Для соблюдения распределения на входящих пачках (P_1, P_2, ...P_n), которое может нарушаться вследствие различного времени распознавания документов, предлагается использовать средства мониторинга характеристик выходного потока. Такого рода средства дают возможность отследить задания, не попавшие на распознавание в свою очередь и, если их количество превосходит некоторую пороговую величину, изменить вероятностные характеристики диспетчера для компенсации отклонений режима диспетчирования.

Настоящие исследования выполнены на базе международной лаборатории ELDIC с 1996 г при поддержке фонда Леонарда Эйлера DAAD.

ЛИТЕРАТУРА

1. *Маматов Ю.А.* Организация работы мультипроцессорных СЦВМ с многоуровневой памятью //диссертация на соискание ученой степени д.т.н.

УДК 658.512

Ю.М.Вишняков, А.Цур, С.Ю.Новиков

КАТАЛОГИЗАТОР ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

Автоматизация документооборота предполагает использование комплекса ввода документов с бумажных носителей. Его ключевым звеном является пакетный документный сканер, а сама пакетная обработка требует автоматизации распределения документов по каталогам (каталогизация). В настоящей работе предлагается одна из реализаций такого каталогизатора, разработанного в рамках работ международной лаборатории ELDIC в области безбумажной обработки информации.

Пользователь создает свою «виртуальную» структуру архива в виде дерева (файлы физически находятся в одном каталоге), структура архива записывается в файл и может модифицироваться. Электронные документы имеют сквозную нумерацию, что позволяет при необходимости переходить на следующий или предыдущий документ в пачке. Предусмотрены автоматический, ручной и полуавтоматический режимы каталогизации.

Повышение интеллектуальности каталогизатора достигается через использование различного рода критериев каталогизации путем подключения соответствующих программных модулей. Методики и критерии каталогизации документов зависят во многом от области приложения. В настоящее время нами апробирована и используется простая каталогизация на основе ключевых слов. Практические исследования показали, что следующий алгоритм полуавтоматической каталогизации имеет неплохие результаты.

1) создать базу понятий для всей группы документов. Здесь отсекаются слова на основе словаря не рекомендуемых для индексирования слов;

2) выполнить частотную фильтрацию. Опытным путем установлено, что слова, не являющиеся словами из словаря не рекомендуемых слов и имеющие максимальную частоту, являются ключевыми. Пользователь сам устанавливает частотный порог, после которого в базе понятий остаются только ключевые слова;

3) задать заголовки тем и входящих в них понятий;

4) автоматически сформировать темы;

5) выполнить в случае необходимости вручную корректировку дерева каталога.

Исследования проведены при поддержке фонда Леонарда Эйлера немецкой службы академических обменов.

УДК 658.512

Ю.М.Вишняков, С.И.Родзин

ТЕХНОЛОГИЯ АССОЦИАТИВНОГО ПОИСКА В ПЕРСОНАЛЬНЫХ ЭЛЕКТРОННЫХ ГРАФИЧЕСКИХ АРХИВАХ

Разрабатывается новая компьютерная технология для организации ассоциативного поиска в персональных графических архивах, которая обеспечивает согласованное выполнение следующих требований:

- формулировку запроса для поиска графического объекта с использованием графических примитивов и вербального текста, задаваемых пользователем;
- интеллектуальный ассоциативный поиск в архиве запрашиваемого графического объекта за разумное время с представлением пользователю права окончательного выбора;
- визуализацию графических образов в полиоконном миниатюрном формате.

Технология базируется на использовании современного аппарата «мягких вычислений» (искусственные нейронные сети и нечеткая логика), при этом время поиска линейно зависит от числа ассоциативных соединений в нейросети и слабо зависит от числа хранимых имен графических объектов.

Предполагается, что пользователь может иметь очень приблизительное представление о том, что именно он ищет, поэтому алгоритм поиска носит недетерминированный характер, а исходной информации, формируемой в виде запроса к системе, присуща неполнота, неточность и неопределенность.

Меню системы предусматривает задание следующих атрибутов поиска:

- тип графического объекта;
- величина и положение объекта;
- цвет;
- тип линий и поверхности образа;
- топологические особенности объекта.