Ввод информации с бумажных носителей

Борисов В.А.
Красноармейский филиал
ГОУ ВПО «Академия народного хозяйства
при Правительстве РФ»
Красноармейск 2009 г.



 Несмотря на все расширяющееся использование компьютеров практически во всех сферах человеческой деятельности, бумага остается одним из основных носителей, используемых для подготовки и хранения документов.



Во многих информационных технологиях в настоящее время бумажные носители используются для подготовки исходных документов, визуализации и придания легитимности отчетам, а для обработки и хранения данных - компьютеры, т.е. постоянно возникает необходимость ввода данных в компьютер с бумажных носителей.



Поле на бумажной форме или ее графическом образе, содержащее данные для ввода, будем называть «Полем считывания данных», а поле на экране компьютера и/или в памяти компьютера, в которое вводятся данные— «Полем ввода данных».



 Если данные вносятся в бумажные формы вручную, будем говорить о ручных методах и средствах подготовки данных.



 Если данные вносятся в бумажные формы с помощью механического или электромеханического устройства, будем говорить о механизированных методах и средствах подготовки данных.



 Если подготовка и/или ввод в компьютер выполняются человеком-оператором для каждого отдельного элемента данных, будем говорить об автоматизированных методах и средствах.



 Если набор данных подготавливается с помощью компьютера и/или вводится в компьютер с минимальным участием человека, состоящим в инициализации программы подготовки или ввода данных, будем говорить об автоматических методах и средствах.



Типовые этапы технологического процесса ввода данных в компьютер с бумажных форм

- Сортировка бумажных документов и разделение их на пачки.
- Выбор и подготовка пачки бумажных документов для ввода.

- Выбор из пачки очередного документа для ввода.
- Определение типа выбранного документа и подготовка программы к вводу данных этого типа.

- Поиск очередного поля считывания данных.
- Считывание и запоминание текста, содержащегося в текущем поле считывания данных.

- Распознавание считанных данных и их ввод в текущее поле ввода данных.
- Сравнение введенного текста с исходным. Если есть несоответствия, то его редактирование.

- Автоматическая проверка корректности введенных в поле ввода данных.
- Проверка окончания ввода данных с текущего документа.

- Автоматическая проверка взаимной корректности данных между полями текущего документа. Если данные некорректны, то их редактирование.
- Автоматический экспорт введенных по текущему документу данных в файл или базу данных.

- Проверка окончания пачки документов.
- Проверка наличия рассортированных пачек документов.

- Проверка наличия нерассортированных по пачкам документов.
- Передача пачек введенных документов в архив и окончание ввода данных.

Технология автоматизированного ввода

 Предполагает ввод оператором данных, записанных на бумажных формах, посредством клавиатуры компьютера и с помощью специально подготовленной программы ввода.



 Компьютер с инсталлированной программой ввода называют автоматизированным рабочим местом (APM) ввода данных.

Недостатки

- Низкая скорость автоматизированного ввода данных;
- Значительные затраты времени на вспомогательные действия по обеспечения ввода с бумажных документов;
- Достаточно много ошибок в данных.



 Создание средств автоматизированного ввода не требует больших затрат, эта технология проста в эксплуатации, однако она не позволяет обеспечить высокую производительность.

Технология сканирования и автоматического распознавания (OCR, ICR)

Предполагает предварительное сканирование бумажных форм с последующим автоматическим распознаванием графического образа текста всего документа либо отдельных его полей, выделенных в соответствии с заранее подготовленным шаблоном и запись распознанных данных в поля ввода в памяти компьютера.



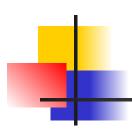
- Обеспечивает распознавание большинства типов печатных документов, заполненных сплошным текстом.
- Однако, главной проблемой остается ввод данных с бумажных форм, содержащих поля, заполненные рукописными текстами.



- Обеспечивает работу с машиночитаемыми формами, печатаются в одной типографии на качественной бумаге и аккуратно заполняются.
- Для подготовки графических образов документов в основном используются сканеры с автоматической подачей бумаги с лотка.



 Технология ICR эффективна только для распознавания данных, подготовленных на машиночитаемых формах (лучше всего заполненных числовыми данными).



 Для обеспечения более высокого качества распознавания, машиночитаемые формы должны быть не черно-белыми, а цветными, обычно красно-оранжевыми (этот цвет хорошо отсекается при сканировании) и совпадать на просвет.

 разработка машиночитаемых форм, регулярная печать их в достаточном количестве в одной типографии и рассылка для заполнения юридическим или физическим лицам, которые должны готовить и сдавать отчеты по этим формам;

 закупка оптического сканера, компьютеров, сетевого оборудования и программного обеспечения для создания комплекса сканирования и автоматического распознавания;

 разработка шаблонов машиночитаемых форм и настройка комплекса ICR на их распознавание;

 разработка программного обеспечения для интеграции результатов распознавания в конкретную информационную технологию;

 обеспечение обучения операторов и вспомогательного персонала для поддержки комплекса в работоспособном состоянии;

 обеспечение регулярного финансирования работ по эксплуатации комплекса.



Технология сканирования и автоматизированного распознавания (Imagewriter)

Функции Imagewriter

- указание параметров для первичной настройки комплекса;
- управление сканированием бумажных документов и формирование каталога их образов;
- открытие каталога ранее отсканированных документов;

Функции Imagewriter

- автоматизированный выбор из каталога образа документа для ввода;
- выбор вида представления образа документа на экране;
- увеличение или уменьшение масштаба образа документа на экране;

Функции Imagewriter

- поворот образа документа;
- автоматизированный выбор и наложение шаблона на образ документа;
- автоматизированный ввод данных в поля ввода на образе документа;



- автоматический контроль корректности введенных данных по правилам контроля;
- автоматический экспорт введенных данных.