

Реализация потокового ввода документов на примере ежегодного собрания акционеров ОАО «Газпром»

А. М. Акименков

В работе описывается система потокового ввода документов на примере системы обработки документов Ежегодного собрания акционеров ОАО «Газпром».

Введение

Система потокового ввода документов является комплексом технических устройств и программных модулей, служащим для ускорения и улучшения качества обработки бумажных документов. В основе программного обеспечения потокового ввода лежит модуль распознавания, который анализирует электронный образ — результат сканирования документа и распознает содержащийся в нем текст (как печатный, так и рукописный), а также другие элементы документа. Результат распознавания документа обрабатывается и помещается в базу данных. Как правило, при обработке документа база данных активно используется для контроля и коррекции распознавания и ввода данных. В случае невозможности уверенного распознавания необходимо участие операторов, использующих визуальный контроль. В системе потокового ввода присутствует модуль отчетов, с помощью которого можно получить отчеты, сводки и другие документы о результатах потокового ввода.

В данной работе в качестве примера системы потокового ввода описывается система потокового ввода документов Ежегодного собрания акционеров ОАО «Газпром» (ниже СПВ).

Все программное обеспечение СПВ создано ООО «Cognitive Technologies Ltd.».

1. Общая характеристика СПВ

1.1. Свойства СПВ, общие для систем потокового ввода

Поскольку в основе обработки бумажных документов при потоковом вводе является их распознавание, первым шагом для организации потокового ввода является стандартизация обрабатываемых документов. СПВ обрабатывает следующие типы документов, используемых при проведении Ежегодного собрания:

- доверенность о передаче голосов доверенному лицу;
- бюллетень голосования.

Каждый из этих типов документов имеет различные варианты.

Вид доверенности различается в зависимости от того, является ли доверенное лицо физическим или юридическим лицом. Кроме того, имеются групповые доверенности, а также еще некоторые варианты доверенностей.

С целью стандартизации вида доверенностей была создана специальная программа для печати доверенностей на персональном компьютере и распространена в подразделениях ОАО «Газпром».

Существуют также различные варианты бюллетеней. Бюллетени различаются по логике обработки и по дизайну. Кроме этого, каждое собрание вносит изменения в вид бюллетеней — меняется число вопросов повестки дня, число кандидатов в совет директоров и т. п. Относительно изменения повестки дня собрания СПВ является масштабируемой системой, т. е. вышеуказанные изменения не требуют перестройки или модификации СПВ. Бюллетени голосования печатаются централизованно Счетной комиссией Ежегодного собрания и вручаются акционерам для голосования. Бюллетени являются персонифицированными, т. е. при печати в них вносится индивидуальная информация об акционере — имя, количество акций, и т. д.

Общий объем документов для обработки может быть весьма значительным, поскольку общее число акционеров ОАО «Газпром» в настоящее время превышает 600 тыс. человек, а комплект бюллетеней для голосования содержит до 10 листов.

Держателем реестра акционеров ОАО «Газпром» является ЗАО «СР ДРАГа». Эта организация ведет соответствующий электронный документооборот и СПВ интегрирована с этим документооборотом. СПВ использует для своей работы базы данных этого документооборота. Физически основным для СПВ сервером баз данных является AS/400. В то же время вся текущая работа СПВ происходит на персональных компьютерах.

1.2. Особенности СПВ

Специфика проведения собрания акционеров определяет некоторые особые требования к СПВ. Основных требования два — ограничение времени для обработки документов, а также высокие требования к точности ввода документов.

Согласно правилам проведения Собрания акционеров, поступившие доверенности и бюллетени заочного голосования должны быть полностью обработаны к определенному сроку. Обычно для такой предварительной обработки отводится 1–1,5 месяца. Что касается бюллетеней, которые участвуют в голосовании непосредственно на собрании, то они все должны быть обработаны в короткий промежуток времени порядка 30–45 минут, включая выдачу отчета по результатам обработки, по которому Председатель счетной комиссии делает доклад на Собрании о результатах голосования.

Требуемая скорость обработки документов обеспечивается благодаря возможности гибкого масштабирования системы. То есть при необходимости в любой момент в СПВ могут быть включены дополнительные сканеры, станции распознавания, рабочие места операторов и т. п.

Второе основное требование — точность обработки документов. Отдельный бюллетень может представлять голоса крупного акционера. Минимальная ошибка ввода такого бюллетеня может повлиять на результаты голосования, что является неприемлемым.

Точность обработки обеспечивается комплексом перекрестных проверок результатов распознавания с использованием базы данных. При сомнении в результате распознавания используется участие оператора. При этом рабочее место оператора является интерактивным, т. е. для оператора подсвечиваются сомнительные места, выдается диагностика возможной ошибки, предлагается информация из базы данных, обеспечивается удобная навигация внутри документа и т. п. Результаты работы оператора также заново проверяются и т. д.

Как показали результаты тестирования, точность обработки документов СПВ намного выше, чем при вводе операторами вручную.

2. Общая схема СПВ

В СПВ можно выделить следующие основные компоненты или модули.

- Комплекс сканирования. В комплекс входит сканер, а также программа, выполняющая первичный контроль процесса сканирования и подготавливающая файлы для дальнейшей обработки.
- Станция распознавания. В результате работы этой компоненты создаются файлы с результатами распознавания сканированных документов.

- Модуль редактирования. С помощью этого модуля происходит проверка, редактирование и запись в базу данных обрабатываемых документов.
- Модуль отчетов. Служит для генерации различных отчетов о результатах работы СПВ.
- База данных на AS/400. Конечно, база данных не является собственной компонентой СПВ, но она активно используется на всех этапах работы.

2.1. Работа комплекса сканирования и станций распознавания

Комплекс сканирования может включать в себя произвольное число сканеров по необходимости. Новые станции распознавания также могут подключаться и отключаться по мере необходимости. В результате работы этих компонент создаются файлы с результатами сканирования и распознавания документов. При работе также проводится проверка форм вводимых документов (определяются случайно попавшие несоответствующие документы, пустые листы и т. п.). После проверок документу присваивается идентификатор, и в базе данных документируются все этапы обработки данного документа.

2.2. Работа модуля редактирования

Модуль редактирования является посредником между результатом сканирования документов и внесением документов в базу данных. С помощью этого модуля организуется проверка, корректировка и внесение в базу данных.

Персональный компьютер с установленным модулем редактирования является рабочей станцией редактирования. Рабочая станция редактирования может работать в трех режимах:

1. Режим фильтра.
2. Режим оператора.
3. Режим администратора.

В режиме фильтра станция редактирования работает автоматически, без участия оператора. Станция обрабатывает поступающие документы и автоматически вносит в базу те из них, которые распознаны отлично и которые не требуют дополнительной обработки оператором.

В режиме оператора работа станции происходит с участием оператора. Оператор может выбрать тип документа (доверенность или бюллетень), который он будет обрабатывать.

Модуль редактирования позволяет оператору видеть отсканированный образ обрабатываемого документа, осуществлять перемещение по

полям документа. При обработке документа модуль редактирования предоставляет оператору различную информацию и возможности, облегчающие, контролирующие и ускоряющие обработку документа. При показе документа выдается диагностика возможных ошибок. Показываемые поля документа имеют различную подсветку, информирующую о корректности или неуверенном распознавании этих полей. Выдается необходимая информация из базы данных, облегчающая контроль и редактирование документа. Кроме того, показывается контролирующая информация, необходимая для принятия решения о корректности документа, например суммарные распределяемые голоса.

Модуль редактирования предоставляет различные возможности удобной навигации по документу: перемещение по сомнительным полям, перемещение только по заполненным полям и т. п.

Если оператор по каким-либо причинам не может принять решение по обрабатываемому документу, или для принятия решения необходимо свериться с исходным бумажным документом, он направляет его для обработки администратором. Чаще всего необходимость в этом возникает при заполнении документа цветными или слишком светлыми чернилами, или при неудачном сканировании.

В режиме администратора оператор имеет те же возможности редактирования и ввода документов, что и в режиме оператора, но он имеет и дополнительные возможности. Например, только администратор может удалить документ из обработки.

2.3. Работа модуля отчетов

Данный модуль выдает отчеты как о текущем состоянии работы, так и итоговые отчеты для доклада об итогах голосования на Собрании акционеров.

2.4. Архивация документов

Большой объем обрабатываемых документов требует технологии, позволяющей быстро найти необходимый документ, в том числе по отсканированному электронному образу найти бумажный документ и наоборот.

Эта задача решается тем, что при сканировании сканер создает на обратной стороне документа надпечатку — уникальный цифровой номер — идентификатор. Этот идентификатор соответствует электронному идентификатору документа, который присваивается ему на этапе сканирования.

Идентификатор содержит два номера — номер коробки и номер листа. Физически документы помещаются в коробки с соответствующими

номерами в порядке возрастания нумерации листов. Данный модуль выдает отчеты как о текущем состоянии работы, так и итоговые отчеты для доклада об итогах голосования на Собрании акционеров.

3. Эффективность работы СПВ

Данный комплекс эксплуатируется уже более пяти лет. В течение этого времени он продолжает развиваться. Появились многостраничные бюллетени, система стала масштабируемой и т. д.

В течение времени своей работы СПВ доказал свою надежность и эффективность.