусловная необходимость наполнения разделов по фундаментальным наукам зарубежной информацией (без которой тематическая дробность подразделов и полнота информации будет крайне низкой) и малый процент российского потока по этим направлениям относительно объемов разделов в целом.

В целом БД ВИНТИ представляет собой в настоящее время информационный массив общим объемом 25 млн. документов ретроспективой с 1981 г., состоит из 29 тематических фрагментов (разделов) и 1 политематической БД (с 2001 г.). Поисковые возможности позволяют ограничивать поиск более дробными тематическими разделами в соответствии с выпусками РЖ или рубриками ГРНТИ или Рубрикатора ВИНТИ. В 2007 г. объем политематической БД составил 550,7 тыс. записей, в т.ч. 206 тыс. отраженных документов (37 %) - на русском языке (включая издания стран СНГ). На обработку в ВИНТИ ежегодно поступает более 80 тыс. экз. научно-технической литературы, около 8 тыс. назв. и 53 тыс. экз. журналов и продолжающихся изданий, 2,2 тыс. названий и 15 тыс. экз. ППИ России, 5,3 тыс. и 36,5 тыс. экз. иностранных ППИ, из них 1 тыс. иностранных журналов обрабатывается из электронных ресурсов. Из 15 электронных ресурсов в обработку поступает 200 тыс. статей. В редакции направляется около 1 млн. статей.

Для ориентации в мировом потоке и для отбора наиболее важной его части в ВИНТИ применяются и разрабатываются формальные и неформальные методы, в т.ч.: известные наукометрические методы (импакт-фактор), методы формальной оценки реферируемого мирового потока, государственные приоритеты развития науки и экономики и т.д. Семь основных приоритетных направлений, все критические технологии и семь направлений основных фундаментальных исследований (ПН-КТ-ОНФИ) в значительной степени охватываются тематическим рубрикатором РЖ/БД ВИНТИ. В то же время, необходимо редактирование Рубрикатора ВИНТИ для достижения более полного соответствия его структуры указанным тематическим государственным приоритетам. При этом в качестве основных принципов формирования входного потока необходимо рассматривать наиболее полное обеспечение ПН-КТ-ОНФИ российским потоком, важнейшими иностранными журналами, имеющими высокие формальные оценки, учитывающие специализацию, научные и экономические приоритеты стран и оценки экспертов.

Резюмируя сказанное и дополняя его, можно сделать заключение, что БД ВИНТИ: а) не дублирует зарубежные БД; б) является наиболее полным источником отечественной опубликованной научно-технической информации; в) представляет собой базу для создания новых ресурсов, в т.ч. для РИНЦ; г) имеет реальную основу для развития в полнотекстовую БД (современная технология копирования статей для реферирования включает одновременно стадию сканирования); д) представляет политематический ресурс, отражающий значительную часть ПН-КТ-ОНФИ.

В то же время в среде быстро развивающихся на новых технологических платформах зарубежных ресурсов для сохранения актуальности, повышения качества и конкурентоспособности информационных продуктов ВИНТИ (БД и РЖ), требуется их безотлагательная реорганизация, основным звеном которой должно стать изменение технологии их подготовки с переходом на полную электронную подготовку и преобразование БД ВИНТИ в основной ресурс. Это позволит: а) дополнять БД постоянно, в том числе ретроспективной информацией, заполняя лакуны независимо от года издания; б) расширить функциональные возможности поисковой системы БД, включив в неё систему перевода с английского и с русского языков; в) включать в БД авторские резюме на английском языке, имея в виду наличие системы перевода; г) сохранить БД в качестве основного государственного информационного ресурса опубликованных документов по стратегически важным направлениям развития науки и технологий; д) сертифицировать БД по международным требованиям и включать ее в международные информационные системы.

Список литературы

1. Антопольский А., Каленов Н. Информационно-библиотечная сфера: Камо грядеши или вопросы стратегии // Информационные ресурсы России. – 2007. – № 6. – С. 11-17.
2. Кириллова О.В. Периодические и продолжающиеся издания в реферативных базах данных: количественная и качественная оценка // НТИ. Сер.1. – 2007. – № 8 – С. 26-33.
3. Печать Российской Федерации в 2006 г.: Статистический сборник / РКП. – М., 2007.