

# ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОЙ РАБОТЫ

УДК [004.57]:002.63 ВИНТИ

Л. Ф. Борисова

## Автоматизированная информационно-технологическая система ВИНТИ по наукам о жизни

### Часть 1. Принципы построения автоматизированной технологии переработки потока документов

*В рамках задачи по созданию системы информационного обеспечения по наукам о жизни разработана и внедрена в опытно-промышленном режиме автоматизированная технология подготовки информационных продуктов по соответствующей тематике.*

*Основой функционирования автоматизированной технологии является концепция технологической БД, логическая структура которой отражает технологические взаимосвязи и состояния всех документальных потоков. Приводится общая характеристика тематического профиля, входного потока и основных системно-обязательных выходных продуктов, а также архитектуры системы. Подробно рассматривается полный технологический цикл переработки информации как поэтапный автоматизированный процесс, сочетающий научную деятельность и операции рутинного типа. Представлены результаты опытно-промышленной эксплуатации автоматизированной системы переработки документов по наукам о жизни.*

Разработка и внедрение автоматизированной технологии подготовки информационных продуктов по наукам о жизни осуществлялись в рамках научно-издательских и производственно-технологических задач ВИНТИ по созданию системы информационного обеспечения исследований в различных направлениях науки и техники.

Система информационного обеспечения понимается нами не только как аналитико-синтетическая переработка и стратегическое накопление всей доступной информации в определенном систематическом порядке. Мы рассматриваем ее как системный комплекс научных, программно-технологических, аналитических, прогнозных, текстово-поисковых, технических и организационных и маркетинговых функций, реализуемых с помощью информационных методов, для формирования входного потока документов, извлечения из первичного документа необходимых текстовых, графических и фактографических данных, упорядочения, хранения и предоставления их в таком виде, в каком они максимально доступны для конкретного потребителя, в том числе индивидуального пользователя [1-5].

В соответствии с современными тенденциями, система информационного обеспечения должна ориентироваться на конкретные пользовательские потребности — с одной стороны, и на интеграцию с другими информационными системами по биологии, медицине, биотехнологии — с другой сторо-

ны. Кроме того, система должна быть очень гибкой и самообучающейся: от источника информации — к предложению новых информационных возможностей и, наоборот, от потребностей пользователя — к источнику информации, к формированию входного потока.

Представляемая серия статей отражает результаты разработки и внедрения новой автоматизированной технологии, подчиненной логике функционирования всей системы информационного обеспечения наук о жизни как гибкого инструмента реализации ее задач.

При разработке автоматизированной технологии за основу была взята система "Ассистент" [6], которая поэтапно внедрялась в ВИНТИ с 1971 г. Ориентация на "Ассистент" позволила обеспечить преемственность разработок и функционально-технологическую совместимость с действующей в ВИНТИ технологией. На сегодня, существующая в ВИНТИ технология в некотором смысле исчерпала свой потенциал и стала тормозом для процесса создания новых информационных продуктов и услуг и повышения качественного уровня информационного обслуживания. Во-первых, эта технология жестко ориентирована на издание Реферативного журнала (РЖ), а не машиночитаемого массива данных, с которого возможен автоматизированный съем информационных продуктов, в том числе и РЖ. Во-вторых, существующая техноло-

гия опирается на независимую обработку всех документов для каждого выпуска РЖ, в котором они могут быть отражены, в соответствии с описываемыми в нем научными аспектами. При этом полностью дублируются все операции: копирование документа, составление библиографического описания, реферирование, редактирование, индексирование ключевыми словами и рубрикационными шифрами, а также наборные работы. И в-третьих, — отсутствует автоматизированный контроль процесса обработки документов, что влечет за собой большой объем ручных операций при учете проводимых работ, а также является причиной недостаточной полноты и оперативности самой системы контроля.

Отмеченные недостатки определили основные задачи развития автоматизированной технологии:

формирование документально-реферативной базы данных как основы для создания широкого спектра информационных продуктов — машиночитаемых и печатных;

исключение дублирования в обработке документов;

обеспечение гибкой структуры описания документов в базе данных и управления доступом к полям описаний документов;

реализация автоматизированного контроля процесса обработки документов.

Кроме того технология должна повлечь сокращение ручного труда, а также сроков переработки материалов. Основная конечная цель развития технологии — это обеспечение возможности создания новых информационных продуктов и услуг, отвечающих современным потребностям ученых и специалистов и повышение качества информационного обслуживания, конкурентноспособного мировым аналогам.

Большинство из этих принципов изложены на уровне задач в новых модификациях системы "Ассистент" ВИНТИ — "Ассистент 3" [7] и "Ассистент 3/94" [8], которые в силу разных финансовых и организационных причин пока не были масштабированы.

Автоматизированная информационно-технологическая система ВИНТИ по наукам о жизни (АИТС "Биология") представляет собой совокупность организационно-технологических и программно-технических средств, позволяющих осуществлять аналитико-синтетическую переработку входного документального потока научной литературы по профилю системы с учетом перечисленных выше целей и задач, а также эффективное информационное обслуживание ученых и специалистов. Функционирование системы основывается на принципах интегральных информационных систем и обеспечивает исчерпывающую аналитико-синтетическую переработку каждого первичного документа, поступившего на вход системы, и многократное многоцелевое использование результатов переработки для формирования различных выходных информационных продуктов и услуг.

Разработка системы проводилась совместно специалистами Отделения научной информации по проблемам наук о жизни и Отдела программных систем ВИНТИ [9–12], а также при участии и финансовой поддержке ПИК ВИНТИ.

## ОБЩАЯ ХАРАКТЕРИСТИКА СИСТЕМЫ

*Тематический профиль* АИТС "Биология" охватывает все разделы Рубрикатора ГАСНТИ по наукам о жизни (табл. 1).

Таблица 1

Тематический профиль системы в терминах рубрик верхнего уровня Рубрикатора ГАСНТИ и Рубрикатора ВИНТИ

Код ГАСНТИ	Код ВИНТИ	Наименование рубрики
41	341	Биология
21	621	Биотехнология
14	314	Биохимия
43	343	Физико-химическая биология
51	151	Психология
81	681	Сельское и лесное хозяйство
61	761	Медицина и здравоохранение

Данная тематическая область реализована в системе в виде двух классификаторов:

Рубрикатора ВИНТИ по наукам о жизни — основан на Рубрикаторе ГАСНТИ, имеет глубину до 9 уровней, содержит порядка 15 тыс. тематических рубрик, ежегодно актуализируется, служит в системе для индексирования документов и автоматизированного формирования запросов на поиск данных, обеспечивающих решение текущих технологических задач;

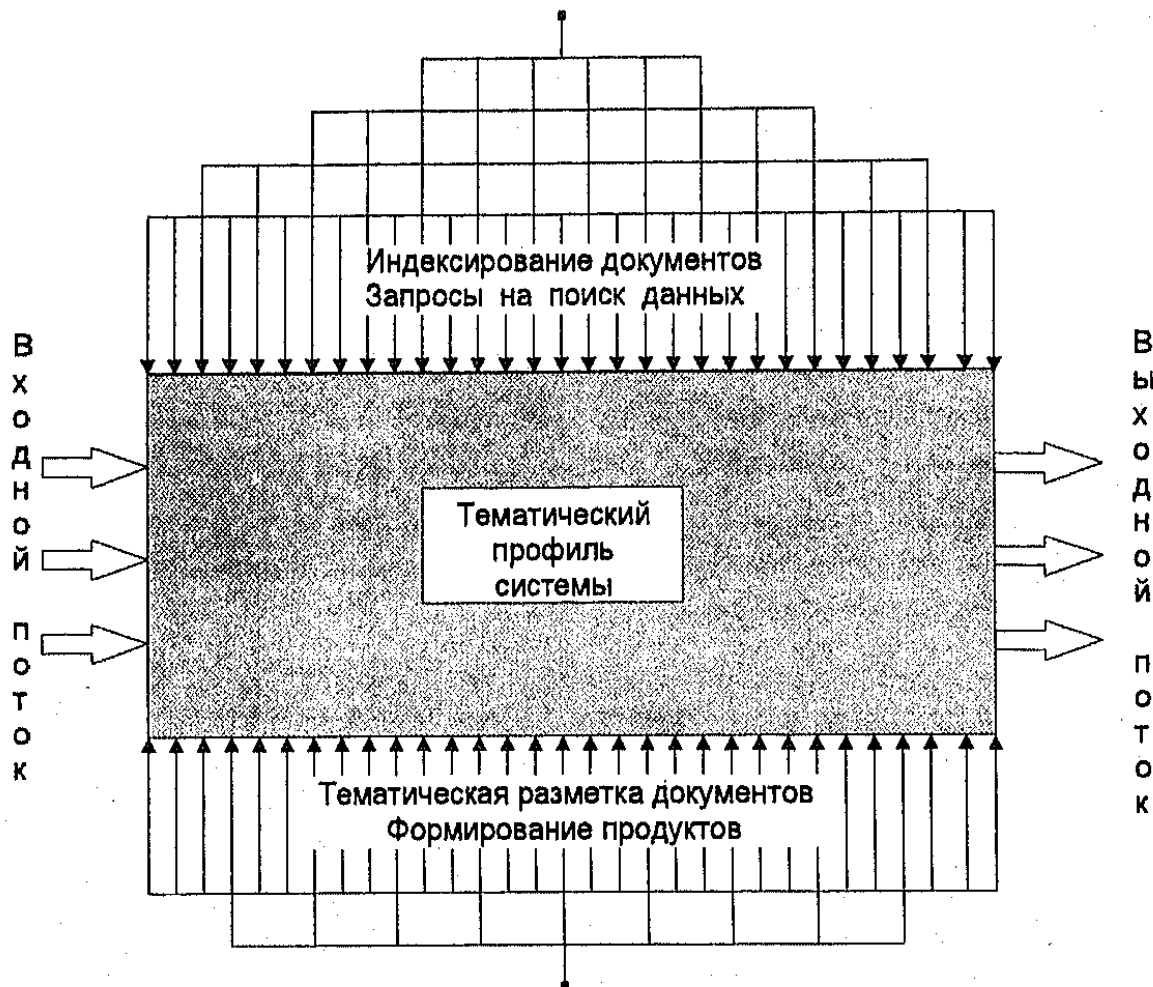
совокупности Рубрикаций машиночитаемых и печатных изданий — базируется на Рубрикаторе ВИНТИ, имеет код своего издания, используется в системе для тематической разметки входного документального потока и формирования выходных продуктов.

В процессах творческой работы специалиста, а также автоматизированной переработки документы проходят сквозь встречные потоки двух классификаторов, что позволяет получать на выходе системы продукты с заданными качествами (рис. 1).

*Входной поток системы* образуют все виды отечественных и зарубежных научных изданий по наукам о жизни. В 1996 г. суммарный поток составил 230 тыс. документов из 2948 первоисточников (2425 — зарубежных и 523 — отечественных), поступающих по различным каналам — это валютная подписка, из Библиотеки по естественным наукам РАН, из Государственной публичной научно-технической библиотеки, по обмену на издания ВИНТИ, а также личные экземпляры ученых и по другим каналам.

*Основным системно-обязательным выгодным продуктом* является документально-реферативная База данных (БД) ВИНТИ по наукам о жизни, годовой объем составляет 260 000 документов. Структура и тематическое содержание БД задаются Рубрикатором ВИНТИ по наукам о жизни.

В процессе создания БД автоматически формируются 55 печатных регламентных выпусков РЖ



Рубрикации выходных продуктов

Рис. 1. Тематический профиль системы, реализуемый Рубрикатором ВИНТИ по наукам о жизни и Рубрикациями выходных продуктов

с указателями, 12 разделов, сводные тома "Биология" и "Медицина", а также 14 проблемных Реферативных сборников (табл. 2). Разделы и сводные тома РЖ не представляют собой формальное слияние массивов отдельных выпусков РЖ, а являются самостоятельными продуктами. Структура и тематическое наполнение печатных изданий задаются соответствующими рубриками, отражающими логику конкретной науки или одной из ее проблем.

Все эти издания по охвату тематической области можно подразделить на политематические и узкотематические, по межтематическим связям —

с полным и частичным дублированием тематики, по назначению — фундаментальные и проблемные. В системе заложена возможность формирования изданий одновременно на русском и английском языках.

Подготовка основных информационных продуктов осуществляется в системе одновременно и требует одинакового сопровождения. Поэтому далее, для упрощения схемы описания системы, в качестве выходного продукта будет указываться традиционный продукт ВИНТИ — Реферативный журнал.

Таблица 2

Тематический профиль системы — печатные издания

Код издания	Уровень	Наименование выпуска РЖ	Периодичность в год
04	св. том	БИОЛОГИЯ	12
04 А	раздел	ОБЩИЕ ПРОБЛЕМЫ БИОЛОГИИ	12
04А1	выпуск	ОБЩИЕ ПРОБЛЕМЫ БИОЛОГИИ	12
04А2	выпуск	ОБЩАЯ ЭКОЛОГИЯ. БИОЦЕНОЛОГИЯ	12
04А3	выпуск	БИОНИКА. БИОКИБЕРНЕТИКА	12
04А4	выпуск	РАДИАЦИОННАЯ БИОЛОГИЯ	12
04Б	раздел	ВИРУСОЛОГИЯ И МИКРОБИОЛОГИЯ	12
04Б1	выпуск	ВИРУСОЛОГИЯ	12
04Б2	выпуск	МИКРОБИОЛОГИЯ ОБЩАЯ	12

Код издания	Уровень	Наименование выпуска РЖ	Периодичность в год
04Б3	выпуск	МИКРОБИОЛОГИЯ ПРИКЛАДНАЯ	12
04Б4	выпуск	МИКРОБИОЛОГИЯ САНИТАРНАЯ	12
04В	раздел	БОТАНИКА	12
04В1	выпуск	БОТАНИКА (ВЫСШИЕ РАСТЕНИЯ)	12
04В2	выпуск	БОТАНИКА (ВОДОРОСЛИ, ГРИБЫ)	12
04В3	выпуск	ФИЗИОЛОГИЯ И БИОХИМИЯ РАСТЕНИЙ	12
04В4	выпуск	РАСТЕНИЕВОДСТВО	12
04В5	выпуск	ФИТОПАТОЛОГИЯ	12
04В6	выпуск	ЛЕКАРСТВЕННЫЕ РАСТЕНИЯ	12
04В7	выпуск	ЛЕСОВЕДЕНИЕ И ЛЕСОВОДСТВО	12
04В8	выпуск	ПОЧВОВЕДЕНИЕ И АГРОХИМИЯ	12
04Д	раздел	ФИЗИКО-ХИМИЧЕСКАЯ БИОЛОГИЯ	12
04Д1	выпуск	БИООРГАНИЧЕСКАЯ ХИМИЯ	12
04Д2	выпуск	БИОФИЗИКА	12
04Д3	выпуск	БИОХИМИЯ (ОБЩИЕ ВОПРОСЫ)	12
04Д4	выпуск	БИОХИМИЯ (АССИМИЛЯЦИЯ АЗОТА И БЕЛКИ)	12
04Д5	выпуск	БИОХИМИЯ (КСЕНОБИОТИКОВ)	12
04Д6	выпуск	МОЛЕКУЛЯРНАЯ БИОЛОГИЯ	12
04Д7	выпуск	МОЛЕКУЛЯРНАЯ ИММУНОЛОГИЯ	12
04И	раздел	ЗООЛОГИЯ	12
04И1	выпуск	ЗООЛОГИЯ ОБЩАЯ	12
04И2	выпуск	ЗООПАРАЗИТОЛОГИЯ	12
04И3	выпуск	ЭНТОМОЛОГИЯ	12
04И4	выпуск	ИХТИОЛОГИЯ	12
04И5	выпуск	ЗООЛОГИЯ НАЗЕМНЫХ ПОЗВОНОЧНЫХ (ОБЩАЯ)	12
04И6	выпуск	ЗООЛОГИЯ НАЗЕМНЫХ ПОЗВОНОЧНЫХ (ОРНИТОЛОГИЯ)	12
04И7	выпуск	ЗООЛОГИЯ НАЗЕМНЫХ ПОЗВОНОЧНЫХ (ТЕРИОЛОГИЯ)	12
04И8	выпуск	БИОЛОГИЯ СЕЛЬСКОХОЗЯЙСТВЕННЫХ ЖИВОТНЫХ	12
04К	раздел	ИММУНОЛОГИЯ. АЛЛЕРГОЛОГИЯ	12
04К1	выпуск	ИММУНОЛОГИЯ. АЛЛЕРГОЛОГИЯ	12
04М	раздел	ФИЗИОЛОГИЯ И МОРФОЛОГИЯ ЧЕЛОВЕКА И ЖИВОТНЫХ	12
04М1	выпуск	МОРФОЛОГИЯ ЧЕЛОВЕКА И ЖИВОТНЫХ (АНТРОПОЛОГИЯ)	12
04М2	выпуск	ФИЗИОЛОГИЯ ЧЕЛОВЕКА И ЖИВОТНЫХ (КРОВЬ, ЛИМФА)	12
04М3	выпуск	ФИЗИОЛОГИЯ ЧЕЛОВЕКА И ЖИВОТНЫХ (НЕЙРОФИЗИОЛОГИЯ)	12
04М4	выпуск	ФИЗИОЛОГИЯ ЧЕЛОВЕКА И ЖИВОТНЫХ (ОБМЕН ВЕЩЕСТВ)	12
04М5	выпуск	ФИЗИОЛОГИЯ ЧЕЛОВЕКА И ЖИВОТНЫХ (ОБЩИЕ ПРОБЛЕМЫ)	12
04М6	выпуск	ФИЗИОЛОГИЯ ЧЕЛОВЕКА И ЖИВОТНЫХ (ЭНДОКРИННАЯ СИСТЕМА)	12
04М7	выпуск	ОБЩИЕ ВОПРОСЫ ПАТОЛОГИЧЕСКОЙ АНАТОМИИ	12
04Н	раздел	ОНКОЛОГИЯ	12
04Н1	выпуск	ОНКОЛОГИЯ ЭКСПЕРИМЕНТАЛЬНАЯ	12
04Н2	выпуск	ОНКОЛОГИЯ КЛИНИЧЕСКАЯ	12
04Н3	выпуск	ОНКОЛОГИЯ (ТЕРАПИЯ ОПУХОЛЕЙ)	12
04П	раздел	ПСИХОЛОГИЯ	12
04П1	выпуск	ПСИХОЛОГИЯ	12
04Р	раздел	БИОТЕХНОЛОГИЯ	12
04Р1	выпуск	БИОТЕХНОЛОГИЯ	12
04Т	раздел	ФАРМАКОЛОГИЯ. ТОКСИКОЛОГИЯ	12
04Т1	выпуск	ФАРМАКОЛОГИЯ ОБЩАЯ И НЕРВНОЙ СИСТЕМЫ	12
04Т2	выпуск	ФАРМАКОЛОГИЯ ЭФФЕКТОРНЫХ СИСТЕМ	12
04Т3	выпуск	КЛИНИЧЕСКАЯ ФАРМАКОЛОГИЯ	12
04Т4	выпуск	ТОКСИКОЛОГИЯ	12
04Т5	выпуск	НАРКОЛОГИЧЕСКАЯ ТОКСИКОЛОГИЯ	12
04Я	раздел	ГЕНЕТИКА. ЦИТОЛОГИЯ	12
04Я1	выпуск	ОБЩАЯ ГЕНЕТИКА	12
04Я2	выпуск	ГЕНЕТИКА МИКРООРГАНИЗМОВ	12
04Я3	выпуск	ГЕНЕТИКА ВОЗДЕЛЫВАЕМЫХ РАСТЕНИЙ	12
04Я4	выпуск	ГЕНЕТИКА СЕЛЬСКОХОЗЯЙСТВЕННЫХ ЖИВОТНЫХ	12
04Я5	выпуск	ГЕНЕТИКА ЧЕЛОВЕКА	12
04Я6	выпуск	ЦИТОЛОГИЯ	12
04Ф	пробл. св. том	МЕДИЦИНА	12
04А5	пробл. сборник	Лучевая диагностика. Часть 1. Ультразвуковая диагностика	12
04А6	пробл. сборник	Лучевая диагностика. Часть 2. Рентгенодиагностика. Магнитно-резонансная томография	12
04А7	пробл. сборник	Лучевая диагностика. Часть 3. Радионуклидная диагностика	12
04А8	пробл. сборник	Эндоскопия	12

Код издания	Уровень	Наименование выпуска РЖ	Периодичность в год
04А9	пробл. сборник	Геронтология. Гериатрия	6
04Б5	пробл. сборник	Туберкулез	12
04К3	пробл. сборник	Бронхиальная астма и другие аллергические заболевания	12
04М8	пробл. сборник	Клиническая эндокринология	12
04М9	пробл. сборник	Реаниматология и интенсивная терапия	4
04М11	пробл. сборник	Гематология	6
04П2	пробл. сборник	Психиатрия	12
04Т6	пробл. сборник	Алкогольная болезнь	12
04Т7	пробл. сборник	Табачная зависимость и курение табака	6
04Т8	пробл. сборник	Медицина труда	12

## КОНЦЕПЦИЯ ТЕХНОЛОГИЧЕСКОЙ БАЗЫ ДАННЫХ

Системообразующим принципом построения АИТС "Биология" является концепция технологической базы данных (ТБД). Согласно этой концепции технологическая БД рассматривается как общий информационный ресурс, совместное использование которого позволяет персоналу системы осуществлять переработку входных документальных потоков и готовить к выпуску информационные продукты. Логическая структура технологической БД отражает технологические взаимосвязи и текущие состояния всех документальных потоков, циркулирующих в системе.

### Классификация, макроструктура и назначение технологической БД

По сфере применимости, выполняемым функциям и характеру хранимой информации технологическую БД можно классифицировать как проблемно-ориентированную операционную базу данных научно-технической информации. По типу модели данных технологическая БД относится к классу полностью структурированных информационных массивов с детерминированной схемой, ориентированной на предварительное определение всех форматов, свойств и связей данных. По особенностям макроструктуры технологическая БД относится к классу составных (интегрированных) баз данных и включает ряд структурно самостоятельных и семантически однородных информационных массивов, которые совместно используются в технологическом процессе, но каждый из которых имеет свое функциональное назначение, свой состав и логическую структуру записей. Состав и взаимосвязи основных массивов технологической БД отражены на рис. 2.

Основное назначение технологической БД связано с хранением рабочего массива аналитических описаний документов (АОД), находящихся на различных стадиях автоматизированной переработки, и обеспечением оперативного машинного доступа

к записям аналитического описания документа в ходе выполнения технологических операций. Технологическая БД обеспечивает процесс автоматизированной переработки документов классификаторами, словарями и другой специальной информацией. Поэтому помимо массива аналитического описания документа в состав технологической БД входят и поддерживаются в актуальном состоянии обеспечивающие информационные массивы:

регистрационно-библиографического описания выпусков сериальных изданий (ТБД/РВО);

кодов тематической разметки документов (ТБД/КТР);

словарей формально-логического контроля полей данных (используемые коды, сокращения, аббревиатуры и пр.) (ТБД/ФЛК);

шифров рубрик Рубрикатора ВИНТИ (ТБД/РВШ);

словарей используемых ключевых слов (ТБД/КС);

рубрикаций выходных информационных продуктов (ТБД/РВЦ).

Состав некоторых обеспечивающих массивов поддерживается в актуальном состоянии соответствующими специализированными службами и подсистемами ВИНТИ. Так, например, актуализация массива регистрационно-библиографического описания выпусков сериальных изданий является одной из регламентных функций подсистемы ведения Электронного каталога первоисточников, действующей на участке "Вход" ВИНТИ.

В связи с тем, что процессы актуализации таких массивов, как словари используемых ключевых слов, содержащих нормированную лексику системы, требуют методико-технологической поддержки, в составе технологической БД предусмотрен вспомогательный рабочий массив новой неопознанной лексики (ТБД/ННЛ). Эти записи содержат сведения о новых лексических единицах, зарегистрированных системой, включая частоту использования, источники и тематическую принадлежность, и применяются для интеллектуальной обработки

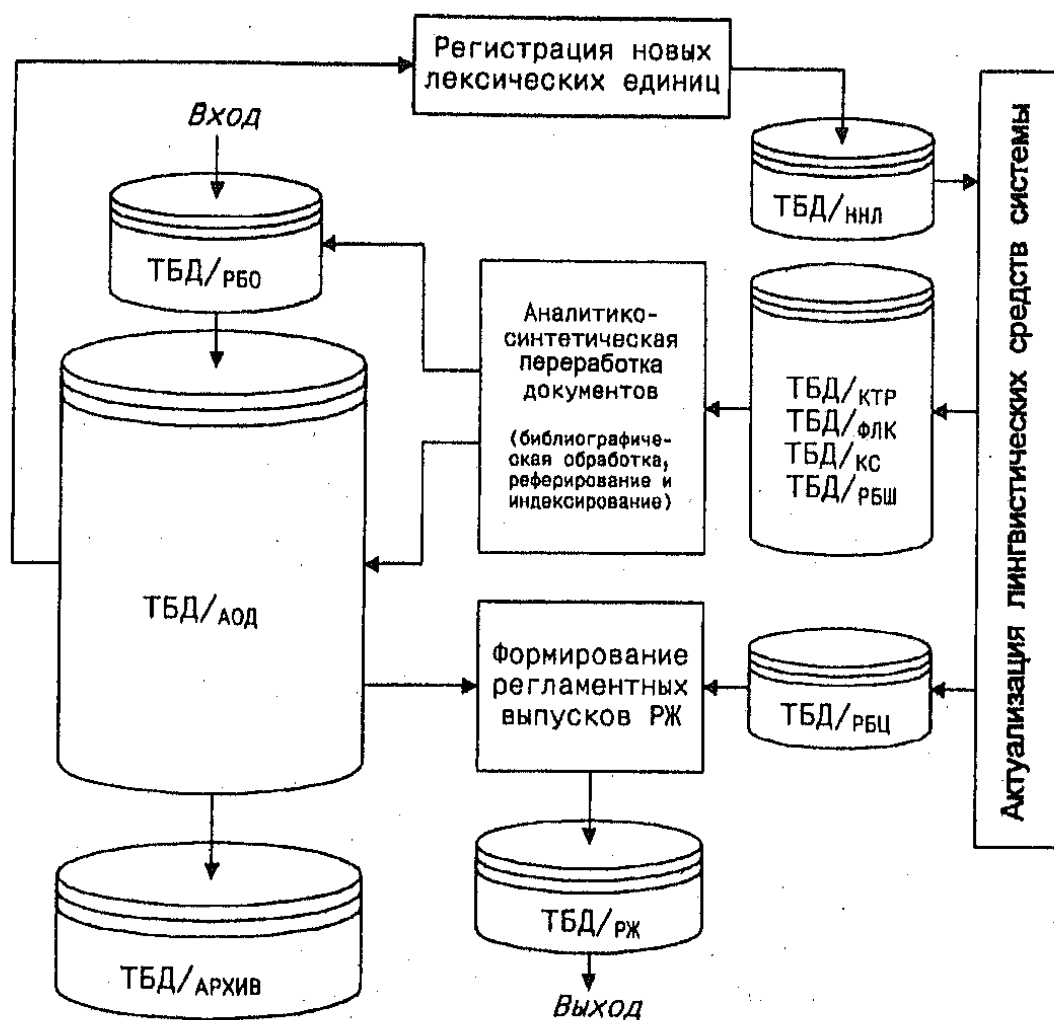


Рис. 2. Макроструктура технологической базы данных: ТБД/АОД — рабочий массив аналитических описаний документов; ТБД/РБО — массив регистрационно-библиографических описаний первоисточников серийных изданий; ТБД/КТР — таблица кодов тематической разметки; ТБД/ФЛК — словари формально-логического контроля; ТБД/КС — словари используемых ключевых слов; ТБД/РБШ — таблица рубрикационных шифров Рубрикатора ВИНТИ; ТБД/РБЦ — массив рубрикаций выходных информационных изданий; ТБД/РЖ — рабочий массив формируемых в системе номеров выпусков РЖ; ТБД/ННЛ — рабочий массив новой непознанной лексики; ТБД/РЖ — рабочий массив формируемых в системе номеров выпусков РЖ; ТБД/АРХИВ — технологический фонд записей АОД

и принятия решения о корректировке словарно-терминологических средств системы.

Массив рубрикаций выходных информационных продуктов используется в процессах автоматизированной сборки регламентных выпусков РЖ.

#### Основной информационный объект технологической БД

Основным информационным объектом технологической БД являются записи аналитического описания документа (АОД).

Общий состав полей данных каждой такой записи представлен в табл. 3. В соответствии с внутренним форматом, принятым в ВИНТИ [2], каждое поле записи обозначено уникальным 3-х значным цифровым кодом — меткой поля. Форматы отдельных элементов записи аналитического описания документа используют либо текстовые поля переменной длины, либо буквенно-цифровые коды. Ряд полей допускает наличие множественных значений данных (например, поле библиографического описания с меткой 001 содержит перечень авторов документа, а поле индексов с меткой 186 содержит список рубрикационных шифров).

Таблица 3

#### Общий состав полей (формат) записи аналитического описания документа

№ п/п	Метка поля	Наименование поля
1	410	Идентификатор документа (ИД)
2	035	Код вида документа (КВД)
2	050	Системный идентификатор документа (СИД)
3	211	Код ВИНТИ
Элементы БО:		
4	001	Автор(ы) индивидуальный(е)
5	003	Сокращенное оригинальное заглавие издания
6	004	Язык(и) текста
7	007	Дата издания, депонирования
8	021	Основное оригинальное заглавие документа
9	042	Код страны издания (1)

№ п/п	Метка поля	Наименование поля
10	043	Страницы/объем
11	052	ISSN (1)
12	074	Том(а) сериального издания (1)
13	076	№ выпуска, ддмм, мм, сезон (1)
14	041	Язык резюме
15	047	Сведения к основному оригинальному заглавию
16	038	Приложение
17	048	Место издания (1)
18	051	Регистрационный № микрофиши
19	012	Коллективный заявитель/патентовладелец
20	018	Обозначение вида патентного документа
21	022	Коллективный автор (сокращенно) (1)
22	024	№ части статьи
23	025	Заглавие части статьи
24	031	Параллельное сокращенное заглавие издания
25	040	Страна выдачи патентного документа
26	046	Место издания
27	053	ISBN (1)
28	054	Том, выпуск, часть
29	059	Сведения к сокращенному заглавию издания
30	065	Часть тома/выпуска
31	077	Часть № выпуска сериального издания (1)
32	083	Дата подачи заявки
33	090	Индекс МКИ
34	092	№ патентного документа
35	094	Регистрационный № заявки
36	095	Дата публикации заявки
37	099	Индекс НКИ
38	009	Бывшее сокращенное заглавие издания
39	013	Редактор(ы)
40	015	Сведения о библиографии
41	017	№ депонированной работы
42	027	Указание о редакторах
43	033	Издательство (1)
44	044	Иллюстрации/карты
45	061	Дата регистрации депонированной работы
46	084	Место хранения/депонирования
47	086	Дата заявки конвенционального приоритета
48	096	Местонахождение организации(й)
49	097	Код страны конвенционального приоритета
50	098	№ конвенционального приоритета
51	101	Бюллетень № (к полю 95 для патентов)
52	008	Обозначение вида диссертации
53	045	Указание о дискуссии и страницы
54	062	Место защиты диссертации
55	078	Дата защиты диссертации
56	091	Ученая степень и область наук
57	204	Указание о параллельном языке текста
58	023	Заглавие тома/выпуска/части
59	060	Общее обозначение материала (микроформа)
60	075	Область серии
61	081	Периодичность
62	082	Цена подписная (за номер)
63	093	№ издания/характеристика
64	218	Место работы автора
65	219	Адрес для корреспонденции
66	010	Бывшее полное заглавие издания
67	011	Составитель(и)
68	014	Индивидуальный заявитель/патентовладелец

№ п/п	Метка поля	Наименование поля
69	016	Параллельное полное заглавие издания
70	026	Указание о составителе
71	028	Страна стандарта
72	030	Место издания (2)
73	032	Издательство (2)
74	049	Сопроводительный материал
75	055	Обозначение ранее действовавшего документа
76	056	Дата утверждения нормативно-технологического документа
77	057	Дата введения и сроки стандарта
78	058	Дата окончания срока стандарта
79	063	Сокращенное оригинальное заглавие сброшюрованного издания
80	064	Код страны издания (2)
81	066	Дата издания (2)
82	067	ISBN (2)
83	068	ISSN (2)
84	069	Том(а) сериального издания (2)
85	070	Заглавие части тома/выпуска
86	071	Числовой масштаб
87	072	Обозначение и № подсерии
88	073	№ выпуска, ддмм, мм, сезон (2)
89	079	Индекс и цифровое обозначение стандарта
90	080	Обозначение классификации стандарта
91	085	Указание о сотрудничестве
92	087	Перечень учреждений сотрудничества
93	089	Индекс НКС
94	201	Полное оригинальное название издания, имеющего подсерию
95	202	Заглавие подсерии
96	203	Сотрудничество (перечень лиц)
97	208	Сведения к заглавию сброшюрованного издания
98	209	Коллективный автор (сокращенно) (2)
99	210	Часть № выпуска сериального издания (1)
100	212	Сокращенное оригинальное заглавие сопроводит. материала
101	213	Заглавие статьи — ответ на дискуссию
102	214	Автор(ы) ответа на дискуссию
103	215	№ микрофиши сопроводительного материала
104	216	№ кадров сопроводительного материала
105	102	Шифр по обмену
106	103	Шифр депонирования
		<b>Код(ы) тематической разметки:</b>
107	286	Код тематической разметки основной (КТРосн)
108	287	Код(ы) тематической разметки дополнительный(е) (КТРдоп)
		<b>Элементы реферативного описания:</b>
109	411	Архивный номер первичного документа
110	037	Код УДК
111	180	Характер документа
112	182	Перевод основного оригинального заглавия (рус.)
113	183	Перевод основного оригинального заглавия (англ.)
114	192	Реферат (рус.)
115	193	Реферат (англ.)
116	019	Автор реферата
117	196	Сведения о библиографии

№ п/п	Метка поля	Наименование поля
		<b>Финансовые поля:</b>
118	170	Шифр референта
119	171	Расценка за реферат
120	172	Шифр редактора (1)
121	173	Расценка за редактирование (1)
122	174	Шифр редактора (2)
123	175	Расценка за редактирование (2)
		<b>Поисковый образ документа (ПОД):</b>
124	220	Метки ключевых слов
125	222	Ключевые слова (рус.)
126	223	Ключевые слова (англ.)
		<b>Рубрикационные шифры (РБШ):</b>
127	186	Список рубрикационных шифров основных (РБШосн)
128	187	Список рубрикационных шифров дополнительных (РБШдоп)
129	188	Расширение гл. рубрикационного кода основного
130	189	Расширения гл. рубрикационных кодов дополнительных
		<b>Поля связи с номером выпуска РЖ:</b>
131	296	Код номера выпуска профильного РЖ
132	297	Коды номеров выпусков непрофильных РЖ
133	291	Номер реферата в профильном выпуске РЖ
134	292	Номера рефератов в непрофильных выпусках РЖ

№ п/п	Метка поля	Наименование поля
135	293	Дата опубликования в профильном РЖ
136	294	Даты опубликования в непрофильных РЖ
		<b>Поля технологического контроля:</b>
137	801	
138	803	
139	804	
140	805	
141	813	Дата клавиатурного ввода реферата
142	814	
143	815	
		<b>Индикатор технологического состояния:</b>
586		Код профильного технологического состояния
587		Коды непрофильных технологических состояний

**Поле идентификатора.** В момент создания каждой записи аналитического описания документа автоматически присваивается уникальное значение идентификатора (ИД), которое не изменяется в ходе всего технологического процесса и используется в качестве основного ключа доступа к основным информационным объектам. В качестве идентификатора принят 11-разрядный буквенно-цифровой код.

**Код вида документа.** Помимо идентификатора каждой записи документа в момент ее создания присваивается единственное значение кода вида до-

Таблица 4

## Виды первичных документов

№ п/п	КВД	Литера	Вид документа	Состав БО	
1	931	П	Статья в сериальном издании	СИД + БО <sub>М</sub> + БО <sub>А</sub>	
2	117		Патент/заявка в журнале	СИД + БО <sub>М</sub> + БО <sub>А</sub>	
3	118		Статья из журнала в журнале	СИД + БО <sub>М</sub> + БО <sub>А</sub>	
4	119		Дискусс. статья в сериальн. изд-и	СИД + БО <sub>М</sub> + БО <sub>А</sub>	
5	120		Статья из журнала с микроформой	СИД + БО <sub>М</sub> + БО <sub>А</sub>	
6	634		Статья в книге/сборнике	БО <sub>М</sub> + БО <sub>А</sub>	
7	733		Статья в сб. депонир. научн. работ	БО <sub>М</sub> + БО <sub>А</sub>	
8	832	Ж	Отдельный выпуск журнала	СИД + БО <sub>М</sub>	
9	113		Вновь вышедший журнал	СИД + БО <sub>М</sub>	
10	436		Книга	БО <sub>М</sub>	
11	111		Диссертация	БО <sub>М</sub>	
12	114		Атлас	БО <sub>М</sub>	
13	115		КРТ	БО <sub>М</sub>	
14	327		ПР	БО <sub>М</sub>	
15	238		ДЕП	БО <sub>М</sub>	
16	139		П	Отдельный патентный документ	БО <sub>М</sub>
17	110		С	Стандарт	БО <sub>М</sub>

КВД — код вида документа по классификации ВИНТИ;

СИД — системный идентификатор документа;

БО<sub>М</sub> — библиографическое описание монографического уровня;

БО<sub>А</sub> — библиографическое описание аналитического уровня.

Некоторые виды документов помечаются в информационных изданиях специальной литерой — буквенным индексом.



кумента (КВД), которое в дальнейшем не изменяется. По классификации ВИНТИ используются 17 видов документов, каждому из которых соответствует 3-х значный цифровой код (табл. 4). Все виды документов подразделяются на два класса: документы, которые имеют только монографический уровень библиографического описания (книга, диссертация, депонированная научная работа и пр.), и документы, которые имеют монографический и аналитический уровень библиографического описания (статья в сериальном издании, статья в сборнике и пр.).

**Поле системного идентификатора.** Системный идентификатор документа (СИД) представляет собой уникальный регистрационный код номера выпуска первоисточника вида сериаль-

ного издания по Электронному каталогу первоисточников ВИНТИ и используется в технологическом процессе для связи записи аналитического описания документа с регистрационно-библиографической записью в массиве технологической БД.

**Поля библиографического описания.** Обязательной составной частью каждой записи является совокупность полей библиографического описания (БО) [13]. Для различных видов документов в состав полей БО входят либо только элементы монографического уровня (БО<sub>М</sub>), либо элементы монографического и аналитического уровней (БО<sub>А</sub>). Общий состав полей БО и состав полей БО на различных уровнях по видам документов представлен в табл. 5.

Таблица 5

Состав полей БО в записи аналитического описания документа — по видам документов

КВД	Метки полей БО <sub>М</sub>
931	[050] [003] 060 031 059 022 046 009 [007] 074 076 077 038 084 052 042
117	[003] [007] 074 076 077 052 084 042
118	[050] [003] 060 059 022 046 [07] 074 076 077 063 208 209 066 069 073 210 084 052 068 042
119	[050] [003] 060 031 059 022 046 009 [007] 076 077 038 084 052 042
120	[050] [003] 046 009 [007] 074 076 077 084 052 068 042
634	[003] 060 059 054 023 065 022 093 [048] [007] 084 053 067 052 042 064
733	[003] 059 054 065 042
832	[050] [021] [201] 072 202 060 016 047 022 046 010 [07] 074 076 077 038 043 044 004 052 084 042
113	[050] [021] [201] 072 202 060 016 047 003 031 059 022 027 013 [048] 033 030 032 [007] 081 052 082 084 042
436	[021] 060 047 054 023 065 070 001 026 011 022 027 013 085 087 203 093 [048] 033 030 032 [007] 043 044 049 075 004 041 053 067 052 084 042 064
111	[021] 060 [008] [091] 054 023 [001] [062] [078] 043 044 049 004 042
114	[021] 047 054 023 065 070 022 001 026 011 027 013 093 [048] 033 030 032 [007] 043 044 075 004 053
115	042
327	[021] 047 022 026 011 027 013 093 071 [048] 033 030 032 [007] 044 075 004 053 042
238	[021] 047 [022] [096] [007] 043 044 004 042
139	[021] 047 054 023 065 070 001 026 011 027 013 [022] [096] [007] 043 044 015 004 084 061 017 042
110	[021] [018] [92] [40] [90] 001 012 014 094 083 095 086 098 097 099 084 004 042
	[021] [079] 028 055 056 057 058 089 080 084 042
КВД	Метки полей БО <sub>А</sub>
931	[021] 047 024 025 001 026 011 085 087 203 043 045 004 041
117	[021] [018] [092] [040] [090] 001 012 014 094 083 095 043 004
118	[021] 047 024 025 001 026 011 085 087 203 043 004 041
119	[021] 001 213 214 043 004
120	[021] 047 024 025 001 026 011 085 087 203 043 212 060 215 216 004 041
634	[021] 047 024 025 001 026 011 085 087 203 043 045 004 041
733	[021] 047 024 025 001 026 011 022 [096] [007] 043 044 015 004 [084] [061] [017]

[...] — отмечены обязательные элементы БО.

Для документов вида 832 и 113 обязательным элементом является или поле с меткой 021, или поле с меткой 201.

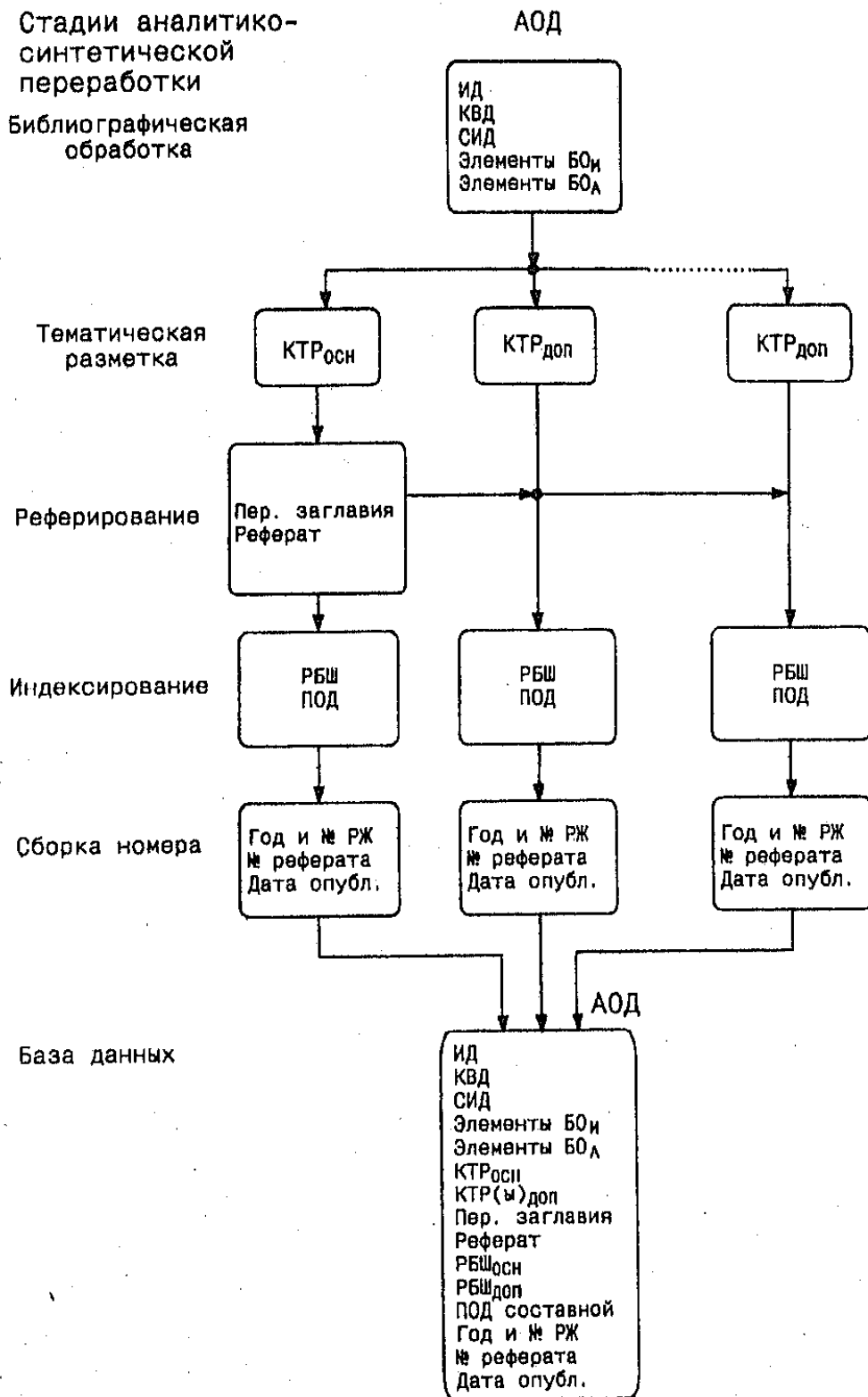


Рис. 3. Динамика структуры записи АОД в процессе аналитико-синтетической переработки документа. ИД — идентификатор документа; КВД — код вида документа; СИД — системный идентификатор; КТР — код тематической разметки; ПОД — поисковый образ документа

**Коды тематической разметки.** Каждый первичный документ может быть размечен специалистом только одним основным тематическим кодом (КТР<sub>осн</sub>) в профильный (основной) выпуск РЖ и факультативно несколькими дополнительными тематическими кодами (КТР<sub>доп</sub>) в непрофильные (дополнительные) выпуски РЖ. Значение основного тематического кода характеризует профильный тематический аспект данного документа и одновременно задает документу техно-

логический маршрут для основной содержательной обработки и производства профильного РЖ. Факультативная разметка характеризует наличие дополнительных тематических аспектов данного документа и задает дополнительные технологические адреса для содержательной обработки и производства непрофильных РЖ. В системе поддерживается актуальный список кодов тематической разметки. Этот список включает все коды выпусков РЖ, которые регламентно производятся в системе.

**Поля реферативного описания.** Каждая запись аналитического описания документа может иметь единственный набор элементов реферативного описания, который составляют только при основной содержательной обработке документа (КРЖ<sub>ОСН</sub>). В общем случае полный состав элементов реферативного описания включает: перевод оригинального заглавия документа, текст реферата.

**Поля индексов.** Каждая запись может иметь только один основной набор индексов (поисковый образ документа и рубрикационные шифры) и факультативно — более одного дополнительного набора индексов.

Рубрикационные шифры (РБШ) присваиваются документу в процессе индексирования рубриками по Рубриктору ВИНТИ. Индексирование документа рубрикационными шифрами осуществляется независимо для каждого тематического аспекта в ходе основной или дополнительной обработки. Если для данного тематического аспекта документ индексирован несколькими рубрикационными шифрами, то первый из них по порядку — главный, а остальные — второстепенные. Главный рубрикационный шифр используется в системе при компьютерной сборке профильного тематического номера выпуска РЖ в качестве критерия для опубликования реферата в соответствующем разделе РЖ, а второстепенные рубрикационные шифры служат для организации полиграфических ссылок на этот реферат из других разделов РЖ. При компьютерной сборке номеров сводных томов РЖ, а также при формировании БД, все рубрикационные шифры, которыми индексирован документ, автоматически объединяются в единый сводный набор.

Каждый поисковый образ документа (ПОД), в соответствии с внутренним форматом ВИНТИ, представляет собой упорядоченный набор связанных пар ключевых слов и их буквенных меток. В ходе основной переработки документа составляют ПОД<sub>ОСН</sub> в соответствии с основным тематическим аспектом документа. ПОД<sub>доп</sub> является факультативным элементом и составляется в соответствии с тем или иным дополнительным тематическим аспектом/ами. В случае отсутствия ПОД<sub>доп</sub> система автоматически использует для данного тематического аспекта значение ПОД<sub>ОСН</sub> (по умолчанию).

**Поля связи с номером выпуска РЖ.** Для автоматического вовлечения аналитического описания документа в технологические процессы подготовки регламентных выпусков РЖ система использует специальный набор полей в структуре записи — поля связи с номером выпуска РЖ.

Каждая запись может участвовать в подготовке нескольких тематических выпусков РЖ в зависимости от количества кодов тематической разметки в ее составе. В ходе компьютерной сборки номеров выпусков РЖ система автоматически заполняет соответствующие поля связи набором значения (в том числе год и № выпуска РЖ, № реферата и дата опубликования документа). Заполнение полей связи служит внутрисистемным критерием для завершения циклов тематической переработки документа. Полное заполнение всех полей связи по всем кодам тематической разметки означает, что документ полностью завершил свой жизненный цикл и подлежит архивации.

**Поля технологического контроля.** Поля технологического контроля записи аналитического описания документа содержат даты клавиатурного ввода и корректировки документа, коды операторов и др. сведения, используемые в системе для решения задач контроля и управления технологическим процессом. Общая логическая структура записи в привязке отдельных структурных частей к соответствующим стадиям переработки документа представлена на рис. 3.

## Пользователи технологической БД

Пользователями технологической БД являются научные и технические сотрудники, которые непосредственно участвуют в автоматизированной переработке документальных потоков и подготовке выходных продуктов. Взаимодействие пользователей с технологической БД осуществляется посредством автоматизированных рабочих мест (АРМ ввода библиографического описания, АРМ ввода содержательного описания, АРМ референта, АРМ редактора и др.), которые реализуются на основе персонального компьютера и объединены в локальную сеть.

Работу пользователей технологической БД регулирует администратор технологической БД на специализированном АРМ.

## ТЕХНОЛОГИЧЕСКИЕ СОСТОЯНИЯ ДАННЫХ

Процессы переработки документальных потоков в системе протекают во времени и находят свое отражение в смене технологических состояний записей аналитического описания документа в пределах их жизненного цикла.

**Жизненный цикл данных.** Жизненный цикл записей аналитического описания документа начинается на стадии библиографической обработки документа с момента клавиатурного ввода нового экземпляра с соответствующими кодами тематической разметки. Далее, по технологическим правилам, принятым в системе, каждая запись проходит все стадии содержательной обработки — независимо и асинхронно по каждому своему тематическому коду. Это означает, что, начиная со стадии содержательной обработки, жизненный цикл аналитического описания документа может развиваться параллельно по нескольким тематическим направлениям, одно из которых — профильное (основное), а остальные — непрофильные (дополнительные).

В ходе основной содержательной обработки запись снабжается такими элементами описания, как перевод заглавия документа, реферат, ПОД и рубрикационные шифры. Завершается цикл основной обработки опубликованием документа в регламентном выпуске профильного РЖ. В ходе дополнительной содержательной обработки запись аналитического описания документа может лишь дополняться элементами ПОДа и рубрикационными шифрами. Завершается цикл дополнительной обработки опубликованием документа в регламентном выпуске непрофильного РЖ. Дополнительная обработка может завершаться и преждевременно в результате исключения того или иного "избыточного" тематического кода разметки из записи

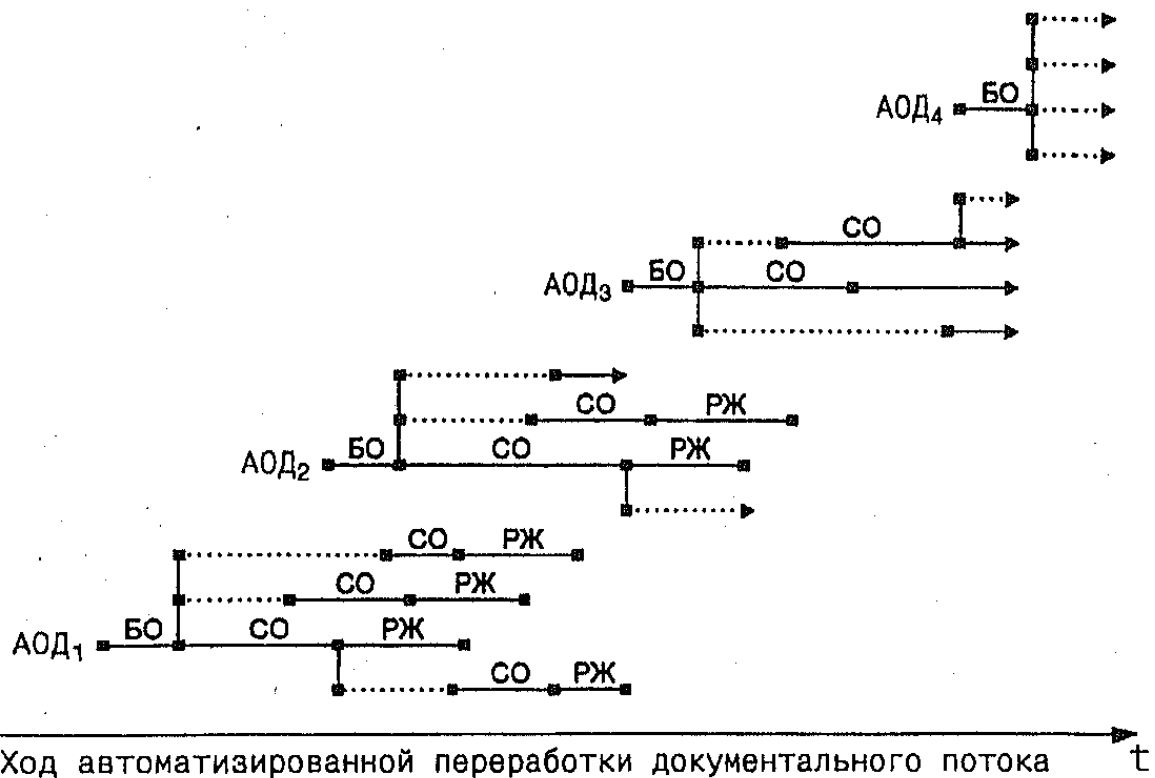


Рис. 4. Типовые примеры развития жизненных циклов записей АОД в ходе автоматизированной переработки документов: АОД<sub>1</sub> — полностью завершённый жизненный цикл; АОД<sub>2</sub> — частично завершённый жизненный цикл; АОД<sub>3</sub> — незавершённый жизненный цикл; АОД<sub>4</sub> — начальное состояние жизненного цикла. БО — библиографическая обработка и разметка; СО — содержательная обработка; РЖ — подготовка к изданию выпуска РЖ.

(см. рис. 3). В ходе основной или дополнительной содержательной обработки в любом документе могут быть выявлены новые тематические аспекты, не включенные в первоначальную разметку. В этих случаях технологические правила, принятые в системе, допускают частичную переметку записей на промежуточных стадиях переработки.

Таким образом, отличительной особенностью жизненного цикла записи является то, что он развивается во времени по принципу "ветвления" на отдельные тематические жизненные циклы соответственно тематическим кодам, присваиваемым документу как в начале, так и в процессе переработки. Причем в каждый момент времени различные записи могут находиться на разных стадиях жизненных циклов. Критерием полного завершения жизненного цикла аналитического описания документа служит факт опубликования документа в профильном и во всех непрофильных выпусках РЖ, соответствующих его кодам тематической разметки. Помимо этого в системе принято принудительное ограничение на продолжительность жизненного цикла аналитического описания документа, соответствующему которому после трех месяцев с момента опубликования по основному коду у документа автоматически исключаются все его незавершенные обработкой тематические коды.

Типовые примеры развития во времени жизненных циклов АОД представлены на рис. 4 в виде условных временных диаграмм.

Таким образом, текущее технологическое состояние каждой записи аналитического описания документа зависит от стадии и результатов переработки документа в системе. Запись может одновременно находиться в различных технологических состо-

яниях по каждому своему коду тематической разметки (по профильному и факультативным кодам).

**Индикатор и виды технологических состояний.** С целью контроля динамики технологических состояний в состав каждой записи аналитического описания документа включен специальный элемент — индикатор технологических состояний, который автоматически меняет свои значения при смене текущих состояний записи. Значениями индикатора служат специальные внутрисистемные коды, дискретный набор которых в совокупности составляет "шкалу технологических состояний". Автоматическая идентификация текущих состояний аналитического описания документа осуществляется по формальному критерию — "формуле" состояния. Этот критерий использует специальный алгоритм контроля за текущими значениями полей данных в каждой записи.

Все значения индикатора подразделяются на три основных вида: рабочие, выходные и аварийные. Рабочие (корректные) состояния имеют те записи аналитического описания документа, которые находятся на той или иной промежуточной (незавершенной) стадии обработки и имеют корректные значения полей данных, соответствующие требованиям внутреннего формата. Выходные состояния имеют записи, которые частично или полностью завершили свой жизненный цикл. Аварийные состояния имеют те записи, у которых выявлены недопустимые отклонения от требований, установленных в системе внутренним форматом (отсутствуют значения обязательных полей для данной стадии переработки, ошибки в значениях нормированных полей и т. п.). Каждое состояние характеризуется кодом, названием и формулой — формальным критерием для его идентификации. Шка-

ла рабочих, выходных и аварийных технологических состояний записей аналитического описания документа представлена в табл. 6.

Таблица 6

**Шкала технологических состояний записи аналитического описания документа**

Код	Наименование и формула состояния
<b>Рабочие состояния:</b>	
W1	Неполное библиографическое описание: ИД + СИД + БО <sub>А</sub> - БО <sub>М</sub>
W2	Библиографическая обработка завершена: W1 + БО <sub>М</sub> или ИД + БО
W3	Основная обработка завершена: W2 + КТР <sub>осн</sub> РЕФ + ПОД <sub>осн</sub> + РБК <sub>осн</sub>
W4	Дополнительная обработка завершена: W2 + КТР <sub>доп</sub> РЕФ + ПОД <sub>доп</sub> + РБК <sub>доп</sub>
...	... /в работе/
<b>Выходные состояния:</b>	
T1	Опубликован в профильном РЖ: ... /в работе/
T2	Опубликован в непрофильном РЖ: ... /в работе/
T3	Полностью опубликован в РЖ: ... /в работе/
...	...
<b>Аварийные технологические состояния АОД</b>	
E1	Отсутствует код профильной тематической разметки: - КТР <sub>осн</sub>
E2	Отсутствует СИД: W1-СИД
E3	Отсутствуют обязательные поля библиографич. описания: W2-БО <sub>овяз</sub>
E4	Отсутствуют обязательные поля основного описания: W3-РЕФ или W3-ПОД <sub>осн</sub> или W3-РБК <sub>осн</sub>
E5	Отсутствуют обязательные поля дополнит. описания: W4-ПОД <sub>доп</sub> или W4-РБК <sub>доп</sub>
...	... /в работе/

**ТЕХНОЛОГИЧЕСКАЯ СХЕМА ПЕРЕРАБОТКИ ДОКУМЕНТОВ**

Полный технологический цикл переработки информации характеризуется как единый поэтапный автоматизированный процесс "конвейерного" типа, обеспечивающий строгую последовательность всех операций творческого и рутинного типов от момента поступления первоисточников до создания выходных продуктов (рис. 5).

**Организация поступлений документальных источников.** Документы из внешних источников по тематике наук о жизни поступают в диспетчерскую службу системы с участка "Вход" ВИНТИ в виде макрокопий первичных документов, снабженные регистрационно-учетными сведениями. При этом все макрокопии документов, являющихся составными частями номеров выпусков серийных изданий, поступают со специальным кодом системного идентификатора документа, который используется в системе в качестве ключа до-

ступа к регистрационно-библиографическим записям в Электронном каталоге ВИНТИ. Средний объем ежедневной порции поступающих в систему документов составляет 1000 макрокопий. После регистрации и учета поступившие документы переходят в группу тематической разметки.

**Тематическая разметка документов.**

Первичная тематическая разметка документов выполняется научными сотрудниками, имеющими специальную профессиональную подготовку и владеющими иностранными языками и проблематикой по тематическому профилю системы. В соответствии с технологическими правилами, принятыми в системе, в качестве кодов разметки используются буквенно-цифровые индексы тематических выпусков (табл. 2). 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. Технологическое оформление решения о разметке реализуется вручную простановкой кода непосредственно на макрокопии документа. Первичная тематическая разметка документа осуществляется по принципу "максимально точного" основного кода и "избыточной" дополнительной разметки. Такой подход обеспечивает максимальное соответствие содержания многоаспектных документов потребностям внешних пользователей системы как по профильной, так и по непрофильной тематикам. При этом избыточные коды, присвоенные первоначально документу, могут как исключаться, так и дополняться на последующих стадиях переработки документа решением научных редакторов или референтов, более компетентных в узких тематических вопросах.

**Библиографическая обработка и прямой ввод данных.**

Макрокопии первичных документов с кодами тематической разметки поступают ежедневными порциями на участок библиографической обработки. Принятая в системе технология библиографической обработки документов реализует принцип интеграции функций интеллектуальной обработки и клавиатурного ввода данных, получивший в специальной литературе наименование "прямой ввод данных" [14]. Такая технология позволяет отказаться от использования предмашинных формуляров на стадии ввода библиографического описания, повышает оперативность обработки за счет сокращения промежуточных операций и одновременно и однократного чтения документа. Вместе с тем повышаются требования к персоналу системы, осуществляющему прямой ввод данных "с листа", который должен иметь необходимую профессиональную подготовку и знать внутренний формат данных.

В системе принят формат данных для библиографического описания документов, который полностью соответствует единому внутреннему формату действующей в ВИНТИ технологии и жестко устанавливает состав, обозначения, наименования и форматы всех полей данных, используемых при библиографическом описании документов разных видов (см. табл. 4). В ходе прямого ввода создаются записи, которым система автоматически присваивает уникальные машинные идентифи-

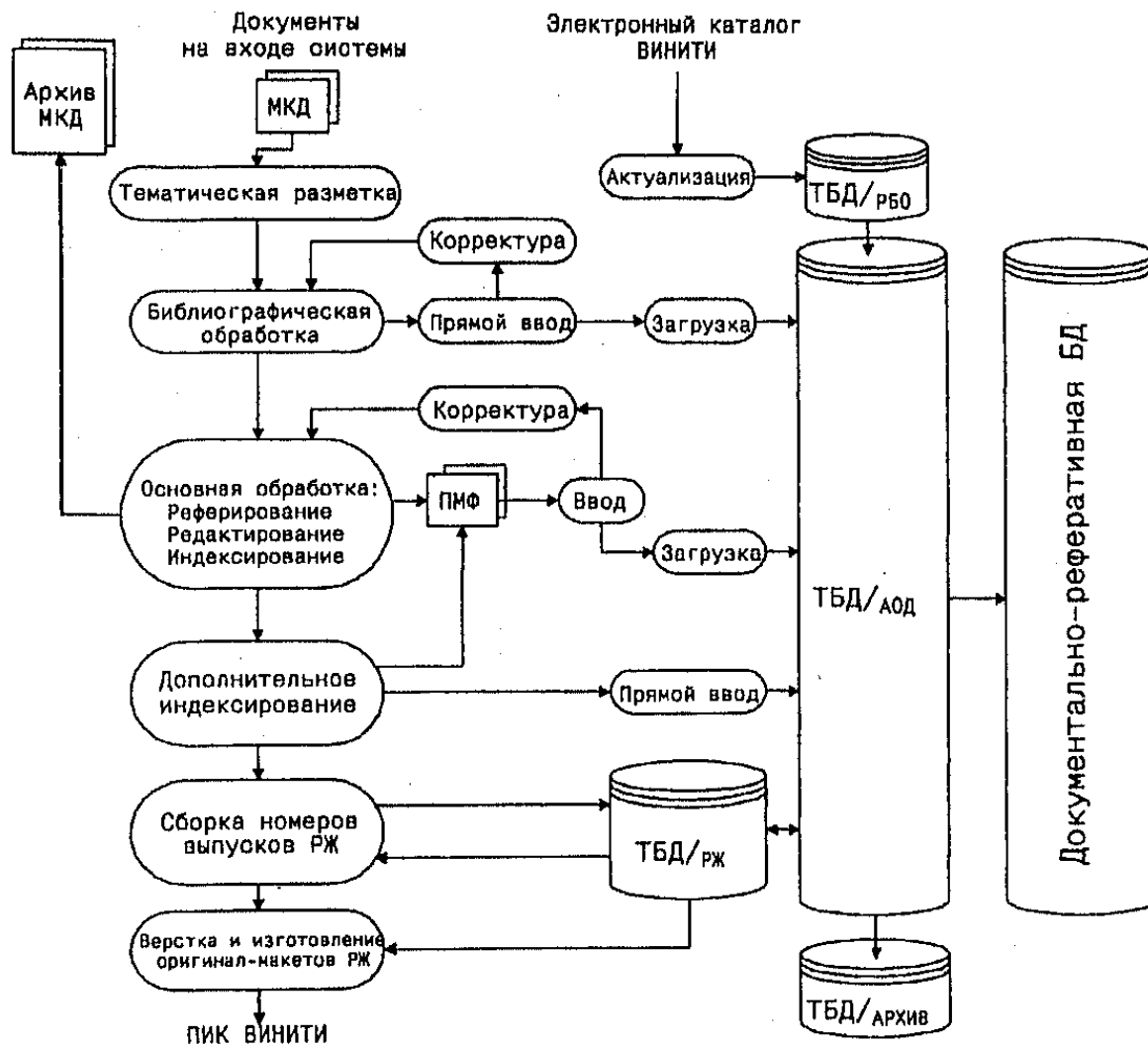


Рис. 5. Технологическая схема автоматизированной переработки потока документов в Отделении научной информации ВИНТИ по проблемам наук о жизни. МКД — макрокопии первичных документов; ПМФ — предмашинный формуляр.

каторы, и которые аккумулируются ежедневными порциями в рабочих массивах на АРМ исполнителей. Эти записи заполняются библиографическими сведениями о документе, общими сведениями о документе и учетно-технологическими данными. Одновременно с библиографическими полями вводятся коды тематической разметки и системный идентификатор.

В ходе библиографической обработки макрокопии первичных документов вместе с распечатками результатов прямого ввода на специальном формуляре поступают ежедневными порциями в группу корректоров. После корректуры файлы с ежедневными порциями записей передаются для загрузки. На этой технологической стадии формируется библиографическая БД. Макрокопии документов, снабженные особым формуляром с распечатками БО, комплектуются в профильные тематические порции и распределяются в соответствии с основным кодом тематической разметки по научным редакторам, ответственным за содержательную обработку документов.

#### Реферирование и индексирование документов.

Макрокопии документов, прошедших цикл библиографической обработки, поступают ежедневными порциями к научным редакторам, которые организуют их реферативную обработку. Аналитическое реферирование и индексирование документов по профильному тематическому аспекту осуществляют специалисты-эксперты. Редактор вы-

полняет научное и литературное редактирование рефератов, индексирование рубрикационными шифрами и ключевыми словами. Результаты обработки, а также финансовые сведения, заносятся в специальный предмашинный формуляр, на котором уже имеется распечатка БО и идентификатор. Таким образом этот формуляр преобразуется в предмашинный формуляр для клавиатурного набора содержательного описания документов.

Автоматизированная технология подготовки РЖ предусматривает возможность создания реферата на АРМ референта с последующим редактированием и индексированием на АРМ редактора, минуя стадию производственного набора содержательных описаний.

Все содержательные описания (СО) документов на машиночитаемых носителях, поступающие ежедневными порциями из ПИК ВИНТИ и с АРМ редактора, загружаются в массив технологической БД, где они автоматически объединяются с соответствующими библиографическими записями по идентификатору и становятся доступными для дальнейшей многоаспектной и многоцелевой автоматизированной обработки.

Дополнительное индексирование рубрикационными шифрами и ключевыми словами документов, имеющих дополнительные коды тематической разметки, по каждому непрофильному тематическому аспекту осуществляет научный редактор соответствующего тематического направления. В системе предусмотрены два варианта дополнительного

индексирования: посредством внесения дополнений в предмашиный формуляр, который заполняется профильным редактором для производственного клавиатурного набора в ПИК ВИНТИ, и в режиме прямого ввода значений индексов непосредственно в записи массива технологической БД на АРМ редактора.

#### Сборка регламентных выпусков РЖ.

Текущий состав массива технологической БД с записями аналитических описаний документов служит основой для подготовки регламентных выпусков РЖ. Каждый очередной номер выпуска РЖ готовит соответствующий профильный научный редактор на специализированном АРМ редактора в отведенные регламентные сроки, используя автоматизированную технологию, получившую название "техническое макетирование РЖ". В ходе компьютерной сборки номера выпуска РЖ редактор имеет возможность отбирать по запросу записи, размеченные тематическим кодом данного выпуска РЖ, просматривать содержимое рабочего массива на экране дисплея, осуществлять необходимые операции и принимать решение о включении тех или иных записей в данный номер выпуска РЖ. По окончании цикла технического макетирования

редактор направляет файл, содержащий сформированный номер выпуска РЖ, на участок верстки и лазерного изготовления оригинал-макета РЖ.

#### Формирование БД.

Одной из основных регламентных технологических операций в системе является ежемесячная подготовка файлов БД, содержащих записи, полностью переработанные и опубликованные в выпусках РЖ, и их передача в АБНД ВИНТИ.

#### Ведение технологического архива записей.

Записи, прошедшие в системе полный жизненный цикл, подлежат удалению из рабочего массива технологической БД и передаются в специальный технологический архив вторичных документов — массив. Технологический архив предназначен для решения разнообразных задач по ретроспективному поиску записей.

### АРХИТЕКТУРА АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ

Общая архитектура основных автоматизированных звеньев АИТС "Биология", реализованная в Отделении научной информации ВИНТИ по проблемам наук о жизни, представлена на рис. 6.

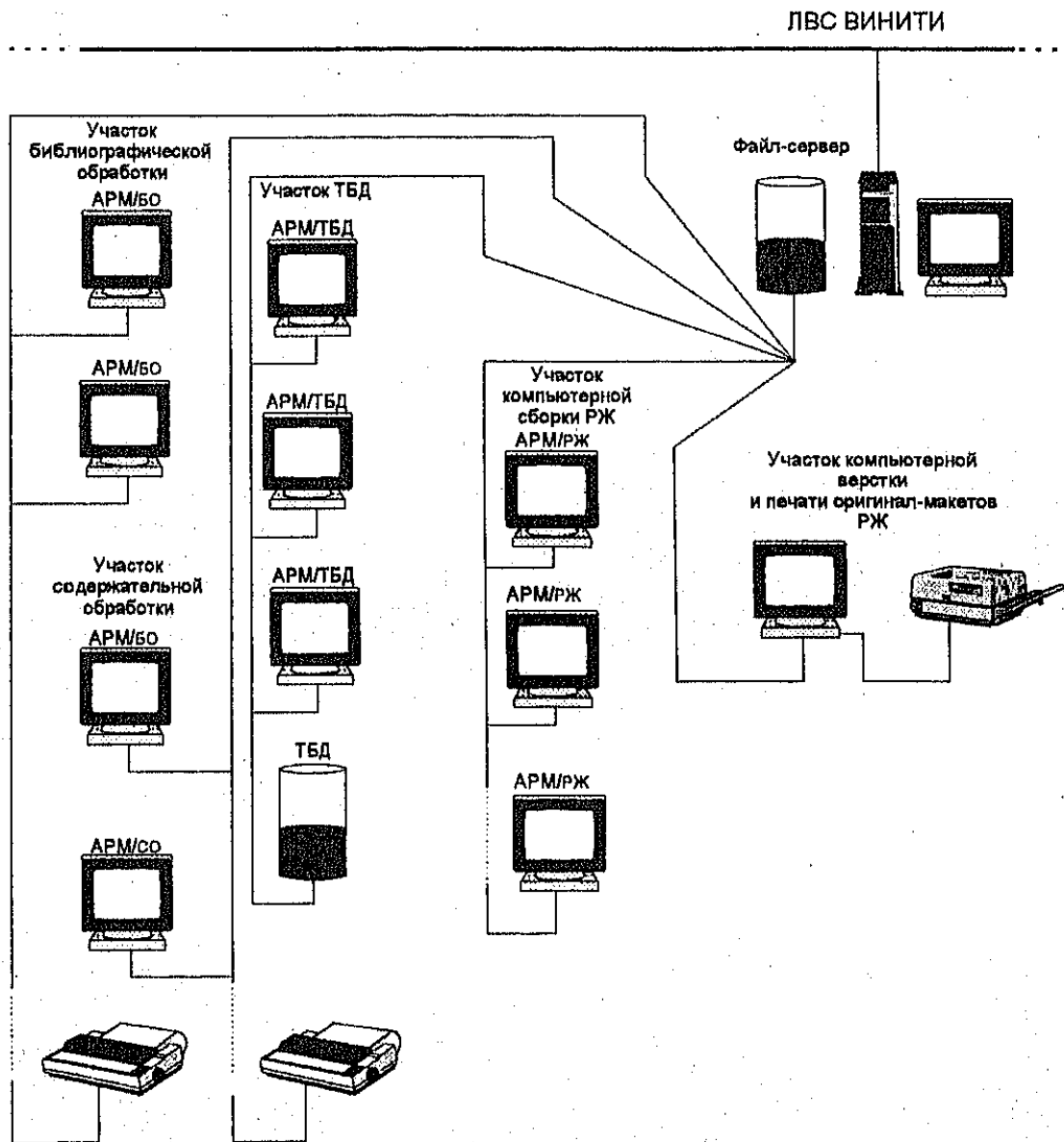


Рис. 6. Архитектура автоматизированной информационно-технологической системы по наукам о жизни. АРМ/БО — автоматизированное рабочее место библиографической обработки и прямого ввода данных; АРМ/ТБД — автоматизированное рабочее место оператора ТБД; АРМ/РЖ — автоматизированное рабочее место редактора РЖ

Система построена на IBM-совместимых ПК (IBM-386 и IBM-486)\*, функционирующих в составе локальной вычислительной сети ВИНТИ в операционных средах MS DOS и Novell NetWare. Все прикладное программное обеспечение системы разработано в Отделе программных систем ВИНТИ с использованием СУБД "CDS/ISIS 3.0", СУБД "Advanced Revelation 1.1 и 1.3", языка программирования "C++" и издательского пакета "Ventura Published".

Основу архитектуры составляет файл-сервер, который подключен непосредственно к основной информационной магистрали ЛВС ВИНТИ и обеспечивает циркуляцию информационных потоков внутри системы и сопряжение с другими автоматизированными подсистемами и службами ВИНТИ. Персонал системы работает на специализированных АРМ, подключенных к общему файл-серверу системы. Схемы подключения АРМ реализуют сетевую топологию типа "звезды" на выделенных сегментах локальной сети с использованием стандарта "EtherNet". Каждый сегмент объединяет группу автоматизированных рабочих мест, функционирующих на одном из технологических участков системы. Все обмены файлами данных по сети жестко регламентированы технологической схемой переработки документальных пото-

ков в системе, осуществляются через общий файл-сервер и используют специально разработанный внутрисистемный коммуникативный формат.

В системе выделены следующие функционально специализированные участки: библиографической обработки, ввода содержательных описаний (функционирует в ВИНТИ, ПИК ВИНТИ и непосредственно у референтов); технологический БД, компьютерной сборки РЖ, компьютерной верстки и лазерного изготовления оригинал-макетов РЖ, — структура которых базируется на специализированных автоматизированных рабочих местах. Подробнее указанные функциональные участки и соответствующие АРМ будут рассмотрены в последующих публикациях.

### СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА АИТС "БИОЛОГИЯ" И "АССИСТЕНТ 3"

Как было отмечено, в основу АИТС "Биология" заложена система "Ассистент 3", но в ходе разработки АИТС "Биология" были внесены существенные дополнения и изменения. Результаты сравнительного анализа этих модификаций отражены в табл. 7.

Таблица 7

#### Основные принципы функционирования автоматизированной информационно-технологической системы по наукам о жизни (АИТС "Биология") в сравнении с интегральной информационной системой "Ассистент 3"

"Ассистент 3"	АИТС "Биология"
<b>Комплектование входного потока</b>	
Исчерпывающий сбор и комплектование всех опубликованных в мире научных документов по естественным и техническим наукам.	Исчерпывающий сбор и комплектование всех опубликованных в мире научных документов по тематическому профилю наук о жизни.
<b>Аналитико-синтетическая переработка</b>	
Одноразовая и исчерпывающая аналитико-синтетическая переработка каждого документа, осуществляемая последовательно: сначала по основному тематическому аспекту, а затем факультативно по дополнительным тематическим аспектам. Формирование по результатам переработки полного аналитического описания на каждый документ, включая обязательное основное описание и факультативные дополнения.	Однократная и исчерпывающая аналитико-синтетическая переработка каждого документа, осуществляемая параллельно: по основному тематическому аспекту и факультативно по дополнительным тематическим аспектам. Формирование по результатам переработки полного аналитического описания на каждый документ, включая обязательное основное описание и факультативные дополнения.
<b>Формирование технологического массива данных</b>	
Формирование и ведение единого рабочего массива, в котором хранятся в единственном экземпляре аналитические описания на каждый переработанный документ — записи АОД; однократный ввод в рабочий массив всех элементов АОД после завершения основной обработки; многократный ввод дополнений к записям АОД по мере завершения дополнительных обработок.	Формирование и ведение единой технологической базы данных, в которой хранятся в единственном экземпляре текущие аналитические описания на каждый перерабатываемый документ — записи АОД; незамедлительный и поэтапный ввод в технологическую базу элементов АОД по мере осуществления последовательных стадий основной и/или дополнительной обработки документа.

\* В условиях резкой ограниченности бюджетного финансирования и выделения средств на техническое обеспечение, вычислительная техника приобреталась в основном за счет прибыли от договорной деятельности. В перспективе, при улучшении финансовой ситуации, планируется осуществить переход к компьютерам класса Pentium.



"Ассистент 3"	АИТС "Биология"
<b>Библиографическая обработка</b>	
Одноразовое составление библиографического описания на каждый документ; отложенный ввод элементов библиографического описания в рабочий массив после завершения всех стадий основной обработки (реферирование, индексирование и др.).	Опережающая и однократная библиографическая обработка каждого документа; незамедлительный ввод элементов библиографического описания в соответствующие записи АОД технологической базы данных.
<b>Тематическая разметка</b>	
Разметка каждого первичного документа кодами рубрик Рубрикатора ВИНТИ в соответствии с выявленными в документе тематическими аспектами, а также кодом профильного подразделения ВИНТИ, где должна производиться основная обработка документа, и кодами непрофильных подразделений, где должна производиться его дополнительная обработка.	Разметка каждого первичного документа кодами выпусков РЖ по Рубрикатору РЖ ВИНТИ в соответствии с выявленными в документе тематическими аспектами, включая обязательный основной код и факультативные дополнительные коды; незамедлительный ввод кодов тематической разметки в соответствующие записи АОД технологической базы данных.
<b>Реферирование и научное редактирование</b>	
Одноразовое составление реферата на каждый документ в ходе основной обработки; научное и литературное редактирование реферата; отложенный ввод текста реферата в рабочий массив после завершения всех стадий основной обработки.	Однократное реферирование каждого документа в ходе основной обработки, научное и литературное редактирование реферата; незамедлительный ввод результатов реферирования в соответствующие записи АОД технологической базы данных.
<b>Индексирование ключевыми словами</b>	
Одноразовое координатное индексирование каждого документа дескрипторами и ключевыми словами в ходе основной обработки; многократное координатное индексирование документов в ходе дополнительных обработок; отложенный ввод ПОД-ов в рабочий массив после завершения всех стадий основной и/или дополнительной обработки.	Одноразовое координатное индексирование каждого документа дескрипторами и ключевыми словами в ходе основной обработки; многократное координатное индексирование документов в ходе дополнительных обработок; незамедлительный ввод ПОД-ов в соответствующие записи АОД технологической базы данных.
<b>Индексирование по Рубрикатору</b>	
Одноразовое индексирование каждого документа по Рубрикатору ВИНТИ в ходе основной обработки; многократное индексирование документов в ходе дополнительных обработок; отложенный ввод рубрикативных кодов в рабочий массив после завершения всех стадий основной и/или дополнительной обработки.	Одноразовое индексирование каждого документа по Рубрикатору ВИНТИ в ходе основной обработки; многократное индексирование документов в ходе дополнительных обработок; незамедлительный ввод рубрикативных кодов в соответствующие записи АОД технологической базы данных.
<b>Подготовка к выпуску информационных изданий</b>	
Многократное использование записей АОД, введенных в единый рабочий массив, для формирования различных тематических выпусков РЖ и указателей к ним, а также других информационных изданий по разным отраслям естественных и технических наук, предметам и межотраслевым комплексным проблемам.	Многократное использование записей АОД, введенных в единую технологическую базу, для формирования различных тематических выпусков РЖ и указателей к ним, а также других информационных изданий по наукам о жизни.
<b>Лингвистическое обеспечение</b>	
Централизованное лингвистическое обеспечение, основанное на ведении Рубрикатора ГАСНТИ, Рубрикатора ВИНТИ, Рубрикатора РЖ ВИНТИ, словарей используемых ключевых слов и дескрипторов, таблиц УДК и МКИ, регулярных сокращений и аббревиатур, применяемых имен и названий, терминологических словарей.	Централизованное лингвистическое обеспечение, основанное на ведении тематического Рубрикатора ВИНТИ по наукам о жизни, системы Рубрикации всех выпусков РЖ, входящих в сводный том РЖ "Биология", словарей используемых ключевых слов и дескрипторов, регулярных сокращений и аббревиатур, терминологических словарей.

"Ассистент 3"	АИТС "Биология"
<b>Нормативно-технологическое обеспечение</b>	
Единые нормативно-технологические предписания для всех стадий, технологических операций и персонала системы; полная технологическая совместимость всех функциональных подсистем и звеньев системы.	Единые нормативно-технологические предписания для всех стадий, технологических операций и персонала системы; полная технологическая совместимость всех функциональных подсистем и звеньев системы.

**Уровень автоматизации технологических процессов**

<p>Ручное выполнение всех стадий аналитико-синтетической переработки документов штатными сотрудниками ВИНТИ и нештатными учеными и специалистами; ручная подготовка всех результатов переработки к преобразованию в машиночитаемую форму и однократный ручной ввод записей АОД в единый рабочий массив; сочетание ручной и автоматизированной корректуры записей АОД; автоматизированная подготовка машиночитаемых РЖ на базе единого рабочего массива с использованием автоматизированных рабочих мест для редакторов; автоматическое лазерное изготовление оригинал-макетов информационных изданий для последующей офсетной печати тиража.</p>	<p>Сочетание ручных и автоматизированных операций при аналитико-синтетической переработке документов штатными сотрудниками ВИНТИ и нештатными учеными и специалистами; ручная подготовка части результатов переработки к преобразованию в машиночитаемую форму и поэтапный ручной ввод записей АОД в единую технологическую базу данных; сочетание ручной и автоматизированной корректуры записей АОД; автоматизированная подготовка машиночитаемых РЖ по текущему содержанию технологической базы данных с использованием автоматизированных рабочих мест для редакторов; автоматическое лазерное изготовление оригинал-макетов информационных изданий для последующей офсетной печати тиража.</p>
--	---

В отличие от системы "Ассистент 3" в АИТС "Биология" единственным документом, регламентирующим разметку и диспетчеризацию документальных потоков, является Рубрикатор РЖ ВИНТИ, а в качестве кодов разметки используются индексы тематических выпусков РЖ. За счет этого совмещаются тематический и технологический аспекты разметки, так как кодировка выпусков РЖ однозначно определяет тематические профили для переработки документов и одновременно служит основой для комплектования порций документов и их адресной маршрутизации. Механизм ведения и актуализации Рубрикатора РЖ, принятый в ВИНТИ, гарантирует устойчивость и актуальность классификационной схемы разметки, а также адресность технологических маршрутов. Последнее связано с тем, что в системе документы комплектуются и рассылаются на переработку непосредственно редакторам, ответственным за тот или иной тематический выпуск РЖ. Требования к персоналу, осуществляющему разметку, сводятся к укрупненной многоаспектной экспертной оценке тематического содержания документа, выражаемой в кодах РЖ. При этом возможная неточность разметки частично компенсируется принципом ее "избыточности", когда каждый документ может быть размечен несколькими кодами РЖ, а также возможностью изменения кодов тематической разметки в ходе дальнейшей содержательной обработки документа.

Существенное различие в принципах функционирования АИТС "Биология" и системы "Ассистент 3" касается ввода данных в рабочий массив системы. В системе "Ассистент 3" предусмотрен однократный отложенный ввод в ЭВМ результатов аналитико-синтетической переработки,

в то время как в АИТС "Биология" ввод результатов переработки организуется по мере осуществления последовательных стадий аналитико-синтетической переработки. Сначала вводятся библиографические описания и коды тематической разметки. Это обуславливает возможность параллельной многоаспектной автоматизированной обработки документов по каждому тематическому коду. Далее, по завершении каждого этапа переработки, незамедлительно вводятся элементы содержательного описания (реферат, ключевые слова, рубрикационные шифры и т. п.), которые автоматически объединяются с элементами библиографического описания в едином аналитическом описании документа. Это дает возможность поэтапно, а также многократно использовать результаты переработки документа для формирования различных выходных информационных продуктов.

#### **РЕЗУЛЬТАТЫ ОПЫТНО-ПРОМЫШЛЕННОЙ ЭКСПЛУАТАЦИИ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ПЕРЕРАБОТКИ ДОКУМЕНТОВ**

Данная технология апробирована в режиме опытно-промышленной эксплуатации в 1996 г. на 100% потоке документов по проблемам наук о жизни при подготовке 12 номеров БД и 55 выпусков сводного тома "Биология" с пономерными указателями к ним. Разработка внедрена в технологический цикл ВИНТИ на полном потоке одного направления и охватывает все этапы подготовки информационных изданий.

Опытно-промышленный процесс эксплуатации автоматизированной системы переработки доку-

ментов в Отделении научной информации по проблемам наук о жизни характеризуется следующими показателями:

- тематическая и библиографическая обработка ксерокопий, набор и корректура БО 260 067 док.
- творческая содержательная переработка документов по основному коду тематической разметки (реферирование и индексирование) 214 931 док.
- обработка рефератов, поступающих от референтов на машиночитаемых носителях 431 док.
- обработка содержательных описаний, поступающих из ПИК ВИНТИ 207 500 док.
- обработка содержательных описаний, поступающих с участка ввода содержательных описаний ВИНТИ 7 000 док.
- загрузка массивов БО и содержательных описаний документов в технологическую БД 473 337 записей
- загрузка описаний из Электронного каталога первоисточников ВИНТИ 18 400 записей
- переметка кодов тематической разметки описаний документов в технологической БД 14 500 ОД
- ведение/пополнение системы рубрикации регламентных выпусков РЖ 55 вып.
- работа на центральной станции системы по автоматизированному поиску и отбору порций описания документов в технологической БД и формированию очередных выпусков РЖ 1 150 часов
- работа научных редакторов по компьютерной сборке РЖ на АРМ редактора в локальной сети системы 7425 часов
- верстка оригинал-макетов изданий 660 изданий
- подготовка оригинал-макетов выпусков РЖ 660 изданий
- передача файлов с финансовыми сведениями в систему "Гонорар" ВИНТИ 660 файлов
- передача файлов БД в АБД ВИНТИ 660 разделов БД

Поскольку срок прохождения документа в системе составляет 2–2,5 мес., не все документы, прошедшие библиографическую обработку, завершили свой жизненный цикл в системе в этом же году (45 136 документов находятся в процессе содержательной обработки).

Результаты опытно-промышленной эксплуатации системы позволяют рассматривать компьютерную технологию подготовки информационных изданий по наукам о жизни как действующий образец будущей технологии обработки всего потока документов ВИНТИ. Экономические расчеты, проведенные совместно с Планово-финансовым отделом ВИНТИ, показывают эффективность промышленного внедрения подобной технологии, которая раскрывает широкие возможности по исключению нерентабельных дублирующихся процессов и наращиванию дополнительных возможностей, а также по созданию новых информационных продуктов и услуг, способствующих адаптации ВИНТИ к функционированию на данном нестабильном этапе экономической ситуации в стране.

#### СПИСОК ЛИТЕРАТУРЫ

1. Арапов М. В. Новые ниши на рынке информационных продуктов и услуг // НТИ. Сер. 2. — 1991. — № 2. — С. 1–6.
2. Арский Ю. М., Гиляревский Р. С., Туров И. С., Черный А. И. Инфосфера: Информационные структуры, системы и процессы в науке и обществе. — М.: ВИНТИ, 1996. — 489 с.
3. Борисова Л. Ф., Петров Р. В. Научно-информационное обеспечение фундаментальных и прикладных исследований по биологии // НТИ. Сер. 2. — 1993. — № 10. — С. 21.
4. Checkland P. B. Information systems and systems thinking: Time to unite? // Int. J. Inf. Manag. — 1988. V. 8. — № 4. — С. 239–248.
5. Hill Michael W. Information policies: premonitions and prospects // J. Inf. Sci. — 1995. V. 21. — № 4. — С. 272–282.
6. Черный А. И. Состояние разработки и перспективы внедрения системы "АССИСТЕНТ" как одного из звеньев сети автоматизированных информационных центров — М.: ВИНТИ, 1997.
7. Черный А. И., Высочанская О. А., Жильцова Л. А., Кузнецова Э. К. Автоматизированная информационная система ВИНТИ "АССИСТЕНТ-3". Общая концепция. Препринт. — М.: ВИНТИ, 1988. — 105 с.
8. Черный А. И. Техническое задание на разработку автоматизированной информационной системы ВИНТИ по опубликованным отечественным и зарубежным источникам по естественным и техническим наукам "АССИСТЕНТ-3/94". — М.: ВИНТИ, 1994. — 27 с.
9. Борисова Л. Ф. Создание автоматизированной системы производства информационных продуктов по биологии / Материалы конф. "НТИ-95". — М.: ВИНТИ, 1995.
10. Борисова Л. Ф., Скрылев К. М., Шапкин А. В. Автоматизированная информационно-технологическая система ВИНТИ по наукам о жизни (принципы построения и практика работы) / Материалы конф. "НТИ-96". — М.: ВИНТИ, 1996.
11. Разработка и развитие автоматизированной технологии производства информационных продуктов ВИНТИ. Отчет по НИР. № ГР 01930005454.