

**Антопольский  
Александр Борисович**

**Лингвистическое обеспечение  
электронных библиотек**

**Учебное пособие по специальности  
«Библиотечно-информационные системы и  
электронные библиотеки»**

Москва 2002

## Содержание

### Введение

#### **1. Теоретические основы лингвистического обеспечения**

- 1.1. Общая модель коммуникативного взаимодействия
- 1.2. Лингвистические аспекты коммуникации
- 1.3. Психологические аспекты коммуникации
- 1.4. Основные понятия семиотики
- 1.5. Основные понятия логики
- 1.6. Теория речевых актов
- 1.7. Пресуппозиция

Литература к главе 1

#### **2. Электронные библиотеки: основные понятия и технологии**

- 2.1. Классификация АИС
- 2.2. Электронные библиотеки как специфический вид АИС
- 2.3. Основные понятия информационного поиска
- 2.4. Пользовательские интерфейсы в электронных библиотеках
- 2.5. Языки разметки

Литература к главе 2

#### **3. Общие сведения о ЛО**

- 2.1. Определение понятия ЛО.
- 2.2. История разработки ЛО в России
- 2.3. Классификация средств ЛО

Литература к главе 3

#### **4. Системы метаданных**

- 4.1. Общие понятия систем метаданных
- 4.2. Языки библиографических данных
- 4.3. Основные системы метаданных
- 4.4.. Дублинское ядро метаданных
- 4.5. MARC - машиночитаемый каталог
- 4.6. Формат ONIX
- 4.7. Метаданные глобальной (правительственной) службы поиска информации
- 4.8. Организация деятельности по созданию метаданных

Литература к главе 4

#### **5. Классификационные языки**

- 5.1. Общие понятия классификации
- 5.2. Универсальная десятичная классификация
- 5.3. Государственный рубрикатор научно-технической информации
- 5.4. Проблемы и перспективы применения информационных классификаций

Литература к главе 5

#### **6. Вербальные языки**

- 6.1. Общее описание и история развития вербальных языков
- 6.2. Лексика и организация лексики в вербальных языках
- 6.3. Информационно-поисковый тезаурус

- 6.4. Принципы создания и ведения тезауруса УИС «Россия»
  - 6.5. Грамматика вербальных ИПЯ традиционных АИПС
  - 6.6. Методика индексирования средствами вербальных языков в традиционных АИПС
  - 6.7. Грамматики вербальных языков современных ЭБ
  - 6.8. Организация поиска с использованием вербальных ИПЯ
- Литература к главе 6

## **7. Лингвистическое обеспечение фактографических и комплексных систем**

- 7.1. Общие понятия фактографии.
  - 7.2. Интегрированные документально-фактографические АИС
  - 7.3. Комбинированные документально-фактографические системы
  - 7.4. Обработка запросов на естественном языке к фактографическим базам данных
  - 7.5. ЛО ситуационного управления
- Литература к главе 7

## **8. Автоматическая обработка текста**

- 8.1. Виды процессов автоматической обработки текста
- 8.2. Морфологический анализ
- 8.3. Синтаксический анализ
- 8.4. Позиционные методы анализа текста
- 8.5. Суперсинтаксический анализ
- 8.6. Семантический анализ
- 8.7. Статистические методы
- 8.8. Требования к автоматическому индексированию

## **9. Лингвистические банки данных и компьютерная лексикография**

- 9.1. Основные типы словарей в АИС
- 9.2. Примеры организации лингвистических банков данных
- 9.3. Обмен словарями и коммуникативные форматы словарей
- 9.4. Лингвистические банки данных в Интернет
- 9.5. Основы компьютерной лексикографии

**Заключение.**

**Приложение.** Термины и определения

## Введение

Настоящая монография посвящена проблемам создания и использования языковых средств такого класса автоматизированных информационных систем (АИС), которые в последние годы стали называть электронными библиотеками.

Актуальность данной задачи не вызывает сомнений. Бурное развитие информатики в последние годы, в том числе невероятно быстрое развитие Интернета породили массовое создание электронных информационных ресурсов по всем областям знания и в самых различных видах деятельности. Специалистам очевидна и необходимость организации этих ресурсов в систематически организованные массивы, в которых было бы удобно и эффективно проводить информационный поиск, в том числе неподготовленному пользователю. Очевидно также, что для этой цели нужны адекватные языковые средства. Интернет, который стал практически единой средой для создания и размещения информационных ресурсов и для поиска в них, с одной стороны, предоставляет множество дополнительных возможностей, с другой стороны, накладывает определенные ограничения на выбор и применение языковых средств. Направление информатики, посвященное проблемам организации электронных ресурсов в условиях глобальных сетей и организации поиска в них, стало в последние годы называться *электронными* или *цифровыми библиотеками*.

Поэтому не вызывает сомнений необходимость в исследованиях, обобщающих опыт работ по созданию таких средств, и ориентирующих разработчиков АИС при выборе языковых средств для электронных библиотек. Соответственно необходимо и вести обучение студентов, специализирующихся в этом направлении информатики.

В то же время подготовка такой монографии явилась очень сложной задачей, причем сложность обусловлена спецификой современного периода развития отечественной информатики.

Дело в том, что современный период последние 10-12 лет явились периодом кризиса отечественной науки вообще и информатики в частности. Если в течение 1970-1980-х гг. исследования в области информационных систем и информационного поиска велись широким фронтом и появлялось множество публикаций, посвященных этим проблемам и, в частности, лингвистическому обеспечению АИС, то с 1990 г. их число резко сократилось, а обобщающих монографий просто не было.

Отчасти это обусловлено тем, что центральным звеном этого направления информатики в СССР была сеть органов научно-технической информации, которая практически рухнула в результате всех событий 1990-х гг.

В настоящее время исследования и разработки в этой области ведутся в различных структурах – государственных и коммерческих, библиотечных и информационных, однако совершенно разрозненно и почти всюду – очень малыми силами.

В результате возник разрыв между достижениями отечественной

информатики прошлых десятилетий и современным состоянием теории и практики информационно-поисковых систем и электронных библиотек.

Автор занимается этой проблемой, можно сказать, с самого начала, с конца 1960-х гг и имел возможность воочию наблюдать все этапы эволюции этого направления информатики, как его подъемы, так и падения. При этом с удручающей регулярностью наблюдаются попытки наступать на одни и те же грабли и вновь изобретать велосипед, что объясняется исключительно незнанием уже давно пройденного пути.

Настоящая книга ставит задачей хотя бы частично заполнить разрыв между нынешним поколением молодых исследователей, прежде всего студентов и аспирантов, и изложить основы создания и использования лингвистического обеспечения АИС с учетом как прошлых достижений, так и сегодняшнего состояния в данной области.

Исходя из данного состояния дел, книга строится следующим образом. Вначале кратко излагаются базовые понятия тех теоретических дисциплин, которые, по мнению автора необходимо использовать при разработке лингвистического обеспечения. Это лингвистика, логика, психология, семиотика, теория коммуникации и др. Эти разделы ориентированы прежде всего на студенческую аудиторию.

Затем приводятся те сведения из информатики, которые являются наиболее актуальными для данной проблемы. При этом автор стремился избегать программно-технологических аспектов, сосредотачиваясь на чисто информационных аспектах проблем представления и поиска информации. Специальный раздел посвящен изложению взглядов автора на существо рассматриваемой дисциплины информатики – «электронные библиотеки».

Новым для работ в области ЛО АИС. является проблематика создания и применения языков разметки. По мнению автора, именно развитие этих языков во многом будет определять перспективы и возможности применения языковых средств в электронных библиотеках, подобно тому, как 20 лет назад возможности поисковых систем определялись теорией и практикой систем управления базами данных.

Начиная с 3-й главы, излагается собственно проблематика лингвистического обеспечения. Общий принцип при подготовке этих материалов был следующий: включать описания результатов тех разработок прошлых лет, которые автору представляются перспективными в настоящее время. В то же время общая структура достаточно традиционна: языки классификационные, вербальные и объектно-признаковые. Принципиально новым является раздел, посвященный системам метаданных. Автор стоит на позиции, что это модное понятие является некоторым обобщением и модернизацией давно известных языков библиографического описания документов.

Несколько слов о проблеме терминологии. К концу 1980-х гг., при участии автора был разработан ряд терминологических стандартов в данной области. Хотя в некоторых частностях система терминов и определений уже устарела, тем не менее, автор счел необходимым привести ее практически

полностью (Приложение), поскольку системность подхода в этих стандартах была достигнута весьма высокая, чего явно недостает терминологии, применяемой в современных работах.

Что же касается наименования основного объекта исследований то термин «электронная библиотека» (ЭБ) автор стремился использовать применительно к работам последних лет. При изложении же истории разработок лингвистического обеспечения прежних лет (до 1990 г.) в основном используется применявшийся в те годы более традиционный, хотя и не очень точный, термин «автоматизированная информационная система» (АИС). В некоторых случаях используется термин «автоматизированная информационно-поисковая система» (АИПС), который имеет более узкое значение. Специально рассмотрению этих понятий посвящено часть материалов главы 2.