

История создания технологий для формирования Базы данных и Реферативного журнала ВИНТИ РАН с конца 1980-х годов (краткая справка)¹

Совершенствованию технологий формирования Базы данных (БД) и Реферативного журнала (РЖ) ВИНТИ РАН всегда придавалось первостепенное значение, так как они являются основными научно-информационными продуктами (НИП) Института.

В 1970-е годы в ВИНТИ была создана Автоматизированная справочно-информационная система по науке и технике (АССИСТЕНТ), которая регулярно совершенствовалась. В конце 1980-х годов эксплуатировалась её версия 2А. Она обеспечивала комплектование и обработку входного потока отечественных и зарубежных публикаций по точным, естественным и техническим наукам, аналитико-синтетическую обработку публикаций в отраслевых отделах, выпуск РЖ. Компьютерные средства применялись для ведения рубрикаций, на стадии издания номеров РЖ с указателями, а также при формировании результирующих БД. В Институте и Производственно-издательском комбинате (ПИК) ВИНТИ использовались ЭВМ Минск-32, ЕС ЭВМ, станции сбора данных, фотонаборные машины ДИГИСЕТ.

Одновременно перед ВИНТИ встали проблемы повышения качества и расширения спектра выпускаемых НИП на базе новых средств – персональной вычислительной техники с гибкими возможностями компьютерных издательских систем, развитыми коммуникационными средствами и т.п. При этом была осознана потребность в оптимизации производственного процесса на основе ресурсосберегающих технологий на различных участках обработки информации. Практически одновременно велись работы по нескольким направлениям: совершенствование системы АССИСТЕНТ, проекты «Гамма», «Пума».

В 1989 г. в Институте был создан Отдел программных систем (ОПС). Целью его работы являлось создание прикладных программных средств, обеспечивающих в ВИНТИ автоматизированный процесс обработки научно-технической литературы (НТЛ) и производства НИП. Эта задача могла быть решена путём модернизации системы производства на основе технологической базы данных (ТБД), реализующей так называемый «процессный» подход.

В 1989–1990 гг. работы базировались на весьма скромных средствах: 8 персональных компьютеров, объединенных в локальную сеть; программные средства – информационно-поисковые системы CDS/ISIS и RBASE, издательский пакет СПНТ (система программированного набора текстов), одноранговая сетевая операционная система RPTINET, пакет КЕРМИТ. В результате был создан алфавит и коммуникативный формат, удобные для использования на персональных компьютерах; построен комплекс прикладных программ для библиографической обработки публикаций; создан Автоматизированный центр обработки первоисточников (АЦОП) и реализована сетевая связь с ним для получения описаний выпусков журналов, зарегистрированных во входном потоке; произведены первые оригинал-макеты номеров РЖ с использованием специализированной издательской системы.

В 1991–1996 гг. была сформулирована Концепция ТБД и выполнено её натурное моделирование. Использовались программные средства: СУБД Advanced revelation, языки программирования С и Pascal, издательская система «Ксерокс вентура паблишер», сетевая операционная система Nowell. Для обеспечения функционирования компонентов ТБД построена первая очередь локальной сети, объединившая три выделенных файл-сервера и около 30 рабочих мест. Важнейшие результаты этого этапа:

- углубление автоматизированной обработки входного потока: от регистрации выпусков изданий в АЦОП – на постатейный уровень; создание экспериментального производственного участка обработки журнальных статей (автоматизированные рабочие места, штатное расписание, нормы выработки); начало издания нового продукта – бюллетеня «Новые поступления. Содержание советских журналов»;

¹ Использованы материалы статьи «Современные технологии в производстве информационных продуктов» / Шапкин А. В. // НТИ. сер. 2. Информационные процессы и системы / ВИНТИ РАН. – 2003. – № 12. – С. 1–18. <http://amb.viniti.ru/sid2/sid2free?sid2=J038659483X>, а также отчеты по НИР «Научно-методическое и программно-технологическое обеспечение интерактивного формирования информационных продуктов. Развитие и совершенствование классификационных систем УДК, ГРНТИ и Рубриката ВИНТИ».

- действующая модель полного цикла обработки потока публикаций и выпуска номеров РЖ, а также БД ВИНТИ, построенная на Концепции ТБД; система была внедрена в эксплуатацию на полном потоке Отдела научной информации по биологии и обеспечивала формирование как штатных выпусков РЖ, так и новых НИП;
- компьютерная система производства НИП, которая может использоваться как во взаимодействии с ТБД, так и независимо; налажено производство оригинал-макетов номеров РЖ по биологии, новых видов НИП, а также предоставление пользователям РЖ в электронной форме;
- экспериментальный участок набора данных на персональных компьютерах в ПИК ВИНТИ.

В 1997–2003 гг. выполнены промышленные разработки компонентов ТБД. Накопленные ранее результаты исследований и натурного моделирования были реализованы в промышленном масштабе на базе более мощных программных и технических средств. Началось использование развитой локальной сети с многопроцессорными серверами баз данных, архитектурой «клиент-сервер», реляционной моделью данных, СУБД MS SQL-Server, системой программирования Delphi и C++, доступом к электронным публикациям через Интернет. Для идентификации объектов начали применяться штрих-коды.

Результатами этих работ стали:

- ввод в промышленную эксплуатацию ключевых компонентов ТБД: системы регистрации входного потока (1999 г.), автоматизированной системы комплектования (2001 г.), системы ведения единого Рубрикатора и Регистра информационных продуктов (2001 г.);
- образование Центра развития информационных систем, в котором была создана техническая база для промышленных разработок автоматизированных технологий; принято решение о разработке новой системы регистрации входного потока взамен АЦОП.
- построение системы ведения единого Рубрикатора ВИНТИ и рубрикации РЖ;
- внедрение технологии однократного реферирования, разработанной в ПИК ВИНТИ.

В 2004 г. в эксплуатацию введены технологии и программные средства аналитической обработки выпусков изданий и тематической разметки документов, созданные в 2002–2003 гг.

Концепция построения функционирующей в ВИНТИ РАН Единой технологической базы данных (ЕТБД) соответствует современным тенденциям развития обслуживания в сфере НТИ. В частности, можно отметить следующие свойства ЕТБД:

- наличие в её составе (в качестве основополагающего компонента) Системы классификационных схем (включающей более 20 классификаторов, которые используются в крупных национальных и международных информационных органах) с установленными смысловыми взаимосвязями между понятиями различных классификаций;
- расширение электронной составляющей входного потока с использованием роботизированных средств съема метаданных из открытых ресурсов сети Интернет и с применением автоматизированных ресурсосберегающих методов библиографической обработки данных;
- автоматическая тематическая классификация публикаций на основе словарно-семантического анализа метаданных с использованием методик искусственного интеллекта;
- широкое внедрение средств интерактивного взаимодействия с исполнителями (научными редакторами, референтами, корректорами) в процессе обработки документов и выпуска НИП;
- автоматическое формирование НИП по гибко задаваемым тематическим профилям (по заказам потребителей).

Формируемые на основе ЕТБД реферативно-библиографические продукты ВИНТИ РАН предоставляются пользователям в следующих видах:

- доступ в режиме онлайн к Электронному каталогу научно-технической литературы;
- в традиционной печатной форме доступно примерно 180 печатных тематических выпусков РЖ;
- на компакт-дисках – номера Электронного РЖ, которые по своему содержанию и поисковому аппарату максимально приближены к печатным изданиям;
- результаты многоаспектного диалогового поиска в информационно-поисковой системе «Сокол» – в режиме онлайн или на компакт-дисках.