

## Занятие №12

*Обработка* – понятие широкое, часто включает в себя несколько взаимосвязанных более мелких операций. К обработке относят операции проведения расчётов, выборки, поиска, объединения, сортировки, фильтрации и другие.

Важно помнить, что *обработка* – это систематическое выполнение операций над данными (информацией, знаниями); процесс преобразования, вычисления, анализа и синтеза любых форм данных, информации и знаний путём систематического выполнения операций над ними. Обычно отдельно выделяют операции обработки данных, информации и знаний.

**Обработка данных** (англ. «Data processing») – это процесс последовательного управления данными (числа и символы) и преобразования их в информацию.

Обработка данных может осуществляться в интерактивном и фоновом режимах. Напомним, что под *данными* понимают сведения, представленные в определённой знаковой системе, а информация – это данные, сопровождающиеся смысловой нагрузкой, помещённые в некоторый контекст; данные, как-либо оцениваемые приёмником информации. Как правило, получение информации связывают с уменьшением неопределённости существующего выбора; ответ на какой-либо заданный или подразумеваемый вопрос. Причём для одних личностей (или с одной точки зрения) данные вполне могут быть информацией.

**Технология обработки информации** – это упорядоченная последовательность взаимосвязанных действий, выполняемых в строго определённой последовательности с момента возникновения информации до получения заданных результатов.

**Обработка информации** представляет собой переработку информации определённого типа (текстовый, звуковой, графический и др.) и преобразования её в информацию другого типа. Например, различают обработку текстовой информации, изображений (графики, фото, видео и мультимедиа), звуковой информации (речь, музыка, другие звуковые сигналы). Использование новейших технологий обеспечивает их комплексное представление. При этом человеческое мышление может рассматриваться как процесс обработки информации. В библиотеках методы и процессы, связанные с обработкой поступающих в них документов с целью создания их описаний (метаданных) получили название *научная обработка литературы*.

Информационная технология обработки предназначена для решения хорошо структурированных задач, по которым имеются необходимые входные данные, известны алгоритмы и другие стандартные процедуры их обработки. Эта технология применяется в целях автоматизации рутинных постоянно повторяющихся операций, что позволяет повышать производительность труда, освобождая исполнителей от рутинных операций,

а порой и сокращая численность работников. При этом решаются задачи: обработки данных; создания периодических отчётов о состоянии дел; связанные с получением ответов на различные текущие запросы и оформлением их в виде документов или отчётов.

При обработке применяют такие информационные технологии, как:

- сбор и регистрация данных непосредственно в процессе производства в форме документа с использованием центральной ЭВМ или персональных компьютеров;
- обработка данных в режиме диалога;
- агрегирование (объединение) данных;
- использование электронных носителей информации.

Классификация технических средств обработки информации представлена на Рис. 57.

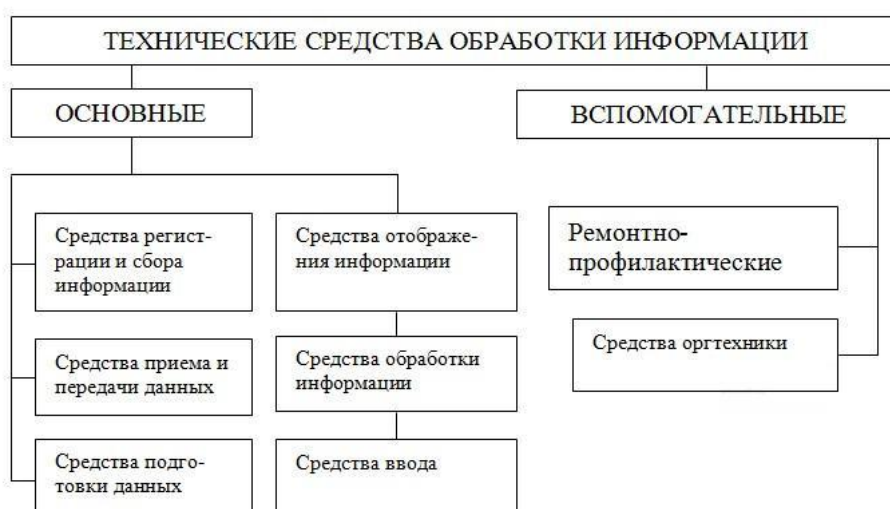


Рис. 57. Классификация технических средств обработки информации

*Студенты предлагают перечни технических средств обработки информации по каждому подвиду.*

Работы, связанные с вводом, редактированием (корректировкой) и т. д. библиографических описаний относят к процессам обработки информации и данных.

**Подсистема сбора и первичной обработки** предназначена для выполнения операций, связанных с предмашинной обработкой информации. В ней осуществляется сбор первичной информации об объектах в естественном виде (на естественном языке), например, описаний книг. Затем – отбор сведений, необходимых для включения в фонд АБИС; перевод в необходимый вид и формат (использование соответствующих форматов ввода, составление аннотаций и БО); фиксирование информации, прошедшей первичную обработку и соответствующим образом формализованной, на носителях (преимущественно бумажных – предмашинный формат).

**Подсистема ввода** используется для переноса информации на машинные носители и ввода её в ЭВМ, а также для контроля за правильностью переноса

информации с целью устранения возникающих ошибок. В качестве устройств ввода эффективно использовать терминалы (дисплей с клавиатурой, подключенный к главной машине). В Библиотеке Конгресса США в начале 1990-х годов обработка поступлений (каталогизация) ведётся на персональных компьютерах её работников в онлайн-режиме. При этом пользователям (читателям) доступны даже предварительно обработанные поступления с пометкой «предварительная запись».

### ***Библиографическая обработка документов в АБИС. Назначение и общая характеристика коммуникативных (обменных) форматов***

Каждая запись в электронном каталоге библиотеки состоит из набора характеризующих документ элементов (библиографическое описание и др.). Способом отображения этих элементов является формат, предназначенный для кодирования элементов библиографических записей.

В любой компьютерной программе, работающей с библиографическими записями, может использоваться свой внутренний формат, особенно это касается СУБД, например, многие первые СУБД имели по умолчанию формат «dbf», а в MS Access используется формат «mdb», о чём свидетельствуют типы файлов (расширения) с БД в этих СУБД. А, например, в отечественной СУБД АИБС «МАРК» использовался собственный внутренний формат «ldb».

Формат может иметь формализованную (как правило, табличную) и неформализованную структуру. Названные выше форматы «dbf» и «mdb» относятся к формализованным, а формат «ldb» – к неформализованным структурам. Использование каждой из них имеет свои достоинства и недостатки. Так, например, в формализованных структурах используется фиксированная длина любого поля электронной библиографической записи. В этом случае, как правило, такая БД обладает значительной избыточностью, то есть её объём больше чем фактическое количество хранящихся в БД данных. Например, для формирования поля фамилии автора необходимо установить его длину, соответствующую максимально возможной фамилии. Пусть длина такого поля выбрана равная 25 символам. Большинство отечественных (славянских) и особенно азиатских фамилий имеют гораздо меньшую длину (от двух до 8 символов). В этих случаях остальные символы в данном поле будут учитываться и восприниматься как пробелы.

В мире проблемами автоматизации библиотечных процессов занимаются достаточно давно. Известно, что в 1965 году Библиотекой конгресса США разработан формат представления в ЭВМ разделённых на записи библиографических данных. Формат получил название **MARC** (англ. «Machine Reading Catalog» или «Machine Readable Cataloging»), что означает машиночитаемый каталог или машиночитаемая каталогизация.

Каждая запись в нём соответствует описанию библиографического документа и содержит набор элементов, обеспечивающих осуществление поиска. Структура формата обеспечивает описание различных видов

информации (книги, сериальные, периодические и многотомные издания, рукописи, карты, ноты и др.), имеет 999 полей с подполями. Причём поле может содержать до 15 подполей, а сам формат является неформализованной структурой. Это означает, что все записи БД представляют собой как бы непрерывную строку, в которой поля, подполя библиографических записей и сами записи разделяются специальными символами. В начале каждой записи в данном формате указывается тип и количество символов в ней, затем идут цифровые коды описанных полей, а за ними – собственно содержание полей с указанием буквенных кодов подполей.

Формат является основой формирования электронных каталогов библиотек и библиографических БД. К концу 1970-х годов в США, Великобритания, Франция, Канада и Япония сформирована концепция внедрения компьютеров в практику библиотечной работы. В этих странах создавались собственные MARC-подобные форматы. Ныне аналогичные форматы широко используются в большинстве библиотечных систем в Европе.

Очевидно, что даже MARC-подобные форматы в различных странах (зачастую и внутри одной страны) будут несколько отличаться друг от друга, поскольку содержат некоторые традиционные для этих стран представления и технологии формирования библиографических описаний. Различия между внутробиблиотечными форматами обусловлены следующими факторами:

- разной практикой каталогизации,
- различиями в технологии обработки поступлений,
- различием в выходных формах (картотеки, каталоги, внутренние документы и т.д.),
- используемым программным обеспечением.

Для решения данной проблемы в 1980 году создаётся универсальный формат обмена данными между разными автоматизированными библиотечными системами, имеющимися в различных странах планеты – **UNIMARC** (Международный формат представления библиографических данных для автоматизированной обработки). Он является коммуникативным форматом внешнего представления обмениваемых данных. В данный формат конвертируются<sup>1</sup> любые MARC-подобные форматы.

Чем ближе внутренний формат к формату обмена, тем легче осуществлять однозначную конвертацию (конвертирование) данных. На американском континенте и в ряде других частей света в качестве внутрисистемного и коммуникативного формата одновременно используется USMARC, который после доработки получил название MARC21.

Хотя данное положение целесообразно учитывать при проектировании АБИС, это не означает, что внутренний формат должен фактически являться (полностью соответствовать) обменным форматом, что в отдельных случаях

---

<sup>1</sup> В данном случае речь идёт о преобразовании данных из одного формата в другой с максимально точным переносом данных из полей одной БД в соответствующие поля другой БД.

имеет место в нашей стране, например, АИБС «Академия» СПбУКИ базируется на обменном формате UNIMARC. Большинство стран Европы ориентируются на UNIMARC как на коммуникативный формат обмена библиографической информацией. **Основной целью создания формата UNIMARC является содействие международному обмену данными в машиночитаемой форме между национальными библиографическими службами.**

Коммуникативный формат MARC должен обеспечивать:

- возможность его применения для всех видов библиотечных документов;
- достаточную гибкость для решения разнообразных задач в дополнение к каталогизации;
- удобство применения в различных автоматизированных системах.

Таким образом, к концу прошлого столетия в мире для АИБС наибольшее распространение получили два формата, фактически выполняющих роль международных стандартов: USMARC и UNIMARC. Существует ряд программ-конверторов, как правило, поставляемых вместе с программными средствами, или встроенных в технологии приёма/передачи информации в корпоративных АИБС. В этой связи целесообразно использовать, например, универсальный конвертор UNIMARC-USMARC.

В России октябре 1995 года было принято решение создать **Российский Национальный формат**, получивший название RUSMARC.

Национальный коммуникативный стандарт RUSMARC создавался на основе формата UNIMARC в редакции 1994 года и ныне активно применяется.

MARC-формат прежде всего стандарт структуры данных, а не стандарт содержания. Содержание записи регламентируется правилами каталогизации, системой предметизации и т. д.

*Пример записей в MARC-формате.*

00934nam 22002297

450000100210000000500170002102000250003804000660006309000700012909

100110019909400130021009700120022309800120023524500760024726000530

032330000110037652001780038752100200056565000220058565300570060799

0004000664-RU/IS/BASE/122547147-20040204005324.7- a5-98032-292-2c20-

00- aУхтинский государственный технический университете19.11.2003- a67-

c67.405e156977;156978;156979;156980;156981fчзЛ;чзВ;абВхТ 78- a67.405- -

b12.11.03- bbiblio4- анаучная- aТрудовой кодекс Российской ФедерацииbС

изм. и доп. на 15 сент. 2003 г.- aМ.бПроф. юрид. системы "Кодекс":

Проспектс2003- a192 с.- aТекст кодекса подготовлен с использованием

профессиональной юридической системы "Кодекс", сверен с официальным

источником и приводится по состоянию на 15 сентября 2003 года.- -

aШирокая публика- aЮридические науки- aТрудовой кодекс; Российская

Федерация; Юриспруденция- b5d12.11.03e5f12.11.03j5t16av5-

Вводимая в АБИС библиографическая информация в формате MARC представлена на Рис. 58а.

The screenshot shows a library search interface with two main search sections: 'Быстрый поиск' (Fast search) and 'Простой поиск' (Simple search). The 'Быстрый поиск' section includes a search input field, radio buttons for 'Все', 'Книги', 'Газеты', and 'Периодика', and a 'Поиск' button. Below it are options to 'Установить Фильтры сессии' and 'Установки активного фильтра'. The 'Простой поиск' section has a search input field, a dropdown menu for 'Автор', and a 'Поиск' button. It also shows 'База данных: ИБ ИГТУ' and a message 'There are 0 titles in your cart.' At the bottom, there is a 'История поиска' (Search history) section with entries like 'Где-нибудь: Автомобильный транспорт И : р' and 'Предмет: социология'.

The main search results area displays a list of records. At the top, it says 'Запись 10 из 174' and 'Вы искали - ИБ ИГТУ - Предмет: социология'. Below this is a table with columns 'Метка поля', 'Ин 1', and 'Ин 2'. The records are listed with their respective MARC fields and values, such as '001 vtls000013693', '003 RU-NoGTU', and '040 9 % 200701221418 %b lada %c 200610131505 %d lada %e 200505231725 %d opus %y 200403232045 %z VLOAD %s RU-NoGTU %b rus %c RU-NoGTU %d RU-ToGU %e P5B0'.

Рис. 58а.

MARC-записи, накопленные за время использования формата, представляют собой огромный массив стандартизированных данных. Задача библиотечных систем – максимально использовать это стандартизированное, структурированное содержание, которое признается не только библиотечным сообществом. Поисковые службы Интернета предпринимают усилия к интеграции традиционного библиотечного содержания и служб в своих поисковых машинах. В качестве примеров можно назвать два проекта Google:

1. Google LibraryThing – веб-приложение для создания каталогов библиотек любого уровня, хранения и обмена библиографическими записями. Основная функция LibraryThing – это импорт данных от книготорговых организаций и библиотек, при этом библиотеки поставляют записи в MARC-формате и DC.

2. Google BookSearch – проект, направленный на создание всеобъемлющего виртуального каталога для поиска всех книг на всех языках при участии издательств и библиотек. В мае 2008 года подписано соглашение между Google и OCLC, в соответствии с которым, в частности, библиотеки-партнеры OCLC передают в BookSearch свои MARC-записи, расширяя возможности доступа к библиотечным коллекциям через веб.