

Лекция 7

Программное обеспечение технологий документооборота

1. Технологии электронного документооборота
2. Функции и классификация систем документооборота
3. Использование систем автоматизации документооборота
4. Технологии машинного перевода текстов

1. Технологии электронного документооборота

- Документ является основным способом представления информации в любой современной организации (предприятии) и одним из видов информационного обеспечения ИС.
- Документ — содержательно законченная уникально идентифицируемая единица информации на естественном языке.

Описание структуры документов

- согласно теговой модели WWW — HEAD и BODY;
- в информационно-поисковых системах — набором атрибутов;
- согласно стандарту ODA (архитектуры учрежденческих документов) как ПРОФИЛЬ и ТЕЛО;
- Описание документов в органи-зациях (на предприятиях) России.

Теговая модель

- Теги HTML-документов образованы с помощью слов английского языка, понятных сокращений и обозначений.
- Тег состоит из имени, за которым может следовать необязательный список атрибутов тега. Атрибуты тега могут иметь конкретные значения, устанавливаемые для изменения функции тега. Текст тега заключается в угловые скобки ("**<**" и "**>**").
- Например, при описании таблицы открывающий тег с атрибутами может выглядеть так:
- `<TABLE WIDTH=570 ALIGN=center>`
- Эта запись означает следующее: таблица шириной 570 пикселей, выровнена по центру.

Теговая модель

- Контейнер HTML состоит из двух других вложенных контейнеров: заголовка документа (HEAD) и тела документа (BODY):
- `<HTML> <HEAD> ... </HEAD>`
- `<BODY> ... </BODY>`
- `</HTML>`
- Тег тела документа BODY, в отличие от тега HEAD, имеет атрибуты.
Атрибут BACKGROUND определяет фон, на котором отображается текст документа. Так, если источником для фона HTML- документа является графический файл image.gif, то в открывающем теге тела BODY появляется соответствующий атрибут:
- `<BODY BACKGROUND="image.gif">`

Информационно-поисковые системы

Атрибутивный, когда каждому документу присваивается набор определенных атрибутов (полей). При сохранении документа в архив поля заполняются определенными значениями, в дальнейшем при поиске проверяется совпадение значений этих полей запросу. К атрибутам документа можно отнести имя документа, время создания, автора, машинистку, имя подразделения, тип документа (факс, письмо, контракт, спецификация). Список таких атрибутов должен быть расширяем. Их совокупность называется карточкой документа.

Информационно-поисковые СИСТЕМЫ

- Наиболее известные и популярные системы поиска:

www.aport.ru

www.yahoo.com

www.rambler.ru

www.yandex.ru

www.altavista.com

www.google.com

Стандарт ODA

- В настоящее время общепринятым и совместно используемым единым стандартом хранения и передачи структурированных данных является стандарт *ODA (Object Data Base Markup Language)*, т.е. такой формат данных, который не привязан ни к платформе, ни к конкретному программному пакету.
- *ODA* – язык описания структурированных данных объектного типа, который разработан на базе языка *XML*, рекомендованного в качестве стандарта. Документы формата *ODA* являются объектными хранилищами данных любого типа, структуры и содержания.
- Структура описания данных *ODA* позволяет представить в данном формате любые данные из существующих информационных систем, хранилищ и баз данных. Благодаря универсальному описанию информационных структур хранение, передача, анализ, интерпретация данных может осуществляться инструментами, создаваемыми любыми независимыми разработчиками без привязки к информационному контексту, но при этом все они будут полностью совместимы между собой.
- Формат *ODA* равнозначно воспринимается всеми операционными системами, в связи с чем, отсутствует платформенная привязанность и сложности с межплатформенными отношениями и поэтому является основой для единого формата хранения и передачи структурированных данных.

ЯЗЫК XML

- Языки разметки прошли путь от первых форм, создававшихся компаниями и госучреждениями, до Стандартного языка обобщенной разметки (Standard Generalized Markup Language - SGML), Гипертекстового языка разметки (Hypertext Markup Language - HTML) и в конечном итоге до XML. SGML может показаться сложным, а HTML (который, по сути, сначала был просто набором элементов) оказался недостаточно мощным для идентификации информации. XML разрабатывался как простой в применении и удобный для расширения язык разметки.

ЯЗЫК XML

- файлы XML состоят из текста и разметки. Большая часть текста помещается в элементы, в которых текст окружен тегами. Например, допустим, нужно создать поваренную книгу в формате XML. У нас есть рецепт под названием *малиновый крем*, который нужно преобразовать в XML. Чтобы разметить название рецепта, заключим его текст в элемент, который начинается и заканчивается тегами. Этот элемент можно назвать *recipeName*. Чтобы отметить начальный тег элемента, поместим его имя в угловые скобки (<>), вот так: <recipeName>. Затем введем текст *Малиновый крем*. После текста поставим замыкающий тег, который представляет собой имя элемента в угловых скобках, плюс косая черта завершения элемента (/) перед именем элемента, вот так: </recipeName>. Эти теги образуют *элемент*, в который можно вводить текст и даже другие элементы.

Описание элемента на языке XML

```
<recipe> <recipeName>Малиновый  
    крем</recipeName>  
</recipe>
```

Описание документов в организациях (на предприятиях) России

- Документы являются примером слабоструктурированной информации.
- Структура документов описывается в организациях (на предприятиях) России согласно унифицированной системе документации. При этом выделяют заголовочную часть (наименование организации (предприятия), код документа по ОКУД, наименование документа и т.д.), содержательную часть (в виде текста, таблиц), оформительскую часть (подписи лиц, несущих юридическую ответственность).

Описание документов в организациях (на предприятиях) России

Основой унифицированной системы документации является системы классификации документов. В российской практике используются системы классификации документов ОКУД, ОКДП (общероссийский классификатор видов экономической деятельности, продуктов и услуг), и др. Согласно той или иной системе классификации, любой документ организации (на предприятии) имеет свой код. Например, согласно ОКУД (Общероссийский классификатор управленческой документации), баланс имеет код — 0710001.

Документооборот

- Документооборот — совокупность информационных потоков организации, структурирующая процесс управления объектом, его производственно-технологическим циклом.
- Внутренние информационные потоки образуются между структурными подразделениями, имеющими как вертикальные, так и горизонтальные связи. Потоки бывают прямые и обратные. Внешние потоки образуются при взаимоотношениях с внешними организациями в процессе деятельности, связанной с отчетностью, рекламой, выставкой продукции, социологическими опросами, подготовкой справок, встречами с населением и пр.

2. Функции и классификация систем документооборота

Базовые операции технологий обработки документов в системах электронного документооборота

- ввод документов в систему;
- хранение, навигация, поиск и фильтрация документов;
- коллективная работа с документами;
- вывод информации из системы.

Системы электронного документооборота

нового поколения включают в себя:

1. Системы управления ходом работ — workflow management. Обеспечивают автоматическое направление документа определенному исполнителю в соответствии с разработанными правилами. Работа основана на картах деловых процессов, которые описывают, кто, когда, на каком рабочем месте (возможно, в удаленном филиале), с помощью каких программ и как должен обработать те или иные данные.

Системы электронного документооборота нового поколения включают в себя:

- 2. Групповые системы:
- workgroup management — системы групповой работы с документами. Обеспечивают возможность предыдущему исполнителю самому определять, кому направить документ на дальнейшее исполнение;
- groupware — системы группового доступа. Обеспечивают возможность параллельной работы над документами. Ориентированы на небольшие рабочие группы, характеризуются поддержкой выполнения одной коллективной задачи и отсутствием структуризации в организации работ.

3. Использование систем автоматизации документооборота

Наиболее популярные СЭД в России

1. Directum (Directum)
2. DocsVision (DocsVision)
3. Globus (Проминфосистемы)
4. Naumen DMS (Naumen)
5. PayDox (Paybot)
6. БОСС-Референт (ГК АйТи)
7. ДЕЛО (ЭОС)
8. ЕВФРАТ (Е1)(Cognitive Technologies)
9. МОТИВ (Motiware).

Наиболее популярные СЭД в России

10. Отечественные системы «ДЕЛО», «Золушка-WIN», LanDocs, «Эффект Офис», Optima Workflow (системы workflow management);
11. Русифицированные версии зарубежных систем DOCS Open, Links Works, Lotus Notes;
12. Системы, созданные на основе зарубежных — «Босс-Референт», «Золушка-Канцелярия», OfficeMedia (системы groupware);
13. Созданные в виде модулей корпоративных информационных систем — «1С: Электронный документооборот», «1С: Электронная почта», «Галактика: Управление документооборотом»,

4. Технологии машинного перевода текстов

Поколения систем машинного перевода

- Технология Transfer основана на алгоритмах последовательного перевода «слово за словом», «фраза за фразой». Особое внимание в них отводилось построению машинных словарей.
- Технология Interlingua основана на создании некоего промежуточного языка, описывающего все структуры входного и выходного языков. Предложение на входном языке сначала анализируется в терминах промежуточного языка, а затем создается предложение на выходном языке.

Поколения систем машинного перевода

- Технология Transfer основана на алгоритмах последовательного перевода «слово за словом», «фраза за фразой». Особое внимание в них отводилось построению машинных словарей.
- Технология Interlingua основана на создании некоего промежуточного языка, описывающего все структуры входного и выходного языков. Предложение на входном языке сначала анализируется в терминах промежуточного языка, а затем создается предложение на выходном языке.

Поколения систем машинного перевода

- Технология — Translation Memory (англ. — память перевода), основанная на использовании базы соответствий смысловых единиц, которая создается или пополняется, если есть готовая, перед началом работы. При переводе предложения или их части, отсутствующие в базе, не обрабатываются вообще, но зато те, что есть, переводятся всегда правильно. Применение технологии Translation Memory оправдано лишь для объемных текстов с многочисленными повторами, например технической документации, справочников и т.п. Наиболее популярными системами являются Promt, «Сократ».