

Administración Electrónica de Documentos en Automatización de Oficinas

Marcos Boracchia¹, Hugo Ramón², Silvia Esponda³, Marina Iglesias⁴, María Bernarda Albanesi⁴, Guillermo Ricci⁴

Laboratorio de Investigación y Desarrollo en Informática⁵

Facultad de Informática

Universidad Nacional de La Plata

Resumen

La Automatización de Oficinas es un área importante de las aplicaciones en Ciencias de la Computación, en la cual los avances tecnológicos (velocidad de procesamiento, comunicaciones, periféricos, capacidad de memoria, etc.) produjeron grandes cambios en los últimos años.

La modelización y organización de una oficina sin papeles distribuida son temas de investigación, especialmente en tres ejes principales: adquisición y digitalización de datos, almacenamiento y recuperación de datos y distribución en tiempo real de información multimedia.

En este trabajo se discuten distintas alternativas tecnológicas, poniendo énfasis en la administración de imágenes digitales en una organización distribuida en el marco del Proyecto Sistemas Distribuidos en Automatización de Oficinas.

¹ Lic. en Informática - Profesor Adjunto Dedicación Semi Exclusiva, Facultad de Informática, UNLP. E-mail: marcosb@lidi.info.unlp.edu.ar

² Magister en Informática - Profesor Adjunto Dedicación Exclusiva, Facultad de Informática, U.N.L.P. E-mail: hramon@lidi.info.unlp.edu.ar

³ Lic. en Informática - JTP Dedicación Semi Exclusiva, Facultad de Informática, UNLP. E-mail: sesponda@lidi.info.unlp.edu.ar

⁴ Becario LIDI – Fac. de Informática. UNLP. E-mail: miglesias@lidi.info.unlp.edu.ar

⁵ Calle 50 y 115 Primer Piso, (1900) La Plata, Argentina, Teléfono/fax +54(221)4227707
<http://lidi.info.unlp.edu.ar>

Justificación

En los últimos años ha surgido el concepto del documento digital, la necesidad de estructurar y tratar "fragmentos" de información (probablemente de distinta naturaleza) como un todo bajo una identificación única mediante una computadora.

Estos fragmentos de información tienen distintos orígenes y vienen en distintos soportes: un escrito en papel, una imagen de microfilm, un fax, un mensaje de correo electrónico, registros de una base de datos.

Una manera de tratar la información visual basada en medios no magnéticos es digitalizarla y administrarla en forma electrónica. [AVE96], [AMR97].

Las tecnologías asociadas a Internet, intranets, imágenes electrónicas y otras, permiten la comunicación, coordinación y colaboración (independientemente de la ubicación geográfica de los usuarios).

En la actualidad los Sistemas de Administración de Imágenes Electrónicas pueden tener distintas tendencias (según el nivel de integración al que apunten) y no es raro encontrar mezclas de tendencias. Existen los sistemas stand-alone, pero existen también paquetes de software que permiten "mejorar" las aplicaciones existentes y agregarles capacidades de manejo de imágenes (image-enabled) y por último la mayor integración, la mayor flexibilidad, pero también la mayor complejidad, SAIEs desarrollados a partir de software de componentes y/o toolkits.

Independientemente de su tendencia los SAIEs actuales permiten obtener algunos de los siguientes beneficios:

- Aumentar la productividad
- Reducir costos
- Mejorar la atención al cliente
- Acceso inmediato y simultáneo al mismo documento
- Mejorar las condiciones de trabajo

Temas de Investigación

Sistemas de cómputo

La arquitectura DCE (Distributed Computing Environment) es un conjunto de programas usados para desarrollar y mantener aplicaciones cliente-servidor a través de plataformas heterogéneas en una red. Cuando se implementa este diseño, toda la red aparecería al usuario como una única computadora rápida y poderosa. Este ambiente provee a un usuario o a una aplicación las siguientes herramientas y servicios:

- RPC (Remote Procedure Call), que permite invocar una aplicación o función en otras estaciones de trabajo o servidores.
- Threads (Segmentos de programa que se pueden ejecutar independientemente), que aceleran los procesos considerablemente pues se pueden ejecutar en paralelo en varias máquinas.
- Medidas de seguridad, que se pueden aplicar automáticamente a toda la red.

En un ambiente distribuido, cuando todo el software está basado en PCs, un sistema de imágenes se puede integrar usando DDE (Dynamic Data Exchange) y/o OLE (Object Linking and Embedding).

La mayoría de los sistemas de imágenes administran dos tipos de datos: bases de datos relacionales (los índices) y los archivos de imágenes. La mayoría también permite integrar documentos creados por otras aplicaciones tales como procesadores de texto o planillas de cálculo. Normalmente, la base de datos se almacena en medios de alta velocidad, tales como discos duros o arreglos de discos (RAID) pero algunos sistemas almacenan los datos del índice en discos ópticos o CDs. Los documentos también pueden almacenarse en medios ópticos más baratos en dispositivos individuales y/o jukeboxes.

Una API es una interface a los servicios y protocolos provistos por una aplicación o sistema operativo. Esto habitualmente abarca un conjunto de funciones públicas que pueden usar los programadores, desarrolladores y aplicaciones. [PIX96], [SNO98]. Una API permite a un sistema histórico existente interactuar directamente con la funcionalidad de imágenes. La utilización de API requiere el código fuente del sistema histórico y tiene varios beneficios.

Los sistemas de imágenes de componentes son relativamente nuevos. En un sentido amplio, las componentes se pueden considerar como módulos periféricos usados para el desarrollo del sistema de imágenes o como toolkits visuales para desarrollar en forma personalizada todo el sistema de imágenes. Un toolkit visual permite a un programador seleccionar funciones (tales como escanear, imprimir, etc.) a partir de botones y arrastrarlos y soltarlos según las necesidades. Por ejemplo Visual Basic (o Delphi o PowerBuilder) tiene "agregados" (plug-ins) denominados controles personalizados (custom controls) que pueden adicionar funcionalidad a la aplicación. Estos controles también se conocen como VBXs, OCXs o ActiveX.

Dispositivos ópticos, discos y otros temas de almacenamiento

Como regla general, el tamaño de una imagen de documento digitalizada y comprimida es de 50 KB para un documento en hoja carta escaneado a 300 dpi. Sin embargo, hay varios aspectos que pueden afectar este tamaño de documento: los gráficos y las líneas del documento incrementan el almacenamiento, los caracteres pequeños requieren resoluciones de escaneo más altas por legibilidad y por lo tanto aumentan el tamaño de la imagen, el fondo y la tasa de compresión también deben tenerse en cuenta.

Después de determinar el medio de almacenamiento, el tamaño del medio y los requerimientos de almacenamiento en línea se pueden calcular por medio de fórmulas. Se debe estudiar la aplicación y se deben evaluar los volúmenes para determinar la cantidad de dispositivos necesarios para soportar la operación diaria del proceso de negocio.

La fórmula para determinar el tamaño de almacenamiento necesario variará dependiendo de los criterios de compresión, del tipo de documentos que se ingresan al sistema, la resolución requerida y el medio propiamente dicho.

Temas de entrada y captura de información

El proceso de captura de documentos abarca varias funciones clave, que deben ser analizadas y discutidas antes de implementar el sistema de imágenes. Estas funciones son:

- Preparación de documentos
- Escaneo/Entrada
- Indexado
- Control de Calidad
- Reescaneo
- Envío a almacenamiento

La preparación de documentos abarca el quitado de ganchos, el alisado de documentos doblados, la colocación de encabezados de lotes, etc. La preparación depende del tipo de escáner.

Un documento mal escaneado puede no notarse hasta la etapa final de control de calidad. Invariablemente en todo sistema habrá un factor de reescaneo. Esto también debe tenerse en cuenta cuando se determinen los requerimientos de escáner.

El escaneo es un proceso que consiste en digitalizar una página y convertirla en una imagen que la computadora pueda manejar. La imagen se puede almacenar en forma binaria o puede convertirse en texto mediante el uso de software de OCR.

En todo ambiente de entrada es importante realizar un paso de control de calidad con el proceso de escaneo o de entrada, donde se ve cada documento en pantalla verificando la claridad u otros parámetros de calidad. En sistemas pequeños, en esta etapa también se realiza el indexado manual.

La selección de los escáneres correctos se basa en el análisis de varios puntos clave: el volumen diario; existencia de días pico; condiciones especiales de los documentos; condiciones del papel (fino, arrugado); consideraciones de indexado (OCR, ICR, código de barras, full text), etc.

Al considerar la transformación de la información basada en papel a datos digitalizados o imágenes, se debe llevar a cabo un cambio de paradigma. Esto sólo se puede lograr entendiendo completamente los parámetros existentes. La clave para un indexado exitoso depende de la habilidad de proveer tantas referencias cruzadas como sean necesarias para todos los potenciales usuarios del sistema y hacerlo con una cantidad mínima de entrada de datos.

Con un buen diseño de índices, los sistemas de imágenes superarán a cualquier otro método de recuperación. Sin embargo, sin un diseño del índice hecho a conciencia, el sistema de imágenes puede fracasar rápidamente.

La razón principal por la cual las imágenes electrónicas son una tecnología única comparada con otras tecnologías de automatización de oficinas es la completa transformación de los medios y métodos existentes (papel/microfilm/archivos) a medios digitales. Ninguna tecnología actual requiere tal compromiso con el cambio y bases tan sólidas para hacer este cambio drástico.

Uno de los obstáculos más grandes para implementar sistemas de imágenes es el problema de la conversión. En la mayoría de las aplicaciones hay un archivo histórico en papel o microfilm que debe ser provisto de la mejor solución posible. Esto requiere una evaluación y análisis cuidadoso de la información histórica y de los requerimientos de uso de la información.

Hay básicamente tres soluciones al problema de la conversión: No convertir, conversión bajo demanda y conversión total.

Comunicaciones y redes

La tecnología de redes es importante para los sistemas de imágenes por varias razones. Primero, porque brinda la posibilidad de distribuir las imágenes y los documentos. Pero simultáneamente con las mejoras en la tecnología de redes, las nuevas aplicaciones como workflow, groupware, videoconferencia, proyectos de trabajo compartidos, etc., dependen de capacidades de red más sofisticadas.

Una página de tamaño carta comprimida como imagen puede ocupar treinta veces más espacio que la misma página almacenada como un archivo de procesador de texto. Las redes permiten almacenar esta página una sola vez y aún ponerla disponible para muchos usuarios. Mover estos grandes

archivos de imágenes de un lugar a otro con rendimientos aceptables, es un reto para las redes actuales.

Resultados Obtenidos e Investigaciones Futuras

Se ha desarrollado un conjunto de aplicaciones que constituyen un sistema de administración de documentos, implementado parcialmente en la Facultad de Informática de la UNLP, utilizado para el manejo de Resoluciones de la Facultad.

Se invirtió mucho tiempo y esfuerzo en analizar y desarrollar la aplicación de consulta de los documentos para que los criterios de búsqueda y recuperación sean creados dinámicamente por el propio usuario a partir de cualquier dato guardado del documento. Esta recuperación inteligente de los datos es la gran ventaja de los sistemas de administración de documentos y es el factor clave para superar exitosamente el cambio de paradigma que implica la implementación de este tipo de sistemas.

Un beneficio más abstracto, pero no por eso menos significativo, es que los sistemas de administración de documentos bien implementados, es decir alimentados con los contenidos adecuados, pueden ser un gran soporte en la toma de decisiones al permitir localizar información en tiempo y forma. Esta información antes era inaccesible o incompleta por no estar en formato electrónico ni centralizada.

Actualmente se está analizando la futura integración del Sistema de Administración de Documentos con los otros sistemas administrativos de la Facultad y extender el uso a otros documentos de la Organización.

También se está analizando en general la manera de integrar otros tipos (en cuanto al medio de soporte original) de documentos (en este caso electrónicos) más allá de los documentos en papel, extendiendo o adaptando los SAIEs.

Otro tema de interés es definir una arquitectura web para los SAIEs.

Bibliografía

[ATD98] "Document Imaging Technology. Certified Document Imaging Architech Study Guide". @doc. 1998.

[AVE96] "Introduction to Electronic Imaging, 3rd Edition". Don M. Avedon. AIIM. 1996.

[SAF95] "The Complete Guide to Document Imaging". Elisa Safdie, Tom Flanagan. The Applied Technology Group. 1995.

[SNO98] "Enhancing Applications with Imaging Technology". Snowbound Software. The Applied Technology Group. 1998.

[AMR97] "The IBM Enterprise Document Management Suite". Donna M. Amrhein. The Rheinner Group. 1997.

[OTG98] "Storage Management". OTG Software. The Applied Technology Group. 1998.

[ATG98] "Document Management. Glossary of Terms". The Applied Technology Group. 1998.

[SAF98] "The Document Life Cycle: A White Paper". William Saffady. AIIM.

[PIX96] "ISIS. An Image and Scanner Interface Specification". Pixel Translations. 1996.