

ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОЙ РАБОТЫ

УДК 004.65:[002:330.133] ВИНТИ

В. Г. Шамаев, А. В. Жаров, А. Б. Горшков

Единая технологическая база данных для подготовки информационных продуктов ВИНТИ

Информационно-технологический комплекс ВИНТИ пополнен Единой технологической базой данных (ЕТБД), которая предназначена для ввода и обработки библиографической и аналитической информации и подготовки информационных продуктов таких, как реферативные журналы в печатной и электронной формах и реферативные базы данных. Описывается технология наполнения ЕТБД и получение на ее основе информационных продуктов.

ВВЕДЕНИЕ

В начале 1990-х гг. все более возрастающий поток научно-технической информации привел к необходимости решения задачи как можно более полной автоматизации процесса информационного обслуживания. К этому времени удачно совпало, что кроме субъективной необходимости, имелись и объективные предпосылки для решения таких задач. Сложилась весьма благоприятная ситуация, связанная с бурным развитием вычислительной техники, программных продуктов, информационно-телекоммуникационной среды и сетевых технологий. В это время в ВИНТИ на новой электронно-вычислительной базе началась разработка Информационно-технологического комплекса.

В настоящее время Информационно-технологический комплекс ВИНТИ, охватывающий все аспекты подготовки информационных продуктов (от подписки на научно-техническую литературу до выпуска РЖ, сигнальной информации, баз данных, серий Итогов науки и техники), обеспечивает автоматизацию процесса подготовки информационных продуктов и среди прочих решает следующие задачи:

- а) сокращение сроков прохождения документов от момента поступления их в ВИНТИ до момента предоставления потребителю созданных информационных продуктов;
- б) повышение качества подготовки основных информационных продуктов ВИНТИ — выпусков Реферативного журнала и баз данных;
- в) создание новых информационных продуктов;
- г) предоставление информационных продуктов как в печатном виде, так и на современных носителях (CD-ROM, DVD и т. п.) и с использованием сетевых технологий (Интранет, Интернет и т. д.).

Все это способствует эффективному оперативному информационному обслуживанию внутренних и внешних пользователей ВИНТИ.

Большое значение при выполнении подобного рода работ имеет унификация комплекса, соблюдение единого внутрисистемного формата данных

ВИНТИ, задаваемого на сегодняшний день Научно-техническим предписанием ВИНТИ 10-2004 [1].

В процессе разработки комплексной системы автоматизированной технологии подготовки информационных продуктов важно, сохраняя специфику технологического процесса ВИНТИ и его информационной составляющей, учесть следующее:

- исключить дублирующие процессы при подготовке информационных продуктов;
- осуществить оптимальную схему прохождения документов (бумажных и электронных);
- использовать возможности Интранет ВИНТИ и Интернет;
- реализовать основной системообразующий и функциональный принцип построения информационной системы — обеспечение одноразовой аналитической переработки документов и многократного и многоцелевого использования ее результатов.

Информационно-технологический комплекс ВИНТИ разрабатывается как развивающаяся система, реализующая существующую в институте идеологию аналитико-синтетической переработки научно-технической информации. На вход поступают документы на электронных и печатных носителях, а также из сетевых ресурсов. Производимые информационные продукты, в том числе: библиографические, реферативные, классификационно-систематизирующие, полнотекстовые — могут предоставляться пользователям также на различных носителях и через Интернет.

- Структурно при этом выделяются три этапа:
- аналитико-синтетическая переработка документов, поступающих на бумажных и электронных носителях;
 - формирование и ведение разнообразных электронных ресурсов (каталоги поступлений, базы данных, электронные реферативные журналы, электронная библиотека);
 - формирование и централизованное ведение классификационно-систематизирующих информационных объектов (массивы рубрикаторов и рубрикаций, словарей ключевых слов и др. терминологии).

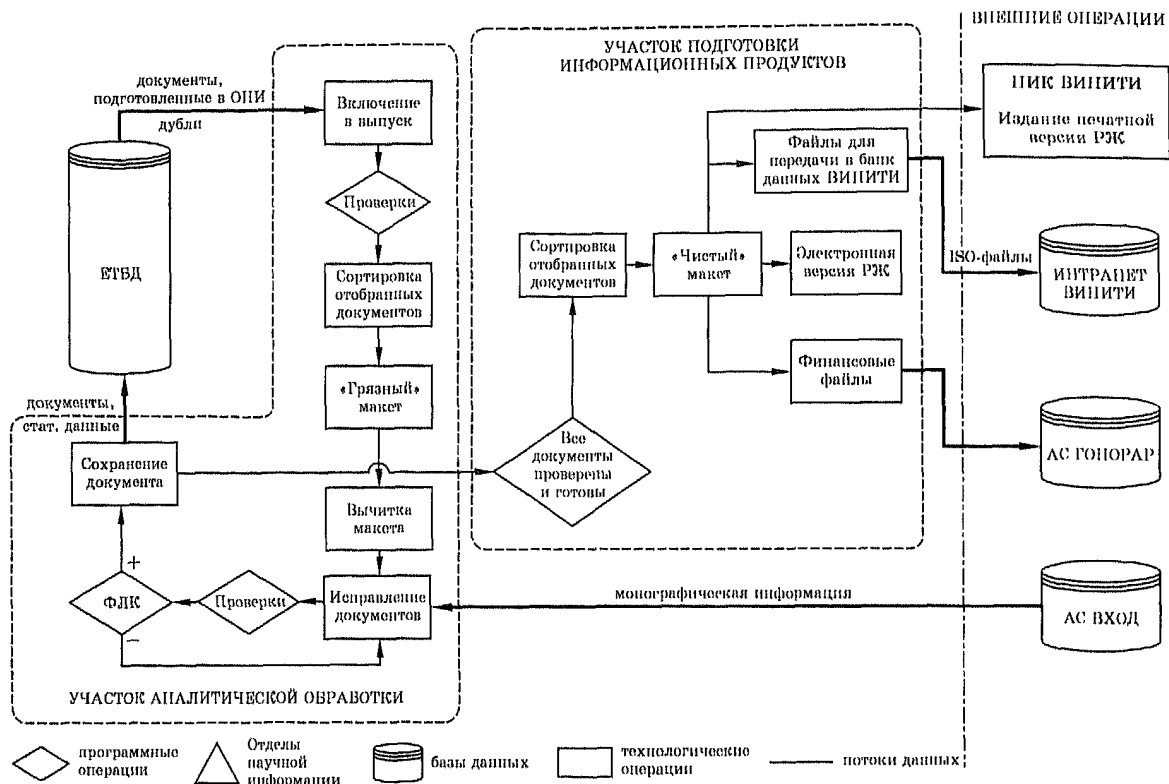


Рис. 2. Технология использования Единой технологической базы данных для получения конечного продукта

В соответствии со схемой на рис. 1 документы, снабженные регистрационно-учетными сведениями и технологическим системным идентификатором документа (СИД-1), становятся доступными по коду отраслевой и тематической разметки редакторам соответствующих выпусков РЖ для работы.

Согласно технологическим правилам, принятым в информационно-технологическом комплексе ВИНТИ, в качестве кодов разметки используются буквенно-цифровые индексы тематических выпусков РЖ. В ходе разметки каждый документ может быть снабжен только одним основным кодом (обязательно) и несколькими дополнительными кодами (факультативно) в зависимости от выявленных тематических аспектов в содержании документа. Дополнительные коды могут быть также добавлены на любом этапе обработки.

Далее первичные документы с кодами тематических выпусков поступают на участок библиографической обработки Научно-технологического отдела (НТО) ВИНТИ для ввода библиографической информации и в ПИК ВИНТИ для набора содержательной части (реферат, рубрики, ключевые слова и т. д.). Возможность набора содержательной части предусмотрена также на участке библиографической обработки НТО.

В ходе ввода библиографической информации в ЕТБД создаются записи описаний документов, которые на этом этапе заполняются не только библиографическими, но и общими сведениями о документе и учетно-технологическими данными. Одновременно с библиографическими данными вводятся коды тематической разметки. На этом этапе вся информация проходит корректуру, при необходимости вносятся исправления и исправленные документы хранятся в ЕТБД.

Содержательная переработка первичных документов предусматривает возможность использования как традиционной - "бумажной" технологии с заполнением формуляров и с последующим набором информации с них на персональных ПК в ПИК ВИНТИ, передачей электронных файлов результатов набора в ВИНТИ и импортом их в ЕТБД, так и автоматизированной "безбумажной" технологии с передачей информации напрямую в ЕТБД.

В первом случае результаты аналитического реферирования и индексирования документов по профильному тематическому аспекту референт наносит в специальный предмашинный формуляр. Редактор осуществляет научное и литературное редактирование рефератов, вносит рубрикационные шифры и ключевые слова, дополняет предмашинные формуляры финансовыми сведениями и направляет их на участок клавиатурного набора в ПИК ВИНТИ. После клавиатурного набора и корректуры файлы порций реферативных записей на машиночитаемом носителе передаются для загрузки в ЕТБД.

Во втором случае референт работает с электронным шаблоном для ввода реферата и пересылает документ по электронной почте или передает на машиночитаемом носителе. Редактирование, рубрицирование и индексирование проводятся редактором в электронной форме.

Результаты содержательной переработки документов, независимо от того, на каком технологическом участке они были произведены, автоматически объединяются с соответствующими библиографическими записями по идентификатору документа (рабочему номеру или СИД-2) и становятся доступными для дальнейшей многоаспектной и многоцелевой автоматизированной обработки и формирования информационных продуктов.

Рубрикационные шифры (РБШ) присваиваются документам в процессе индексирования по Рубрикатору ВИНТИ, независимо для каждого тематического аспекта, в ходе основной или дополнительной переработки. Если какой-либо тематический аспект документа индексирован несколькими РБШ, то первый из них по порядку — считается главным, а остальные — второстепенными (дополнительными). Главный РБШ используется в Информационно-технологическом комплексе ВИНТИ при компьютерной сборке тематического выпуска РЖ в качестве критерия для опубликования реферата в соответствующем разделе/подразделе РЖ. Дополнительные РБШ служат для ссылок на этот реферат в других разделах/подразделах данного выпуска РЖ. Каждый документ может быть помещен в нескольких тематических выпусках РЖ в зависимости от количества кодов тематической разметки в составе записи документа. При компьютерной сборке номеров сводных томов РЖ и политематической Базы данных ВИНТИ все РБШ, которыми индексирован документ, автоматически объединяются в единый сводный набор РБШ. Аналогично осуществляется дополнительное индексирование ключевыми словами.

СОСТАВ И СТРУКТУРА ЕТБД

На участке ввода (см. рис. 1) каждый документ заносится в ЕТБД после выполнения следующих операций:

- импорт монографического описания (*обязательно*);
- ввод библиографического описания (*обязательно*);
- ввод реферативной части, рубрикационных шифров, поискового образа документа (ПОДа) и т. д. (*необязательно* — эта информация может быть добавлена на других стадиях обработки).

Для выполнения этих операций предназначенное автоматизированное рабочее место ввода (АРМ Ввода), с помощью которого решаются и следующие вспомогательные задачи:

- ввод данных по оплате труда референта, операторов набора и корректоров;
- ввод корректуры БО и рефератов;
- а также производятся служебные операции: вывод набранных документов на печать для передачи корректору; сбор статистических данных по набору и корректуре.

В отдельный поток выделены операции импорта данных из внешних систем:

- электронные библиотеки,
- внештатные референты,
- Производственно-издательский комбинат (ПИК ВИНТИ) и т. д.

Эти операции выполняются с помощью автоматизированного рабочего места администратора (АРМ Администратора), что позволяет импортировать и экспортировать документы в ЕТБД, проверять их, удалять, собирать статистическую информацию и др.

На участке аналитической обработки (см. рис. 2) в отраслевых отделах ВИНТИ выполняются следующие операции:

- редактирование рефератов;

- ввод и корректура рубрикационных шифров, поисковых образов документов (ПОД) и т. д.;
- подготовка выпусков РЖ;
- вспомогательные операции: составление, дополнение и корректура списков ключевых слов; подготовка рубрикатов, их дополнение и корректура.

На участке подготовки информационных продуктов (см. рис. 2) производится:

- подготовка данных к выгрузке в Банк данных ВИНТИ;
- подготовка оригинал-макетов РЖ для печати;
- подготовка выпусков электронного РЖ;
- подготовка финансовой информации и передача ее в АС "Гонорар".

ОСНОВНЫЕ СТРУКТУРНЫЕ ЭЛЕМЕНТЫ ЕТБД И ИХ КРАТКОЕ ОПИСАНИЕ

Единая технологическая база данных (ЕТБД) представляет собой совокупность библиографической и содержательной информации (реферат, РБШ, ключевые слова и т. д.) и служебных данных, характеризующих обрабатываемый поток научно-технической литературы с точки зрения прохождения определенных технологических этапов.

Документы организованы в виде совокупности таблиц:

ЕТВД_А — таблица, содержащая системный идентификатор издания (СИД-1), библиографические данные, сведения о вспомогательных операциях;

ЕТВД_Х — таблица, содержащая системный идентификатор документа (СИД-2), данные о тематической разметке документов, содержательную информацию и служебную информацию о стадиях обработки, пройденных каждым документом;

ЕТВД_М — таблица, содержащая системный идентификатор издания (СИД-1), монографическое описание и служебную информацию.

Помимо основных, динамически заполняемых таблиц, в ЕТБД включены служебные таблицы и таблицы, играющие роль справочников при вводе информации в ЕТБД и при проведении различных проверок:

NTP_Lab — список элементов данных согласно НТП ВИНТИ (метка элемента данных, его название, формат отображения в РЖ);

NTP_Doc — распределение элементов данных по видам документов и по уровням описания.

Users_ACT — список пользователей АРМов ввода, администратора и редактора (идентификатор пользователя, его Ф. И. О., права доступа и др.);

Correctors — список корректоров (идентификатор и Ф. И. О. каждого корректора);

List_Country — список стран (полное название страны, сокращенное название, код страны и т. д.);

List_Language — список языков (полное наименование языка, сокращенное наименование и т. д.);

List_Character — список возможных значений элемента данных "Характер документа".

При обработке документов также используются специализированные базы данных, входящие в интегрированное информационное пространство ВИНТИ:

БД рубрикатов;

БД монографической информации (АС "Вход").

Данные каждого документа поступают в ЕТБД в соответствии с технологической цепочкой, представленной рядом АРМов: АРМ Ввода (библиографическое описание); АРМ Администратора БД (загрузка рефератов, импорт документов из других источников), АРМ Редактора (присвоение документам издательских номеров, ввод РБШ и ключевых слов).

На этапе ввода библиографического описания в ЕТБД происходит формирование записи в таблице ЕТВД_А.

Каждый отдельный вид документа представлен своей формой ввода. Вся библиографическая информация, вводимая в форму, объединяется в интегральный массив в виде XML-строки и заносится в отдельное поле таблицы ЕТВД_А. В массиве каждый элемент данных снабжен меткой в соответствии с НТП ВИНТИ.

Помимо библиографических данных для каждого документа вводится СИД-1 (системный идентификатор документа), который заносится в поле SID1 таблицы ЕТВД_А и служит ключом для поиска монографической информации. Если монографическая информация документа найдена, то в таблице проставляется признак готовности записи к следующему этапу обработки и формируется запись в таблице ЕТВД_Х.

Дата появления сведений о документе в таблице фиксируется в полях DTCreate таблицы ЕТВД_Х и Date_А таблицы ЕТВД_А. Шифр оператора, отвечающего за ввод данных, заносится в поле Operator таблицы ЕТВД_А.

АРМ Ввода предполагает возможность тематической разметки документа, т. е. формы ввода содержат специальные поля набора для задания кодов первичной и вторичной тематической разметки, а также поля набора для ввода шифров основного и дополнительного выпусков.

Если тематическая разметка проводится на этапе ввода библиографии, то на текущий документ в таблице ЕТВД_Х — заполняются следующие поля:

FirstTCod1 — код первичной тематической разметки;

FirstTCod2 — шифр основного выпуска;

IDN_А — ссылка на запись о документе в таблице ЕТВД_А;

DTCreate — дата появления записи в таблице ЕТВД_Х (в данном случае совпадает со значением в поле Date_А таблицы ЕТВД_А).

Коды вторичной тематической разметки и шифры дополнительных выпусков записываются в поле IRFieldsX таблицы ЕТВД_Х в XML-формате.

Следующий технологический этап формирования документа — подготовка и ввод в ЕТБД содержательной части — также поддерживается АРМом Ввода.

Помещенные в окна ввода реферат, фамилия автора реферата, перевод заглавия и список ключевых слов с полями дескрипторов добавляются в поле IRFieldsX таблицы ЕТВД_Х в XML-формате.

Предложенная технология содержательной переработки предполагает также возможность ввода в ЕТБД содержательной информации, набранной с формуляров в ПИК ВИНТИ и передаваемой в виде файлов в формате ISO2709. В этом случае производится пакетная загрузка данных в таблицу

ЕТВД_Х, минуя технологические этапы ввода содержательной информации. Загрузка данных осуществляется в АРМе Администратора БД. Связь с библиографическими данными в таблице ЕТВД_А осуществляется с помощью системного идентификатора, присваиваемого каждой записи таблицы ЕТВД_А.

ИСПОЛЬЗОВАНИЕ ЕТБД ДЛЯ ПОЛУЧЕНИЯ КОНЕЧНОГО ПРОДУКТА

Документы в ЕТБД, прошедшие этап ввода реферата, становятся доступны редактору через АРМ Редактора при просмотре готовых рефератов и направлении документов в текущий выпуск.

АРМ Редактора обеспечивает выполнение следующих функций (см. рис. 2):

- просмотр, отбор и редактирование рефератов;
- формирование выпуска;
- генерация файла оригинал-макета выпуска;
- генерация файлов выпуска для передачи в Банк данных ВИНТИ.

Редактору доступны для обработки документы, размеченные для его тематической области кодом отдела и шифром выпуска.

Просмотр назначаемых в выпуск документов осуществляется на двух уровнях. Первый (основной) уровень — просмотр документов, изначально размеченных шифром выпуска редактора. Для этих документов редактор может при необходимости ввести данные по оплате реферата и присвоить документу рубрику. Редактор также имеет возможность внести правки в реферат, перевести или поправить заглавие и ввести или поправить ключевые слова. Заполняются следующие поля таблицы:

DT5 — дата помещения документа в текущий выпуск;

User5 — идентификатор редактора.

После того, как редактор заканчивает работу с документом на первом уровне, документ может быть "размножен" в таблице ЕТВД_Х. Это происходит в том случае, если на этапе тематической разметки были заданы коды вторичной тематической разметки и шифры дополнительных выпусков. При "размножении" документа в ЕТВД_Х появляется $N = K + V$ новых записей (K — количество кодов вторичной разметки, V — количество дополнительных выпусков).

У всех записей—"дублей" в поле IDCopy стоит ссылка на "родительский" документ (т. е. документ, послуживший источником для "размножения"). Признаки прохождения этапа включения в выпуск отсутствуют (не заполнены поля DT5, User5).

Записи, полученные после "размножения", служат материалом для просмотра документов на втором уровне, т. е. отобранных по кодам дополнительной разметки. Для этих документов редактор должен ввести новую рубрику и ключевые слова и, если необходимо, отредактировать реферат и перевод заглавия.

Заключительный этап технологии реализуется функцией пакетного формирования выпуска в АРМе Редактора. На этом этапе происходит группировка готовых к выпуску документов по рубрикам и формирование файла оригинал-макета выпуска в формате LaTeX. Созданный таким образом файл

оригинал-макета может быть направлен в ПИК ВИНТИ для подготовки тиража реферативного журнала. Кроме того, на этом этапе происходит вывод документов выпуска РЖ в файл в формате ISO2709 для передачи в Банк данных ВИНТИ, а также подготовка файлов с данными об оплате для загрузки в АС "Гонорар".

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Информационно-технологический комплекс ВИНТИ, состоящий из АС "Комплектование", АС "Вход", БД Рубрикатора, отраслевых отделов научной информации, Банка данных, пополнился Единой технологической базой данных. ЕТБД с приданными ей функциями формирования конечных информационных продуктов значительно расширяет возможности комплекса и придает ему законченный вид. Процесс, начинающийся с комплектования потока НТЛ и заканчивающийся созданием информационных продуктов и услуг, в настоящее время полностью автоматизирован. По важной составляющей комплекса — ЕТБД, разработаны:

- структура ЕТБД;
- таблицы и связь между ними;
- система записи библиографической и содержательной информации в формате XML;
- автоматизированные рабочие места ввода б/о и редактора;
- система формально-логического контроля (ФЛК) на всех участках обработки;
- средства администрирования ЕТБД;
- средства сбора статистической и бухгалтерской информации;
- средства подготовки конечных информационных продуктов:
 - оригинал-макетов РЖ;
 - данных к загрузке в Банк данных ВИНТИ;
 - электронных журналов.

Помимо этого обеспечено взаимодействие с другими информационными системами ВИНТИ:

- использование БД монографической информации ВИНТИ;
- использование БД рубрикатора ВИНТИ;
- передача данных в АС "Гонорар";
- передача данных в Банк данных ВИНТИ.

Таким образом, с вводом ЕТБД решены или появилась возможность решить следующие задачи:

- использование входного потока научно-технической литературы (НТЛ) не только на бумажных, но и на электронных носителях (CD-ROM, DVD, Интернет, электронные архивы и др.);
- полная и однократная библиографическая переработка входного потока документов;

- многоаспектная переработка документа на содержательном уровне (редактирование, индексирование, рубрицирование);
- использование описания документа в различных информационных продуктах посредством индексирования на основе предметных и системных классификаторов;
- унификация по форме, наполнению и процедурам формирования основных информационных продуктов ВИНТИ (РЖ, БД) и централизация системы выпуска основных информационных продуктов;
- обеспечение самостоятельности отраслевых отделов научной информации в технологии формирования (на основе ЕТБД) дополнительных информационных продуктов ВИНТИ;
- контроль формирования структурированных записей описания документов на всех этапах переработки (в соответствии с НТП);
- организационно-технологический контроль прохождения документов на всех этапах переработки;
- централизованное управление всем технологическим циклом обработки потоков НТЛ на разных уровнях;
- Web-технология обработки НТЛ на основе Web-доступа референтов и редакторов к процессу аналитико-синтетической переработки документов;
- оперативное обеспечение доступа внутренних (участвующих в процессе обработки НТЛ) и внешних (пользующихся произведенными информационными продуктами) пользователей к электронным продуктам и хранилищам данных: Каталог поступлений НТЛ — на аналитическом уровне, выпускам Электронного РЖ, электронному архиву копий первоисточников (Электронной библиотеке) и др.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Представление элементов данных во внутрисистемном формате ВИНТИ. Нормативно-техническое предписание НТП ВИНТИ 10-2004. — М.: ВИНТИ, 2004. — 104 с.
2. Шамаев В. Г., Жаров А. В., Батурина О. Н., Горшков А. Б., Лось Е. К., Лукашевич Н. Л., Максимов И. Н., Седякина А. Н., Старцева О. Б., Щербина-Самойлова М. Б., Ягельницкая О. А. Разработка технологии создания единой технологической базы данных для подготовки информационных продуктов ВИНТИ. — М., 2005. — 72 с. Ил. 41. Библиогр. — 2 назв. — Рус. — Деп. в ВИНТИ 07.11.2005, № 1430-В2005.

Материал поступил в редакцию 22.12.05.