

Gerenciamento Eletrônico de Documentos *GED*

document imaging



www.laserfiche.com

Copyright © 2001 pela Compulink Management Center, Inc.
Todos os direitos reservados.

Laserfiche é uma marca registrada pela Compulink Management Center, Inc. Todas as marcas mencionadas neste manual são de propriedade de suas respectivas empresas. Nenhuma parte desta publicação poderá ser reproduzida, fotocopiada, armazenada em algum sistema de pesquisa ou transmitido sem o consentimento por escrito do publicador.

LaserFiche Document Imaging
Uma Divisão da Compulink Management Center, Inc.
3545 Long Beach Blvd.
Long Beach, CA 90807
USA

Gerenciamento de Documentos Eletrônicos
Publicado por LaserFiche
Primeira edição em Português Outubro de 2001
Impresso no Brasil.
Venda Proibida

Conteúdo

Introdução.....	1
O que é Gerenciamento de Documentos Eletrônicos - GED	2
Trazendo documentos	3
Arquivando documentos.....	5
Indexando documentos	8
Recuperando documentos.....	10
Controlando acesso	11
Benefícios de Gerenciamento de Documentos Eletrônicos	12
Implementando Gerenciamento de Documentos Eletrônicos.....	13
Avaliando suas necessidades	13
Avançando de projeto piloto a solução empresarial...	13
Instalação	13
Treinamento.....	14
Assuntos Legais	15
Suporte e Manutenção	16
Características adicionais	17
Facilidade de uso.....	17
Anotações	17
Imprimir / Fax / e-mail.....	17
Opções de procura usando texto.....	18
Internet/ Intranet.....	19
Segurança.....	19
Portabilidade e CDs	20
Terceirização de Escanerização.....	21
Recuperação pós Desastre.....	22
Processamento em Lote.....	22
Código de Barras.....	22
Zona OCR	23
Workflow	23
Escalabilidade.....	24
Compatibilidade de Sistemas.....	24
Sistemas de Rede.....	25
Arquitetura Cliente/Servidor.....	25
Formato de Arquivo Não Proprietários.....	26

Conteúdo

Perguntas Mais Frequentes	27
Geral.....	27
Escanerização / Importação / Armazenamento.....	28
Visualização / Impressão / Exportação.....	31
OCR: Reconhecimento de Caracter Óptico.....	32
COLD: Computer Output to Laser Disc.....	33
Glossário de Termos	34
Sobre o Autor	43

Introdução

"Tenho feito buscas que poderiam provavelmente ocupar-me por três dias de trabalho em dez minutos. Nossos arquivos são tesouros históricos, este é um dos motivos que nós usamos LaserFiche e pessoas utilizam nossos documentos históricos para pesquisas. Nós queremos arquivar os materiais originais sem risco de deteriorização."

Linda Butler, City Clerk
Flagstaff, AZ

Papel

Nós todos necessitamos dele para fazer nosso trabalho mas papel acumula-se rapidamente. Nossos arquivos engordam rapidamente, arquivos e sistemas de alimentação facilitam a busca de documentos. Gerentes de registros organizam, arquivam e recuperam nossa informação mas o montante de papel continua crescendo. Arquivos de papel são frequentemente difíceis de encontrar

Documentos podem não estar na pasta apropriada ou podem estar "emprestados" para alguém ou ainda perdidos na mesa de outra pessoa. Estudos mostram que profissionais frequentemente perdem até 500 horas por ano apenas procurando documentos.

Esses dias se foram. Gerenciamento Eletrônico de Documentos oferecem uma melhor maneira de gerenciar registros sem dependência..

Documentos Eletrônicos devem:

- Gerenciar milhares de registros e recuperar o que você necessita em segundos.
- Fácil de utilizar, idenpendente se você é a pes soa que necessita do documento ou se você é o gerente de sistemas.
- Permitir que você compartilhe documentos com colegas enquanto protege informações confidenciais.
- Permitir que você envie e-mails ou envie faxes dos documentos com um clique do mouse.
- Prover um fácil acesso para compartilhar documentos com outros escritórios ou levando-os consigo.
- Adaptar-se com a maneira que você trabalha sem forçar mudanças.

Desde 1987, nós da LaserFiche temos dedicado horas incontáveis para construir o Sistema de Gerenciamento Eletrônico de Documentos. Desenvolvido por experiência, nós criamos esse guia para explicar o que Sistema de Gerenciamento de Documentos Eletrônicos é, o que encontrar nele e como fazer de armazenameto e recuperação de documentos um processo simples e transparente.

O que é Gerenciador de Documentos Eletrônicos?

GED é a conversão de documentos em papel para imagens eletrônicas em seu computador. Uma vez em seu computador esses documentos podem ser recuperados em segundos. Todas organizações geram um enorme volume de papéis e documentos eletrônicos. Nós temos desenvolvido nosso próprio jeito de arquivar documentos importantes. Todos sabem a frustração de não ser capaz de encontrar um documento certo quando ele é muito necessário. Métodos tradicionais de arquivamento de papel e registros eletrônicos requerem um grande esforço para gerenciar, distribuir e encontrar esses documentos. Proporcional ao número de arquivos que cresce, o tempo e esforços necessários para gerenciá-los também aumentam.

GED revoluciona o arquivamento de informação e provém meios de rapidamente recuperar e compartilhar todos documentos em seu sistema. Todos os sistemas GED devem possuir os seguintes cinco componentes básicos:

- Ferramentas de escanear para trazer os documentos para o sistema
- Métodos de arquivamento e armazenamento de documentos
- Ferramentas de recuperação para encontrar documentos
- Controle de acesso para prover documentos para pessoas autorizadas

Como a Internet explodiu em popularidade por motivo do rápido acesso que provem às informações armazenadas em web pages, sistemas GED provem um imenso valor pelo seu rápido acesso a informações armazenadas entre documentos de organizações.

GED constroe pontos fortes em documentos de papéis: Arquivos são escanizados ou eletronicamente convertidos e uma alta resolução de fotografia é armazenada em um disco rígido ou óptico. Eletrônico “cartões de índices” podem anexar informações para um documento como autor, número de referência ou data de criação. Arquivos podem ainda ser visualizados, impressos, compartilhados e armazenados e as imagens adicionam uma enorme vantagem oferecido aos documentos um conteúdo ativo.

Não mais apenas imprimir uma página, documentos textos são “lidos” por tecnologia de reconhecimento de caracteres ópticos, Optical Character Recognition (OCR). Um sistema deve permitir que você recupere arquivos procurando por qualquer palavra ou frase no texto, por localização de pastas ou por informações em “cartões de índice”. Quais documentos podem ser acessados e lidos e quais ações e modificações pode-se executar nesses documentos, depende do nível de segurança de acesso o qual deve ser controlado pelo sistema GED.

Selecionar o adequado sistema GED pode ser uma tarefa excitante. Existem vários aspectos a considerar para assegurar que adaptar-se-á às necessidades de sua organização. Seguem as descrições dos cinco componentes básicos em o quê deve ser avaliado na escolha de seu sistema.

Trazendo os Documentos

Existem três métodos primários para trazer arquivos de documentos para um sistema GED:

- Escanear arquivos de papel
- Conversão, criando arquivos de imagens inal teráveis
- Importação, criando versões modificáveis de documentos eletrônicos

Escaneando

Escanear um documento produz uma imagem que pode ser armazenada em um computador. Quando escolher um scanner, é importante considerar o total do orçamento e o tamanho e volume de documentos a serem escanizados. A capacidade de utilizar uma grande variedade de scanners é uma das características de um bom sistema de imagem.

Um scanner para sistemas GED deve possuir ADF (Automatic Document Feeder), alimentador de documentos automático. Esse dispositivo permite que uma pilha de papéis seja colocada na bandeja do scanner e automaticamente alimentar o scanner página por página,

acelerando assim o processo de escaneamento. Scanners sem ADF são primariamente designados para imagens gráficas e requerem que cada página seja colocada manualmente, uma por uma.



Scanner com ADF

Scanners podem suportar uma variedade de tamanho de papéis, de cartões de visita à desenhos de engenharia. A maioria dos departamentos apenas necessitam escanear documentos até tamanho (8" x 14"). Para organizações ou departamentos que utilizam blueprints, planos e desenhos arquitetônicos, existem os scanners de largo formato que suportam documentos e-sized (34" x 44"). Em geral, quanto maior o tamanho de papel que o scanner pode suportar, mais caro ele custa. Outras opções como coloridos ou escalas de cinza, utilizado para fotografias, podem aumentar o preço do scanner.

A velocidade do scanner é outro fator a considerar. Scanners para GED podem suportar de 10 à 200 páginas por minuto. São disponíveis em modo simples e/ou duplo. Scanner de modo duplo (duplex) permitem que os dois lados do documento sejam escanizados ao mesmo tempo em uma única passada pelo scanner. Como as outras opções, alta velocidade de scanners também aumentam o preço do scanner. Algumas vezes é mais barato comprar dois scanners de 20 páginas por minuto do que um scan-

ner de 40 páginas por minuto. Essa opção é apenas compatível com sistemas GED que suportam múltiplas estações de escaneamento.

Se existe um enorme número de documentos a escanear, por exemplo, milhares de páginas, pode ser mais prático e econômico utilizar um serviço terceirizado de escaneamento. Para suportar essa opção, o sistema GED deve acomodar facilmente a sincronização entre as páginas escanerizadas, o serviço terceirizado e as páginas escanerizadas dentro da organização. O volume de dados que contém imagens e informações de índice necessitam ser modulares e facilmente portáteis. Isso assegura que os documentos escanerizados pelo serviço terceirizado podem ser incorporados no sistema “ao vivo” sem interrupções de sistema e sem interrupções e/ou reindexação do trabalho existente. Essa opção é frequentemente referida como “volume portáteis”.

Se uma organização tem vários escritórios e necessita compartilhar os documentos escanerizados um por um, volumes portáteis tem a capacidade e a maneira de facilmente distribuir arquivos.

Conversão

Converter documentos é um processo de transformar documentos do processador de textos ou planilhas eletrônicas para um permanente formato de imagem para ser armazenado com um sistema de imagem. Aplicações Windows como Microsoft Word, Excel ou Autodesk AutocadCAD podem “imprimir” arquivos existentes em uma imagem inalterável do doc-

umento. Essas imagens são usualmente armazenadas com arquivo de qualidade TIFF (Tagged Image File Format). O processo de conversão também gera um completo arquivo de texto enquanto mantém o formato visual do arquivo original. Esse arquivo de texto também pode ser usado para indexação de texto do documento para auxiliar em pesquisas e recuperações futuras.

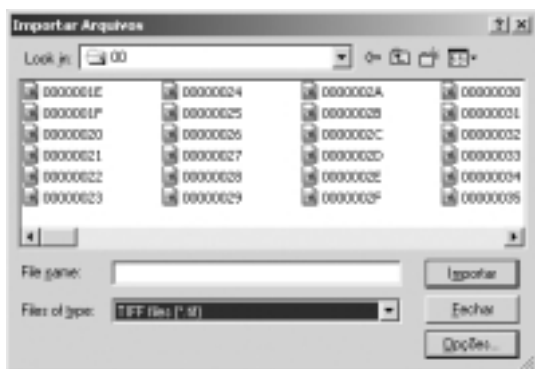
Converter documentos eletrônicos descongestiona o uso do scanner e economiza papel e recursos de impressão além de produzir uma imagem clara do que scannear documentos de papel. Esse método é o melhor para arquivos permanentes.



Convertendo documentos em imagens

Importando

Importar também é conhecido EDM – Eletronic Document Management, é o segundo melhor método para trazer documentos eletrônicos como documentos do Office, gráficos, clips de audio ou filmes para um sistema GED. Arquivos podem ser clicados e arrastados para um sistemas GED mas são modificáveis e mantém seu formato nativo. Esses arquivos podem ser visualizados no seu formato original sendo inicializados pela aplicação original ou utilizando pelo próprio sistema GED.



Importando arquivos

Arquivando documentos

Uma vez que os documentos foram trazidos para o sistema, eles devem ser arquivados. Sistemas GED devem incluir mudanças tecnológicas, aumento de volumes de documentos e testes de tempo. A necessidade e orçamento para armazenamento de imagens são melhores determinadas por organizações individuais envolvidas.

Um bom sistema GED deve ser capaz de utilizar qualquer dispositivo de armazenamento atualmente disponível, e os que estarão sendo lançados futuramente, para manter uma longa duração para armazenamento de documentos. Isso permite que você selecione o equipamento que melhor adaptar-se com sua necessidade, tanto agora como no futuro.

Para assegurar a habilidade de leitura no futuro, se um sistema GED não é utilizado para arquivo digital, os arquivos devem ser em um forma-

to não proprietário. A indústria de informática avança mais rapidamente do que armazenar imagens de documentos ou arquivos de textos, um formato proprietário pode deixar uma organização refém das aplicações proprietárias em capricho a uma única empresa.

Nesse momento, existem cinco opções primárias de armazenamento;

- Mídia magnética (Discos Rígidos)
- Armazenamento Óptico-magnético
- Discos compactos
- DVDs
- WORM

As vantagens e desvantagens de cada uma são descritas a seguir.

Mídias Magnéticas (Discos Rígidos)

A cada vez mais aumentam a velocidade de resposta ao acesso na recuperação de documentos e uma grande redução do preço fazem que as mídias magnéticas, tanto como os discos rígidos isolados ou um sistema RAID (Redundant Array of Independent Disks) são uma escolha popular.

Sistemas RAID significam o agrupamento de vários discos rígidos de mesmo tamanho e modelo configurados para que apareçam como um único disco no armazenamento de grandes volumes. Esse sistema protege dados contra perdas provendo redundância de dados entre os discos e tolerante a falhas de hardware contra os possíveis problemas em discos. Esses disposi-

tivos não possuem um custo alto e podem ser agrupados para armazenar grandes volumes de documentos e prover um rápido acesso.

A desvantagem dessa opção é que enquanto não são caros ainda possuem a possibilidade de conter problemas de hardware como falhas mecânicas. Arquivos de dados também são facilmente deletados, então operações de backup desses dados é fator muito importante para assegurar os dados pois podem ser restaurados.

Armazenamento Óptico Magnético

Anos anos atrás, as mídias Ópticas Magnéticas (OM) como disquetes e discos foram uma solução popular para fazer backup de arquivos pessoais de um computador. Como o termo diz por si só, um dispositivo OM utiliza tecnologias tanto óptica como magnética que obtém ultra e alta densidade. Um típico OM é um pouco maior que o convencional disquete de 3.5" que podem armazenar até 1.44 Mb de dados, um OM pode armazenar até 100Mb até alguns gigabytes (Gb).

Um sistema OM arquiva os dados em alta densidade utilizando laser e combinação entre leitura e gravação magnética. Ambos, laser e magnético são utilizados para adicionar os dados no disquete. O laser aquece a superfície do disquete e então pode ser facilmente magnetizado. Dados podem ser deletados e/ou sobregravados sem limite de vezes, como um convencional disquete de 3.5".



Disquete Óptico Magnético

OM drivers incluem conveniência, custo modesto e confiabilidade. Arquivos de dados podem ser completamente apagados. Com a baixa do custo dos discos rígidos, a popularidade dos OM caiu. Os discos OM podem ser instalados em jukeboxes que suportam centenas de discos.

Discos Compactos

São pequenos discos plásticos usados para armazenar informações digitalmente. Desenvolvidos inicialmente para sistemas de áudio e como uma alternativa para registros fonográficos, CDs são atualmente também utilizados para armazenar dados de sistemas provenientes de computadores. Informações digitais são gravadas no CD codificando como uma série de microscópicos pontos na superfície reflexiva do alumínio do disco. O disco é coberto com um plástico transparente que é executado em um equipamento que utiliza uma interface laser para ler as estampas dos pontos na área de superfície dos discos. Os CDs não têm a habilidade de gravar no processo de execução quando acessados. Formatos padrões de CDs incluem CD-ROM (Compact Disc-Read

Only Memory), um CD gravável, CD (CD Recordable) que permite a gravação uma única vez e ainda CD-RW (CD Rewritable) que permitem várias gravações em um mesmo CD.

CDs oferecem um seguro e confiável meio que provém um termo longo de armazenamento de imagens, em alguns casos até 100 anos. CDs utilizam especificações ISO-9600 que significa que os dados podem ser lidos em várias plataformas de computadores como por exemplo PCs, MACs....

A desvantagem básica dessa solução é o limite de armazenamento é a capacidade do CD e podem ser acessados pelos dispositivos de CD-Roms, Torres de CD e jukeboxes até 500 discos fazendo disso um método conveniente para armazenamento de grandes números de documentos.



Torre de disco compacto

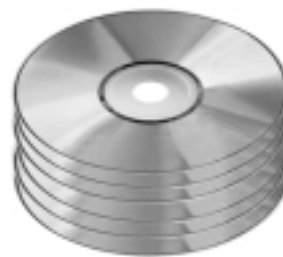
DVDs

DVD significa Digital Video Disc ou Digital Versatile Disc que é a próxima geração de tecnologia de armazenamento em discos ópticos. É essencialmente maior, mais rápido e podem armazenar maiores volumes de informações que os CDs e ainda capacidade para vídeo e audio. DVD apontam para diversão e lazer,

computadores e informações em negócios com um único sinal. Irão substituir os CDs, fitas de videos e talvez os cartuchos de jogos.

DVDs têm a capacidade de arquivar mais dados do que o espaço físico em um CD. Como um CD ou DVD, permite acesso randômico em qualquer parte do disco sem a necessidade de passar para a frente ou para trás como uma fita . Como um OM , um DVD nunca encosta a mídia no momento do acesso, no caso de música é tocada por um feixe de raio laser, sendo assim não produzem arranhos ou gastos na mídia se você persistir acessando o mesmo dado. A superfície plástica é sensível a toque de dedos , poeira e sujeira. Cuide deles com os mesmos cuidados dos CDs mas sem tratamentos especiais. Isso significa que DVDs podem ser executados milhões de vezes e continuam a representar a melhor opção de longa duração para um seguro armazenamento de documentos GED.

As desvantagens dessa opção são os altos custos e o diferente padrão entre unidades de DVDs e unidades de CD-ROMs como no caso dos videos VHS e Beta , diferentes empresas estão usando diferentes formatos para gravar DVDs.



Discos Digitais

WORM

WORM significa , Write Once Read Many , ou seja, Grave uma vez e leia várias. É uma tecnologia de disco óptico que permite que grave dados em um disco apenas uma vez. Os dados são permanentes e podem ser lidos inúmeras vezes. Esse formato de mídia requer um especial dispositivo de hardware e software para executá-lo. Diferente dos CD-ROMS eles apenas podem ser lidos em drivers compatíveis. Não existe muita aceitação, embora tenham encontrado um pequeno mercado de arquivo de mídia. Por motivo de um limitado número de empresas que provém, materiais e suporte para tecnologia WORM não é altamente recomendado.



Disquete WORM

Indexando Documentos

Quando documentos em papéis são recebidos em um escritório, eles devem ser organizados para serem utilizados. Eles são normalmente etiquetados, organizados, indexados, grampeados e localizados em pastas em armários. Sem esses passos nada poderia ser encontrado em um local de trabalho. Documentos eletrônicos não são diferentes. Um sistema GED deve prover diferentes métodos de organizar as informações para um uso futuro. Não importando qual o tipo de combinação de metodologia de indexação é utilizada, necessita ser fácil de usar e de fácil entendimento por pessoas que recuperam os documentos bem como por aqueles que armazenam os mesmos.

Existem várias idéias de como introduzir a metodologia para adicionar os documentos no sistema GED. Em geral quanto mais um sistema GED puder adaptar-se no sistema atual da empresa, menor será as mudanças internas e a necessidade de treinos.

Existem três maneiras primárias para organizar documentos em sistemas GED:

- Arquivos de índices
- Indexação de texto
- Estrutura de pastas

Arquivos de índices

Indexando documentos utilizando campos e palavras-chaves é um método tradicional utilizado com papel que traduz muito bem aos sistemas eletrônicos. Um sistema GED deve permitir usuários a customizar modelos de índices, criar vários modelos e ter diferentes tipos de índices de dados com cada um desses modelos como data e caracteres numéricos e alfabéticos. Campos de indexação podem ser usados para criar categorias de documentos, rastrear data de criação e/ou retenção, descrição do documento entre usuários. Em adição um sistema GED deve permitir características de caixas de opções para acelerar a entrada de dados nos campos e ter ferramentas disponíveis para auxiliar a entrada automática de dados nos índices de informação..

Indexação de textos – Full-text

Provedo indexação de texto, sistemas GED eliminam o tempo necessário para pessoas qualificadas ler e manualmente indexar documentos usando palavras-chaves. Para fazer isso, o software deve ter a capacidade de executar OCR, Optical Character Recognition. Esse processo “lê” uma página escanizada e então indexa cada palavra. Isso reduz dramaticamente os custos de indexação enquanto provém capacidades de pesquisa. Com as indexações de texto completo você pode localizar os documentos utilizando qualquer palavra mesmo quando a(s) palavra(s) não estão contidas em um índice de palavras.

Tipicamente, quando um computador utiliza OCR, o documento utiliza como padrão o idioma Inglês. Se a utilização de diferentes

idiomas é requerido, um sistema GED deverá suportar OCR e indexação em texto nos mesmos idiomas em questão.

Para prevenir trabalho extra, um bom sistema GED deve permitir o OCR e indexação de texto automático sem requerer envolvimento humano

Estrutura de pastas

Com campos indexados e texto completo indexado, um sistema GED deve prover um método visual para localizar documentos. Na maioria dos escritórios, arquivos são localizados procurando por pastas ou em gavetas específicas. Um sistema GED deve ter a habilidade de eletronicamente recriar esse sistema através de vários níveis de pastas

A flexibilidade da estrutura de pastas facilita o preenchimento dos documentos eletrônicos e faz do sistema de GED mais eficiente.



Exemplo de estrutura de pasta/arquivo

Recuperando Documentos

Recuperação de documentos é onde um potente sistema de indexação é pago. Usuários necessitam utilizar ferramentas para encontrar documentos entre sistemas baseados no que eles conhecem. Em alguns casos, isso significa procurando através de pastas, em outros casos isso pode significar conduzir buscas em campos de indexação. Se tudo que se sabe sobre o documento que se quer buscar é uma palavra que ele contém, uma indexação de texto completa irá auxiliar a encontrar esse arquivo relevante. Entretanto, o método de recuperação deve ser simples e de fácil utilização.

Usuários que são familiarizados com documentos textos devem ser capaz de utilizar as informações para encontrar o que eles necessitam. Alguns sistemas podem apenas encontrar páginas baseadas em palavras chaves. Esse método não é sempre o melhor porque a pessoa que seleciona as palavras chaves não necessariamente é a mesma pessoa que faz a busca. Para ser totalmente útil, um sistema GED deve utilizar recuperação de texto completo.

Como isso, utilizar um nome de documento e pastas para localizar e encontrar um documento pode auxiliar e ser intuitivo mas nem sempre é o melhor método. As vezes uma pessoa irá saber exatamente quais documentos ela necessita, mas não em que pasta eles estão e/ou o nome do documento.

Utilizando as informações de campos de indexação para encontrar um documento em particular pode ser útil. Um sistema com características completas irá possuir campos de modelos definidos pelo usuário. Campos de indexação para encontrar um específico documento pode ser útil. Campos de indexação permitem usuários a combinar milhões de registros em segundos para encontrar seus documentos. Claro que uma pessoa necessitará conhecer como o documento foi categorizado e a que campos de modelo foram associados a ele.

Para maximizar a eficácia da busca, um sistema de busca inteligente deve ser capaz de combinar buscas pelos modelos com textos completos com os nomes dos documentos e a localização nas pastas. Um bom sistema GED faz a recuperação de documentos relevantes rápida, fácil e eficiente.

Controlando Acesso

O componente final e obrigatório de um sistema GED é o controle de acesso. Em muitos ambientes de computadores, diferentes pessoas utilizam diferentes tipos de computadores em diferentes localidades para buscar informações. Um sistema GED completo deve prover a esses diferentes usuários apropriados níveis de acesso, sem comprometer confidencialidade e/ou segurança. Para fazer isso, um sistema deve possuir duas características fundamentais:

- Ampla disponibilidade
- Segurança adequada

Ampla disponibilidade

Um sistema GED deve oferecer diferentes caminhos para acessar arquivos. Um amplo nível de acesso economiza recursos financeiros, capital intelectual e sistema de rede. O método mais comum de acessar é pelo computador do usuário. Todo sistema GED deve prover uma interface cliente-servidor que permita escaneização, indexação e recuperação de documentos. Sem essa interface básica, o sistema não pode funcionar.

Para prover ampla disponibilidade e flexibilidade de acesso, sistemas GED hoje em dia devem combinar as necessidades dos escritórios com suas diversas utilidades e lotações remotas. Sistemas GED não são mais um processo dentro do escritório. Muitos usuários requerem portabilidade para trocar documentos com outros colegas e com outros escritórios em outra localidade. Isso é frequentemente feito através de

CDs , notebooks e/ou documentos via e-mail. Sistemas GED sem essa flexibilidade limitam as habilidades para o usuário.

Em adição, compartilhar documentos através da Internet e Intranet permitem aos administradores de sistema a desenvolverem um sistema GED através da rede corporativa e/ou para o público. Usuários devem ser capazes de buscar, recuperar e visualizar documentos com um Web Browser. Documentos acessados pelo browser remove as limitações de localizações e de plataformas de computadores como (Windows, Macintosh , Unix , etc)

Segurança Adequada

As organizações utilizam sistemas GED para arquivar uma enorme variedade de documentos, tanto público como privados, um sistema de controle de acesso necessita estar presente. Um adequado sistema de segurança deve permitir ao administrador do sistema controlar em que pastas e documentos os usuários podem visualizar, que ações podem executar nesses documentos (editar, copiar, deletar, etc). Esse sistema deve controlar acesso as pastas, documentos e até imagens com tarjas e textos de uma maneira simples. A habilidade de desenvolver GED para uma grande variedade de usuários requer um robusto sistema de segurança combinado com a interface do usuário.

Um bom sistema de acesso irá fazer um sistema GED disponível para todas as pessoas autorizadas tanto em um escritório como em uma localidade remota pela web, tudo sem comprometer a segurança do sistema.

Benefícios de Gerenciamento de Documentos Eletrônicos

Milhões de organizações ao redor do mundo utilizam GED todos os dias ao invés de sistemas de papéis. GED oferece um número de benefícios sob papel e microfimes

- **Recuperação Rápida** – permite que você encontre documentos rapidamente sem deixar a sua mesa de trabalho. Papel e microfilme são lentos porque usuários devem ir até os arquivos e acessá-los manualmente.
- **Indexação Flexível** – pode indexar documentos em várias diferentes maneiras simultaneamente. Indexar papel e microfilme em mais de uma maneira é inoportuno, caro e consome tempo.
- **Busca em texto- full text** – GED podem recuperar arquivos por qualquer palavra no documento, recurso impossível com papel e microfilme.
- **Sem perda de arquivos** – documentos escanerizados mantêm-se nas suas pastas quando são visualizados, nenhum é perdido ou não localizado. Ainda mais, modelos de índice e buscas no texto completo podem localizar documentos se eles foram acidentalmente movimentados. Perdas de documentos são caras e consomem tempo para repor.
- **Arquivo Digital** – O risco de perda ou danificação de papel para documento eletrônico e reduzido com um sistema GED. Mantendo as versões de arquivos em

sistemas GED, auxilia a proteger documentos em papel de serem sobre carregados e mantém os documentos em um formato não proprietário.

- **Compartilhe arquivos facilmente** – sistemas de GED fazem o compartilhamento de documentos facilitado entre colegas de trabalho e cliente pela mesma rede de computadores, por um CD ou pela WEB. Documentos de papel normalmente requerem uma fotocópia para serem compartilhados e microfilmes requerem conversões para papéis.
- **Melhora de Segurança** – GED pode prover melhor e mais flexibilidade de controle sobre documentos. Controles de segurança nas pastas, documento individual, nível de palavras e/ou para diferentes grupos ou indivíduos. Em contraste, todos documentos de papéis arquivados em um armário possuem o mesmo nível de segurança.
- **Economize Espaço** – GED irá ajudar a recuperar valioso espaço físico no escritório tirando as pilhas de papéis.
- **Recuperação de desastres** – GED provém um fácil caminho para fazer um backup dos documentos que podem ser localizados fora do escritório. Papéis são caros e é uma cara maneira de fazer cópia de documentos também sendo vulneráveis a incêndios, enchentes e roubos.

Implementando Gerenciamento de Documentos Eletrônicos

Avaliando suas necessidades

Quando decidir para um sistema GED, existe um número de questões a considerar.

- Quantos documentos devem ser armazenados, considerando tanto o número de documentos existentes como os documentos adicionados manualmente? Essa informação determina quanto espaço de armazenamento é necessário, a configuração de hardware e o custo do sistema.
- Quantos usuários irão utilizar o sistema ao mesmo tempo? Isso determina os custos de softwares preliminares e o tamanho do servidor.
- Quais departamentos irão utilizar o sistema e o que o público terá acesso? Isso determina quais características de especificação e níveis de segurança que serão necessários
- Quais os problemas sérios devem absolutamente ser resolvidos, e quais itens devem serem encaminhados para especificar características e níveis de segurança necessários
- Você necessita de uma solução padrão ou customizada? Isso determina o número de consultores, instalações, treinos, configurações e suporte que serão necessários.
- Que tipo de rede esta sendo atualmente utilizada, NT, NLM, LAN ou outras, irá manter-se? Isso determina restrições, configurações de sistemas e atualizações de estações de trabalho.

Avançando de Projeto Piloto a Solução Empresarial

Muitas vezes, escritórios encontram a melhor maneira de iniciar um projeto piloto envolvendo um ou mais departamentos antes de expandir o sistema por toda a empresa. Muitas empresas iniciam com Departamentos de Documentação mas GED pode ser implementado em qualquer departamento que seja necessário. Projetos pilotos permitem as organizações desenvolver completamente e testar os procedimentos de imagens antes de comprometer-se com uma solução empresarial.

Instalação

O primeiro passo de uma instalação deve ser a inspeção do local pelo revendedor do software para determinar o equipamento necessário e os problemas de conectividade de redes. Instalar hardware envolve em desempacotar, conectar e configurar todos os componentes como instalando os sistemas operacionais e drivers. Isso também inclui testar os equipamentos para assegurar a funcionalidade adequada do hardware e as conectividades de rede.

Depois de testar hardware, a instalação de software consiste em instalar o sistema GED no servidor e nas estações clientes e testar. Geralmente o revendedor irá executar essas atividades.

Treinamento

Programa de treinamento deve direcionar diferentes níveis de usuários e diferentes preocupações.

Usuário Final

Isso envolve ensinar usuários finais os mecanismos do sistema. Esse treino deve ser no local de trabalho. Cada grupo de usuários devem receber instruções necessárias para assegurar conforto com o novo sistema GED. O tempo de treinamento necessário irá depender do nível dos usuários em utilizar aplicações Windows, o sistema GED facilita o uso das novas mudanças para procedimentos existentes fornecendo uma operação amigável e poucas mudanças nos procedimentos. A maioria dos usuários aprendem muito rápido. É aconselhado classes com até 10 usuários e que os participantes não sejam interrompidos.

Administração do Sistema

Para assegurar que o sistema GED funciona tranquilamente é importante treinar alguns indivíduos como administrar o sistema. O treinamento no próprio ambiente de trabalho é o mais recomendado porque aumenta a familiaridade com detalhes específicos do sistema GED.

Implementação de Consultoria

Esse envolve aqueles que são responsáveis por gerenciamento de registros e documentos que desenvolvem estratégias para traduzir os documentos existentes em papel para sistema

eletrônico. Os gerentes de atividades como registros de documentos necessitam entender a diferença nessa atividade entre papel e eletrônico, decisões de retenções de tempos, armazenamento e metodologias de trabalho devem ser definidas antes de iniciar o trabalho. A duração do treinamento depende da complexidade do sistema.

Operação Prática Supervisionada

Isso envolve a supervisão do escritório utilizando personalizadas. Isso permite aumentar o conforto com o sistema e a fazer as perguntas diretamente ao técnico. Essa forma de treino é um excelente caminho de fazer as pessoas confortáveis com o novo sistema.

Assuntos Legais

Como sistemas GED são triviais hoje em dia, várias leis reconhecem valor em arquivos de documentos eletrônicos. Muitas agências de governo, afirmam que documentos originários de papéis podem ser destruídos uma vez que:

- Registros devem ser arquivados em um formato que não permite alteração como CD, DVD ou WORM
- O sistema tenha controles sensatos para assegurar integridade, precisão e confiabilidade.
- O sistema deve prover algum tipo de auditoria para detectar criações , adições , deleções de registros.
- Uma completa e precisa transferência de arquivos pode ser feita.

- O sistema possui confiáveis controles para prever e detectar deteriorização de registros
- Existir um sistema de indexação que permite localizar registros
- Permite imprimir os documentos
- O sistema deve permitir regerências cruzadas entre outros sistemas de registros.
- O sistema deve possuir documentação explicativa de como o produto funciona e como pode ser configurado.

A legalidade dos sistemas de GED variam de cada agência federal, estado, país e departamento envolvido. Organizações devem consultar advogados para especificar os estatutos de governo para sua área.

Suporte e Manutenção

Como manter um carro funcionando tranquilamente, sistemas GED requerem suporte externo. Revendas devem oferecer diversos tipos e níveis de suporte de atualizações de softwares a manutenções em clientes. Vários fatores afetam o nível de suporte que a organização necessita:

- Tamanho do sistema comprado
- Quantidade do sistema em uso
- Sistemas de missões críticas que devem ser operacionais 24 horas por dia em 7 dias da semana
- Pessoa certificada com um nível de experiência em sistema GED
- Acesso a Internet
- Mudanças na rede e/ou infraestrutura da organização
- Mudanças entre pessoas

Suporte pode incluir qualquer uma das seguintes:

- Atualização de softwares
- Linha telefônica “Hotline”
- Acesso remoto via dial-up no seu sistema
- Software patches disponíveis via FTP
- Boletins técnicos regularmente publicados
- Visitas de manutenções em clientes
- Adicional e avançadas sessões de treinos
- Suporte de hardware

Quando comprar hardware como servidores, dispositivos de armazenamento e estações de trabalho, empresas devem escolher revendas com boa reputação para serviço e suporte. Enquanto o custo inicial pode ser mais alto, os benefícios incluem menos tempo com equipamentos sem uso por problemas e maior consistência na precisão de sua funcionalidade.

Características Adicionais

Sistemas GED devem prover o básico em escanerização, recuperação e visualização. Embora um sistema GED seja designado para vários usuários e/ou muitos documentos terão necessidades mais rigorosas. Essa sessão discute esses requerimentos e várias maneiras de abordá-los.

Facilidade de uso

Uma das mais importantes características de sucesso de um sistema GED é a facilidade de uso. Um sistema apenas será utilizado se a inclusão de documentos for fácil e simples. Os melhores sistemas são flexíveis, possui interface gráfica intuitiva e adapta-se com a maneira que as pessoas já trabalham .

Anotações

Anotações provém informações adicionais sobre um documento sem alterar sua imagem original. Registros de papéis são geralmente anotados selecionando , adicionando mensagens, tarjas ou notas e registros de imagens podem suportar essas anotações. A segurança de um sistema GED deve controlar quem pode visualizar anotações como seleções, notas e quem pode ler através das tarjas.

As anotações típicas de um sistema GED incluem:

- Selecionar imagens e textos em várias cores e enfatizar palavras e sessões

- Adicionar tarjas (preto e/ou branco) nas imagens e preservar a integridade do documento original
- Adicionar imagens como “confidencial”, “importante” , “enviado via fax” nos documentos denotando aprovação ou rejeição
- Adicionar notas que contenham comentários adicionais.

Todas anotações devem ser sobre-postas e não podem mudar a imagem original. Deste modo, um documento pode ser impresso com ou sem as anotações. E mais importante ainda, de um ponto de vista legal, um documento arquivado em um sistema GED pode oferecer frequentemente cópias atualizadas dos documentos sem alterar sua forma original.

Imprimir/ Fax / E-mail

Sistemas de GED devem prover caminhos para obter informação do sistema. Imprimir, enviar faxes e e-mails são algumas maneiras de fazer isso. Para maximizar sua utilização, sistemas GED devem suportar as mais comuns impressoras e faxes e ainda serem capazes de imprimir imagens, textos e anotações.

Com o crescimento em popularidade da Internet, mais pessoas usam e-mail para se comunicarem e enviarem informação. Organizações terão uma grande economia se transmitirem seus documentos via e-mail ao

invés de fax ou serviços de correio. Sistemas GED deveriam ter uma opção que permitisse facilmente o envio de imagens para qualquer sistema de correio eletrônico compatível com MAPI (Mail Application Program Interface), e que fosse lido por quem não possui GED.

Opções De Procura Usando Texto

Para maximizar a eficácia da procura de texto, existem várias opções de ajuda. Essas opções consideram possíveis erros de OCR, oferecem ferramentas para restringir a busca, provê linhas de contexto para as palavras procuradas, e facilitam a busca da palavra quando o documento é visto.

Lógica “Fuzzy”

A procura de um texto assume que as palavras procuradas foram escritas corretamente e o OCR está correto. Infelizmente, as pessoas erram na escrita e não existe um pacote de OCR 100% perfeito. A lógica “Fuzzy” compensa esses erros através da procura de variações na forma como a palavra é escrita. Um sistema GED deve permitir que o usuário controle a quantidade de “fuzziness” da busca através da escolha do número de letras que podem estar erradas na palavra ou qual o percentual uma palavra pode estar errada. Por exemplo, uma busca utilizando a lógica “fuzzy” da palavra “caneta” poderia achar a palavra “caneta”, “careta”, “gaveta”.

Wildcards

Wildcards são caracteres, como * (asterisco) e ? (interrogação), que podem ser usados em um texto completo ou no índice de busca de palavras-chave para compensar por erros de escrita, ou quando a escrita é incerta. O asterisco representa qualquer caracter ou caracteres, enquanto que a interrogação representa um único caracter. Por exemplo, procurando por “b*m”, você acharia as palavras “bem”, “bom”, “bombom”. Procurando por “b?m” apenas encontrará as palavras “bem” e “bom”.

Operadores Booleanos

Quando uma busca de texto for realizada, normalmente vários documentos irão corresponder aos critérios de busca. Os operadores booleanos (E, OU, e NÃO) ajudam a ajustar a busca e reduzem o número de documentos não relacionados. Por exemplo, para encontrar documentos relacionados ao Ministério da Educação, a busca seria “Ministério E Educação”.

Buscas por Aproximação

Buscas por aproximação podem ser usadas para limitar os resultados da busca. São utilizadas para encontrar palavras que ocorrem dentro de um certo número de palavras, sentenças ou parágrafos. Por exemplo, para achar documentos relacionados a processos contra tabaco, mas não sobre plantação de tabaco ou processamento de tabaco, usuários poderiam procurar por “tabaco” dentro de uma sentença que tenha “processos”.

Linhas de Contexto

Até buscas específicas geralmente resultam em vários documentos possíveis. Além de prover usuários com uma lista de documentos que atendem aos critérios de busca, alguns sistemas GED apresentam linhas de contexto que mostram como cada ocorrência da palavra é usada em cada documento encontrado. As linhas de contexto permitem ao usuário encontrar o documento apropriado sem a necessidade de ver todos os documentos na lista de resultados.

Highlight de palavras encontradas

Uma vez que o documento foi selecionado, as palavras de busca precisam ser localizadas dentro dele. Para ajudar esse encontro, alguns sistemas de GED mostram a página apropriada do documento e ressaltam a palavra de busca no texto e em imagens. Dessa forma fica fácil para o usuário encontrar imediatamente a parte do documento que o interessa e ampliá-la, ao invés de ter que procurar pelo documento inteiro.

Internet/Intranet

Um GED deve ser capaz de prover uma forma simples de publicar informações na Internet ou Intranet. Isso permite que organizações compartilhem informações com outros departamentos, escritórios remotos, clientes ou para o público. Sistemas na WEB devem ser totalmente aptos à buscas de palavras e também suportar protocolos de segurança como sistemas de rede. Idealmente, um sistema GED não requer HTML ou códigos complexos para publicação de páginas na WEB.

Segurança

A segurança de um sistema de GED é um ponto crítico para uma implementação de sucesso. Enquanto que segurança pode não ser a primeira preocupação para uma instalação departamental, ela se torna mais importante à medida que o sistema se expande e permite o acesso a diferentes departamentos e acesso público aos arquivos. Um GED deve prover segurança em vários níveis, permitindo que cada instalação utilize o método que melhor se adaptar às suas necessidades. A segurança do sistema deve ser capaz de realizar esse nível de segurança e ser simples de administrar.

Direitos de Acesso

Um sistema GED deve permitir que organizações controlem o acesso às pastas e documentos individuais, tanto no nível de grupo quanto individual. A utilização de grupos e direitos adquiridos permite aos administradores rapidamente assinalar privilégios de leitura, enquanto que a segurança a nível de usuário permite que usuários específicos como gerentes vejam documentos que o resto do grupo não vê. Por exemplo, o departamento de Polícia pode ter acesso aos próprios arquivos e aos do de Planejamento, mas o departamento de Planejamento pode apenas ver os próprios arquivos.

Direito a Funções

Um sistema GED deve também permitir que a organização controle o direito às funções em relação às pastas e documentos individuais, em ambos os grupos e a nível individual. Enquanto

que direitos de acesso controlam quais pastas ou documentos um usuário pode ver, direito a funções controla as ações que o usuário pode exercer em um documento, tais como adicionar, editar, copiar ou apagar arquivos. Por exemplo, enquanto departamentos diferentes podem ter privilégios de leitura de minutas da Prefeitura, apenas os funcionários da prefeitura podem modificar esses arquivos.

Tarjamento de texto

O tarjamento de texto (em preto ou branco) permite que a segurança seja controlada a nível de palavras. Um sistema GED deve oferecer a habilidade de bloquear imagens ou textos. A habilidade do usuário de ver textos bloqueados irá depender do seu acesso de segurança. Por exemplo, relatórios de crime podem estar disponíveis para múltiplos departamentos, mas apenas o departamento de polícia é capaz de ver informações de identificação como nomes.

Auditoria

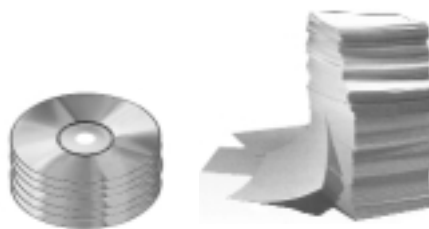
Como um nível adicional de segurança, um sistema GED deve oferecer a habilidade de rastrear quem está usando o sistema, quais documentos estão sendo lidos, quais ações estão sendo aplicadas aos documentos, e quando essas ações aconteceram. A habilidade de auditoria é especialmente importante quando documentos são confidenciais e quando existem vários usuários.

Portabilidade e CDs

Sistemas GED podem permitir que usuários carreguem documentos importantes para qual-

quer lugar, de tal forma que possam ler os documentos em um outro computador.

Quando as pessoas viajam a negócios, frequentemente é necessário trazer os documentos nas viagens. Carregar vários documentos em papel é praticamente impossível, e assim informações importantes são deixadas para trás. Com um sistema que suporta volumes portáteis, documentos podem ser copiados para outros locais. Pastas contendo documentos relevantes podem ser transferidas para outros bancos de dados rápida e facilmente utilizando CDs que permitem a procura de documentos, e podem armazenar até 12.000 páginas cada.



*Discos Ópticos pesam muito menos
que arquivos em papel.*

Se um sistema GED não fornecer esse tipo de portabilidade, usuários desse sistema vão achar difícil trazer seus documentos em viagens, além de transferir arquivos entre escritórios distintos. Pastas transitórias e volumes portáteis ajudam usuários a transferir seus documentos para outros escritórios ou clientes, rápida e facilmente.

Pastas Transitórias

Para usuários que possuem uma cópia do sistema de GED em seu laptop ou escritório remoto, vários sistemas permitem a usuários simplesmente mover as pastas apropriadas para uma pasta transitória e transferir essa pasta para o laptop ou sistema remoto.

Volumes Portáteis

Volumes portáteis são como enorme pastas transitórias e permitem atualizações constantes para compartilhamento do banco de dados de imagens em diferentes localizações. Essa habilidade é proveitosa para organizações que utilizam um bureau de serviços ou para aquelas com vários escritórios. Em vários sistemas de imagem de grande escala, os arquivos de documentos são armazenados em múltiplos drivers ou volumes na rede. Volumes portáteis permitem que volumes inteiros contendo imagens de documentos e texto sejam transferidos em massa para outros bancos de dados.

Distribuição de Documentos

A maioria das organizações precisa compartilhar documentos com seus parceiros de negócios ou clients. Com um sistema de papel, muitas cópias deverão ser impressas, talvez recuperadas e então enviadas. Um sistema de GED permite à organização copiar com rapidez o arquivo desejado em um CD e então enviá-lo, economizando no custo de impressão e de postagem.

Para ser mais eficiente um sistema de GED deveria oferecer Royalty-free na publicação do CD e prover um visualizador que permita as pessoas que não possuem um sistema GED de pesquisar os documentos dentro do CD distribuído.

Terceirização de Escanerização

Algumas vezes organizações acham mais rápido ou mais eficiente financeiramente ter um bureau de serviços realizando a conversão dos documentos para meio digital. Geralmente, o sistema GED é mantido pela organização e o bureau de serviço entrega regularmente CDs contendo os documentos escanerizados. Além do armazenamento das imagens e informação sobre os textos, esses CDs devem também carregar dados descrevendo os nomes dos documentos, campos de indexação, pastas, etc. Caso a organização modifique os documentos existentes e crie novos ao mesmo tempo, não é possível simplesmente sobrepor o banco de dados com o novo providenciado pelo bureau de serviços.

O sistema GED deve ser capaz de fundir os novos dados do bureau de serviços com os dados existentes da organização. Volumes portáteis fazem isso automaticamente.

Recuperação pós Desastre

Desastres podem ocorrer a qualquer momento e danificar ou destruir os documentos de uma organização. Para ajudar na recuperação após um desastre, é aconselhado manter cópias de segurança dos documentos com a ajuda de um sistema GED. Volumes portáteis que permitam a recriação de um sistema do zero podem simplificar esta tarefa.

Processamento em Lote

Organizações que lidam com vários arquivos de Organizações que lidam com vários arquivos de imagens diariamente irão rapidamente entender a importância do processamento em lote. Quando há necessidade de trazer um largo número de documentos para dentro do GED, é ineficiente processar um a um individualmente. Um sistema GED deve permitir que os documentos sejam trazidos de uma vez só para o sistema para aumentar a velocidade do processo.

A lentidão do sistema durante a escanerização ou conversão de documentos individuais devido ao recomençar/parar do processo, pode ser evitado se os documentos forem trazidos para o GED como um único grande lote. Uma vez que todas as imagens tenham sido trazidas, o sistema deve permitir aos usuários facilmente agrupá-las nos documentos apropriados antes

de indexar palavras-chave e movê-los para as pastas apropriadas. O sistema deve permitir que as páginas sejam rearranjadas, retiradas ou adicionadas a um documento para corrigir qualquer engano que possa haver ocorrido na organização do arquivo

Código de Barras

Em operações de alto volume de escanerização, separar e indexar documentos automaticamente utilizando código de barras pode salvar tempo e dinheiro. O código de barras pode ser usado para indexar documentos através da extração de campos de um banco de dados externo, através do preenchimento de campos com valores pré-assinalados, ou através da associação de certos documentos a um particular índice de fichas. O código de barras pode até funcionar como marcadores que indiquem o começo de um documento, automatizando a separação entre documentos. Enquanto que o código de barras requer alguma preparação do banco de dados, os benefícios podem ser enormes. Por exemplo, em 2.000 registros de votos, 500 indagações e 2.500 páginas de minutas legislativas tem de ser escanerizadas, marcas de código de barras podem ser colocadas em cada documento. O sistema então poderá ler cada marca, determinando o início de cada documento, assinalando o tipo correto de índice de fichas para cada documento, e preenchendo informações das fichas.

Zona OCR

Organizações que processam os mesmos formulários repetidamente podem desejar utilizar a zona de OCR para salvar tempo e memória de sistema. A zona de OCR economiza tempo pela automação da indexação de documentos, através da leitura de certas regiões (ou zonas) de um documento e depois colocando o texto no campo de índice apropriado. A quantidade de espaço para armazenamento necessário também é reduzido porque apenas as respostas que foram entradas são indexadas e processadas pelo OCR.

Para minimizar erros, o sistema deve permitir ao usuário setar um percentual mínimo de nível de acuracidade para o OCR. Se qualquer porção do formulário não atender ao padrão, o sistema deve notificar o usuário, e um membro da organização poderá ler o formulário e entrar o campo correto de informação manualmente.

Workflow

Workflow pode aumentar os benefícios de um sistema GED através do roteamento de documentos para várias pessoas. Enquanto essa funcionalidade adicional pode não ser válida durante uma fase piloto, ela se torna fundamental a medida que o sistema expande. O Workflow deve notificar automaticamente usuários específicos de eventos específicos do sistema, baseado em uma lista criada pelo administrador do sistema. Uma vez que o evento é detectado, a

comunicação é estabelecida com o servidor de e-mail existente para o envio de notificação do evento para o recipiente. O Workflow deve incluir uma série de tabelas condicionais que incluam o uso de recibo de retorno e de respostas pré-estabelecidas. Se uma tabela condicional for usada, o servidor de imagens deve enviar uma mensagem lembrete ou uma 2a. mensagem para um recipiente alternativo. Essas tabelas adicionais ajudam a eliminar gargalos e otimizar processos de negócios.

Um componente essencial em um sistema de workflow é a automação de documentos.

Um Workflow de documentos baseado em regras requer que a informação se movimente em um sistema hierárquico sem intervenção externa. O Workflow deve ser capaz de automaticamente mover, copiar ou eliminar documentos dentro do banco de dados de imagem, baseado em uma série de regras pré-determinadas. O administrador de sistemas usa a lista de regras para estabelecer os protocolos de roteamento e condições.

Qualquer ambiente de escritório pode ser considerado ideal e prático, quando se trata de automação de escritório e distribuição de trabalho. O sucesso de qualquer Workflow não está na habilidade de seguir as características exatas de roteamento ou relatórios gerados de um sistema totalmente automatizado, mas em gerenciar as exceções às regras que aparecem. Um sistema de workflow eficiente deve prover acesso completo para o roteamento de documentos e informações através da sua estrutura de pastas e sistema de segurança. Utilizando o acesso de segurança como a chave para implementação de

sistemas permite ao Administrador de Sistemas facilmente modificar os direitos de acesso da forma necessária para acomodar o que ocorre realmente em um ambiente de trabalho dinâmico. Um sistema de workflow deve oferecer aos administradores a simplicidade de copiar e arrastar, uma interface GUI simples e uma estrutura de pasta fácil de entender.

Aplicações de workflow devem ser compatíveis com ODBC para conectar o banco de dados de imagem a banco externos de terceiros e aplicações customizadas para criar um workflow completamente funcional. Como um componente final, o workflow deve prover uma segurança adequada através de relatórios da função de auditoria.

Escalabilidade

A escalabilidade do sistema determina quanto o sistema GED irá crescer com as necessidades de sua organização. Para uma escalabilidade completa, o sistema deve ter os seguintes atributos:

- Suportar até 1000 usuários concorrentes
- Armazenar pelo menos 100 milhões de páginas por banco de dados
- Arquitetura de sistema robusta
- Armazenar informação através de múltiplos drivers ou servidores
- Suporte a múltiplos bancos de dados
- Expansão para Web
- Publicar informação para CD ou DVD

Compatibilidade de Sistemas

Compatibilidade é a habilidade de um sistema de GED trabalhar com computadores e sistemas de rede existentes. Para maximizar a probabilidade dessa compatibilidade com os sistemas existentes, um sistema GED deve:

- Trabalhar com sistemas operacionais existentes como servidores Novell ou Windows NT, e Windows desktop.
- Comunicar-se usando protocolos de rede populares como IPX/SPX ou TCP/IP
- Usar uma arquitetura aberta e um banco de dados não-proprietário
- Usar uma arquitetura cliente/servidor com compressão/descompressão de imagens do lado do cliente e procura e indexação do lado do servidor para minimizar o volume de tráfego na sua rede.
- Armazenar arquivos em formatos padrão de indústria.

Sistemas de Rede

Em qualquer escritório documentos são utilizados para transmitir informações entre as pessoas. Para um sistema GED ser realmente utilizado em um ambiente de escritório, documentos devem ser acessíveis para qualquer um que tenha autorização. É importante para um sistema GED ter um repositório central de registros acessível de qualquer PC. Armazenar documentos em PC individuais prejudica o fluxo de informação entre trabalhadores e desperdiça tempo valioso e energia.

Sistemas de rede também podem suportar certas funções de imagens mais eficientemente que PCs individuais. Por exemplo, o Reconhecimento de Caracter Óptico (OCR) de uma imagem requer um grande poder computacional. Leva em torno de 10 segundos para uma típica estação de trabalho para OCR cada página de um documento. Em uma instalação GED de rede, um lote de documentos podem ser OCR através de processamento paralelo com múltiplas estações de trabalho.

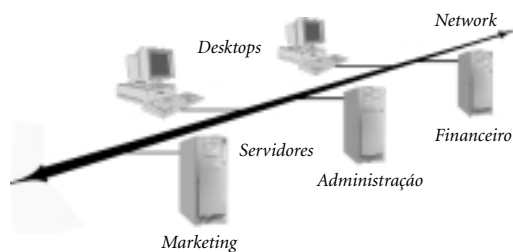
Por exemplo, uma organização pode escanear 600 páginas de um documento em um dia típico. Com uma única estação de trabalho tentando OCR todos os documentos, pode levar até 100 minutos para o computador “ler” todas as páginas. Se um servidor com o dobro de velocidade tentar OCR os documentos, poderá levar apenas 50 minutos. Entretanto, se cinco estações de trabalho forem usadas para OCR esses documentos, a tarefa pode ser finalizada em 10 minutos

Arquitetura Cliente/Servidor

Aplicações de imagens consomem recursos computacionais: os arquivos são enormes e os bancos de dados devem lidar com um grande número de registros. Com uma arquitetura cliente/servidor de verdade, tarefas como indexação, OCR e buscas são distribuídas entre o cliente (a estação de trabalho – PC) e o servidor para otimização da performance. Algumas tarefas possuem uma performance melhor no cliente, enquanto outras são mais eficientes se realizadas no servidor. Aonde tarefas específicas serão realizadas depende do sistema GED. Quando usuários de um sistema cliente/servidor querem procurar em um banco de dados, o PC cliente envia uma requisição ao servidor. O servidor então procura no banco de dados central e envia uma “resposta” de volta ao cliente. A corrupção de dados por problemas na rede é eliminada porque os arquivos nunca deixam o servidor. O tráfego da rede é minimizado porque o servidor apenas envia os documentos requisitados de volta para estação de PC.

Além do mais, funções como OCR, exibição de imagens e busca requerem um poder computacional extensivo. A arquitetura cliente/servidor se torna um requerimento quando um certo número de pessoas necessitam acessar documentos de um sistema GED. Mesmo que uma instalação comece com um projeto piloto de apenas um usuário, é importante garantir que o sistema será capaz de suportar um crescimento futuro.

Buscas podem ser realizadas mais rapidamente no servidor, que é normalmente mais poderoso que uma estação individual. No entanto, pelo tradicional “compartilhamento de arquivos” de sistemas GED, uma cópia do banco de dados é enviada através da rede para o PC, e esse performa as buscas. Esse método leva a: (a) corrupção dos dados devido à problemas na rede; (b) aumento do tráfego na rede, por exemplo, o banco de dados possui 800MB em tamanho; e (c) tempo de busca depende da velocidade da estação PC. Sistemas de compartilhamento de arquivos podem ser mais baratos para começar, mas suas limitações restringem expansão futura e flexibilidade.



Exemplo de um sistema Cliente-Servidor

Formatos de Arquivo Não Proprietários

Preocupação com a capacidade de leitura futura faz com que vários gerentes de registros hesitem em implementar um sistema GED. Com as rápidas mudanças na indústria de computadores, é difícil prever o que acontecerá com os computadores em dez ou vinte anos. Entretanto, a necessidade de recuperação rápida

e melhoria no gerenciamento de registros significa que muitos escritórios necessitam encontrar uma solução GED hoje.

Para endereçar essas preocupações, sistemas GED devem usar formatos de texto e imagem não-proprietários. Com os exemplos de processadores de texto mostram, documentos salvos em WordStar, WordPerfect antigo ou até o antigo MS Word já são difíceis de ler. Sendo que cada empresa de processadores de palavras utiliza formatos proprietários para seus documentos, adquirir o mais recente software para ler um formato antigo pode ser frustrante ou uma tarefa cara. O mesmo se aplica no mundo das imagens.

Os formatos não-proprietários disponíveis para armazenar informação de documentos são poucos, mas estáveis. ASCII tem sido um padrão para informação de texto desde 1963 e se tornou um bloco de construção básico para praticamente qualquer programa envolvendo texto. TIFF tem sido usado como um padrão, um formato gráfico não-proprietário desde 1981. É usado amplamente para transmitir informação de documento através de sistemas de imagens, máquinas de fax e software, e por projetistas.

Dado que os padrões ASCII e TIFF tem prevalecido desde então, compradores de sistemas se sentem confortáveis, e não importa qual novo paradigma apareça no futuro, os desenvolvedores de um novo formato terão um interesse único em prover a conversão para esses padrões. Com formatos de documento proprietário, não há essa garantia.

Perguntas Mais Frequentes

Geral

P: O que é um “documento?”

R: Um documento pode ter de uma a várias páginas, e pode incluir imagens e/ou texto, além de anotações, e fichas (cartão de índice)

P: Eu posso editar ou alterar imagens?

R: Um sistema de imagens não deve fornecer nenhuma ferramenta para edição ou alteração de imagens. Isso é importante pois muitos usuários consideram que imagens não devem ser tocadas e que qualquer alteração irá determinar a integridade do sistema. Além do mais, o sistema deve fornecer uma auditoria para manter registros de quais usuários acessaram quais documentos e quando.

P: Sistemas de imagem suportam auditoria?

R: A auditoria de um sistema de imagem deve gravar o nome de um usuário, data, hora, nome do documento e a ação quando o usuário acessar o banco de dados ou documento. Vários níveis de detalhes de registros de auditoria e acompanhamento de atividades devem estar disponíveis. O sistema deve ter capacidade de selecionar esses registros por classificação e filtragem.

P: O que é ICR (Reconhecimento Inteligente de Caracter)?

R: ICR é o reconhecimento de um texto escrito a mão e também é conhecido como reconhecimento de escrita a mão. Textos escritos a mão são mais difíceis para o computador reconhecer e os resultados possuem uma taxa de erros maior do que os textos impressos. Reconhecimento de ICR geralmente atua melhor em texto confinado, o que significa letras impressas em blocos com uma letra em cada caixa. Um reconhecimento preciso de uma escrita a mão requer que a ferramenta de ICR seja treinada para reconhecer o estilo de escrita de cada usuário.

P: O que é OMR (Reconhecimento de Marca Óptica)?

R: OMR, também chamado de Reconhecimento sensível a marcas, é o reconhecimento de marcas comumente usados em formulários, como múltipla escolha, círculos, e bolhas. OMR pode ser uma parte importante do sistema para organizações que processam muitos formulários padrão. Pesquisa de consumidor é um exemplo do OMR em ação.

P: Qual é o formato padrão usado para armazenar imagens?

R: Imagens em preto e branco são mais comumente armazenadas em arquivos padrão TIFF usando CCITT compressão grupo 4 (duas dimensões). Escala em cinza e imagens colori-

das são frequentemente armazenadas em arquivos TIFF com compressão JPEG.

P: Quais os tipos de sistemas operacionais são normalmente suportados?

R: A maioria dos sistemas GED possuem aplicações cliente que podem rodar aplicações em Windows 95, 98 e Windows NT. Sistemas de Internet/Intranet podem rodar em plataformas adicionais, como Macintosh, Unix, dentre outras.

P: Quanto espaço em disco um sistema GED tipicamente requer?

R: O software no cliente normalmente requer cerca de 20MB de espaço livre em disco para instalar. O Software no servidor requer apenas cerca de 5MB de espaço livre em disco no servidor, e 15MB para o banco de dados. Espaço adicional pode ser necessário para o banco de dados real.

P: E se o meu banco de dados for muito grande para caber em um volume de dados?

R: Um sistema GED irá permitir que dados e imagens sejam armazenados através de múltiplos volumes com cada volume residindo em um diretório diferente ou em um drive, disk array, CD ou MO diferente.

P: Quanto de RAM um sistema GED necessita?

R: O software no cliente geralmente requer de 16 a 20MB de RAM para rodar, com requerimentos maiores para escanização e OCR. A maioria dos sistemas recomenda ter 64MB ou mais.

P: Há necessidade de placa gráfica ou monitor especial?

R: A maioria dos sistemas trabalha com qualquer placa gráfica compatível com Windows e monitor VGA (ou melhor), e é recomendável que se use um monitor de pelo menos 15" com resolução de 800 x 600 dpi pelo menos.

Escanerização/ Importação/ Armazenamento

P: Quais os fabricantes de scanners para GED?

R: Alguns dos top fornecedores de scanner incluem Ricoh, Fujitsu, Panasonic, Bell & Howell, Canon, Hewlett Packard e Avison. Scanners para sistemas de GED possuem alimentadores de alta velocidade para escanizar grandes quantidades de documentos.

P: Quais são as interfaces de hardwares e softwares para scanners mais comuns?

R: Driver ISIS para scanner da Pixel Translations e Kofax Image Controls (<http://www.kofax.com>) são as interfaces de scanner mais comuns. Muitos scanners conectados a uma placa Adaptec SCSI ou à placa Kofax de processamento de imagem TWAIN não é geralmente uma boa interface de scanner para alta velocidade e alto volume de escaneamento. Enquanto que a maioria dos scanners incluem o driver TWAIN, cada driver é escrito por uma companhia diferente e trabalha um pouco distinta. Se um scanner suporta uma interface TWAIN apenas, ele é provavelmente desenhado para um baixo volume de escaneamento e editoração, não para GED.

P: Como eu posso escanerizar cheques?

R: Várias empresas, incluindo Visionshape, fabricam escaners especialmente desenhados para cheques que lêem o número MICR codificado magneticamente na parte de baixo do cheque. Se você não tem um desses scanners, a maioria dos cheques pode ser escanerizado com um scanner regular para GED e pela função OCR normalmente, apesar de que os números MICR não serão lidos.

P: Como eu posso escanerizar documentos de formato grande?

R: Vários fornecedores, incluindo Context, Océ and Calcomp, fabricam scanners especificamente projetados para documentos de formato grande, até tamanho E(34"x 44") e A-0 (33"x 46.8"). Se você não tem um desses, o documento pode ser reduzido em tamanho através de uma copiadora e depois escanerizado com um escaner normal, ou enviando para um bureau de serviços que tenha scanner para formato grande.

P: Qual resolução de imagem eu devo usar?

R: A maioria dos sistemas GED podem suportar documentos escanerizados em várias resoluções, de 50 a 600 dpi (ou mais), dependendo do seu scanner. Dependendo do propósito e do tamanho da página, a maioria dos documentos são escanerizados em preto e branco com 200, 240 ou 300 dpi.

P: E arquivos coloridos ou fotografias?

R: Sistemas GED devem suportar imagens em preto e branco, cinza e colorida. Arquivos coloridos podem ser escanerizados com um scanner colorido, importados ou convertidos para o sistema GED. Fotografias tendem a escanear melhor em escala cinza, uma técnica de cópia com "pixel decimation" que ocorre quando a imagem é mostrada em uma tela com baixa resolução. As bordas dos caracteres podem se tornar irregulares com pontos insuficientes. A escala cinza simplesmente preenche esses pontos marginais com cinza para melhorar a imagem, fazendo com que apareçam fáceis de ler. Sendo que é apenas uma melhoria na visualização, não há melhoria no OCR ou na qualidade de impressão.

P: Como eu posso escanerizar documentos de dois lados?

R: Um sistema GED deve prover dois modos diferentes de se fazer isso. Ele deve suportar scanner duplex, que simultaneamente escaneriza ambos os lados da página. Com um scanner simplex, o usuário deve poder escanerizar todas as páginas da frente, virar os documentos e escanerizar o verso da página, e o sistema deve automaticamente colocar as páginas na ordem correta.

P: Posso escanerizar páginas landscape e portrait juntas em um lote apenas?

R: A maioria dos sistemas permite que você troque a orientação das páginas à medida que você escaneriza ou após a escanerização. Muitos sistemas também incluem a opção de checar automaticamente e corrigir a orientação das páginas.

P: Como imagens “enviesadas” são tratadas?

R: Imagens enviesadas (curvas ou inclinadas) podem afetar a precisão do processo OCR, de forma que sistemas GED incluem software que reconhecem imagens enviesadas e as compensam. Isso é particularmente importante quando escanizando cortes pressionados em um scanner flat bed ou quando escanizando documentos através de um ADF (alimentador de documento automático) usado ou projetado provavelmente.

P: Quais formatos de arquivos a maioria dos sistemas importa?

R: A maioria dos sistemas pode importar texto ASCII em 20 ou mais formatos gráficos, incluindo TIFF 4, TIFF 3, TIFF Raw, TIFF LZW, PCX, BMP, CALS, JPEG, GIF, PICT, PNG e EPS visualização de imagem. A maioria dos sistemas converte arquivos gráficos em preto e branco para o formato Grupo 4 TIFF durante a importação, e imagens cinzas/coloridas para TIFF JPEG.

P: Quanto espaço de armazenamento eu necessito?

R: Documentos escanizados tipicamente requerem cerca de um gigabyte para cada 17.000 páginas, assumindo papel tamanho carta, resolução 300 dpi, e escanização em preto e branco. Isso inclui o espaço para o texto

gerado pelo OCR. Após a compressão, uma imagem típica requer 50KB, apesar de poder variar de 16KB a 150KB, dependendo do tamanho do papel, resolução e complexidade da página. Um texto OCRizado tipicamente requer 2KB de espaço por página.

P: E se o meu banco de dados for muito grande para caber em uma localidade?

R: Vários sistemas GED permitem que imagens sejam armazenadas através de múltiplos volumes com cada volume residindo em um diretório diferente ou em um drive, disk array, CD ou MO diferente.

P: Qual é a diferença entre CD ou jukeboxes/changer de DVD e torres?

R: Em um jukebox/changer, existe mais slots e discos do que drivers. Mecanismos de robôs automaticamente colocam o disco correto em um dos drivers quando o disco é necessitado. Em uma torre, vários drivers de CDs ou DVDs estão empilhados juntos em uma única unidade, e cada disco está sempre em um drive. Torres provêm um rápido acesso ao dado, mas custam muito mais por disco e não carregam muitos discos. Jukeboxes/changers custam menos por disco e podem carregar até 500 discos, mas são mais lentos devido ao consumo de tempo na troca de discos nos drivers.

Visualização / Impressão / Exportação

P: Posso ver combinações de imagens, textos e campos de indexação lado a lado?

R A maioria dos sistemas GED permite que a tela de visualização seja configurada para mostrar qualquer uma das duas a qualquer hora: texto, imagens, campos de indexação, imagens pequenas ou documentos com várias páginas ou nada. O tipo de fonte e o tamanho do texto pode ser alterado para servir às necessidades individuais.

P: Posso abrir e visualizar mais de um documento ao mesmo tempo?

R: A maioria dos sistemas GED permite a visualização de vários documentos, sendo a memória disponível a única limitação para o número de documentos abertos ao mesmo tempo.

P: Como posso fazer uma nova sequência de páginas?

R: Se as páginas estão fora de ordem e necessitam ser re-sequenciadas, a maioria dos sistemas permite vistas das páginas de um tamanho pequeno ("Thumbnails") para serem simplesmente arrastadas para a posição requerida. Da mesma forma, páginas individuais podem ser selecionadas e apagadas, sujeitas ao controle de acesso de segurança apropriados e privilégios.

P: Eu vou precisar de um monitor especial para visualização de imagens?

R: Não, a maioria dos sistemas roda perfeitamente bem em monitores padrão VGA e melhores. Um display de 15" usando uma controladora Super VGA pode ser considerado o mínimo display prático para um usuário ad hoc do sistema. Usuários frequentes devem usar um monitor de 17", e usuários cujo trabalho é escanear e rever documentos com imagens o tempo todo podem considerar um monitor de 19" ou 21".

P: Qual a vantagem de um monitor grande para usuários com poder?

R: Para pessoas que usam um sistema GED intensivamente, o tamanho da tela pode ser um fator crítico. Se os usuários quiserem mover entre as páginas com a facilidade do papel, eles tem que ser capazes de ver a página inteira de uma vez de tal modo que o texto seja visível. Se uma página de 8 1/2" x 11" é o tamanho dominante de papel, então um monitor de 21" capaz de mostrar 1600 x 1200 será ótimo. Usando um monitor padrão VGA de 14" irá necessitar movimentar pelo documento para visualizar a imagem toda no seu tamanho normal.

P: O que é importante além do tamanho do monitor?

R: A resolução da tela e a velocidade de varredura do monitor também são importantes. Geralmente, quanto maior o tamanho do monitor e maior sua resolução, mais difícil será obter a maior velocidade de varredura que é requerida para sustentar a visualização sem que a tela cintile. O melhor ponto inicial para a menor cintilação de tela é geralmente considerada a taxa de varredura horizontal de 72MHz em um monitor de 21". A taxa de varredura máxima é uma função do monitor e da controladora gráfica.

P: Eu vou precisar de uma impressora especial para imagens ou texto que utilizaram o OCR?

R: Geralmente não. A maioria dos sistemas GED suporta a maioria das impressoras compatíveis com Windows, mas se recomenda que utilize uma impressora laser com no mínima 4MB de memória RAM. Se você está usando um sistema instalado na rede e imprimindo um alto volume de páginas em uma impressora de rede, você pode considerar instalar uma impressora laser localmente ou no seu próprio segmento de rede para minimizar o tráfego da rede durante impressões pesadas de várias páginas

P: Em quais formatos posso exportar documentos?

R: A maioria dos sistemas GED pode exportar imagens para vários formatos gráficos incluindo TIFF 3, TIFF 4, TIFF Raw, BMP, GIF, CALS e JPEG, e C.

OCR: Reconhecimento de Caracter Óptico

P: O que é OCR?

R: OCR significa Reconhecimento de Caracter Óptico, é como um computador que converte palavras de uma imagem escaneizada sem capacidade de busca em um texto com capacidades de busca. OCR é normalmente necessário para se usar indexação de texto e buscas, e deve estar incluído em um sistema GED. Ferramentas OCR podem geralmente reconhecer textos digitados ou impressos, não escritos à mão..

P: O quanto o OCR é preciso?

R: A precisão em uma página recém impressa a laser é tipicamente melhor que 99.6%. A precisão em documentos enviados por fax, sujos ou degradados será naturalmente menor, mas a maioria dos sistemas GED possuem tecnologia de limpeza de imagem que pode melhorar a precisão do OCR.

P: Eu terei que corrigir os erros de OCR?

R: Não se o sistema GED suportar lógica "fuzzy", que encontrará palavras mesmo se a ferramenta OCR tiver feito alguns enganos.

P: Qual a velocidade do processo de OCR?

R: A performance do OCR e processos de indexação é totalmente dependente de fatores tais como a velocidade e configuração do sistema e do conteúdo da imagem. Um Pentium 133MHz geralmente necessita aproximadamente de 6 segundos por página, enquanto que um Pentium II 450MHz levará cerca de 2 a 3 segundos por página.

P: Qual a diferença entre OCR e indexação?

R: OCR é processo de conversão de imagens escanizadas para texto, e indexação é o processo de fazer o arquivo texto resultante fácil e rápido de ser pesquisado..

P: Um texto que utilizou OCR pode ser exportado e reutilizado em um processador de palavras?

R: Sim, você pode copiar textos entre o sistema GED e uma aplicação Windows, ou você pode exportar arquivos de textos inteiros (todas as páginas de um documento) para um diretório e abri-lo com seu processador de textos favorito.

P: Posso corrigir manualmente erros de OCR?

R: A maioria dos sistemas permite que os usuários corrijam os erros de OCR dentro do sistema. Entretanto, com capacidades de busca da lógica fuzzy, isso normalmente é desnecessário.

COLD: Computer Output to Laser Disc

P: Qual a diferença entre COLD e GED?

R: GED realiza a escanização, compressão, armazenamento, indexação, funções de OCR, busca e recuperação de milhões de páginas de documentos em papel ou documentos eletrônicos armazenados como imagens permanentes. COLD realiza armazenamento, indexação, busca e impressão de relatórios de enormes arquivos de texto gerados por mainframes, mini-computadores e aplicações de sistemas antigos em grande volumes de impressão. COLD armazena arquivos de texto enormes e extrai chaves de indexação em disco rígidos, cartuchos ópticos ou CD-ROM ao invés de imprimir toda a informação em papel ou armazená-la em microfilme.

P: Quantos campos de indexação pode um servidor COLD extrair de cada relatório?

R: O número de campos de indexação é normalmente ilimitado. Entretanto, o maior número de campos extraídos de cada relatório, mais lento o processo de extração irá rodar e maiores serão os arquivos de indexação.

Glossário de Termos

ADF

Alimentador automático de documento. Esse é o meio pelo qual um scanner alimenta um documento de papel.

Anotações

São as mudanças ou adições realizadas em um documento utilizando notas adesivas, um marcador, ou outra ferramenta eletrônica. Imagens ou textos em um documento podem ser marcadas em cores diferentes, carimbadas (por exemplo “CONFIDENCIAL” ou “FAX”), ou ter notas adesivas eletrônicas anexadas. Anotações devem ser sobrepostas e não alteram o documento original.



Exemplos de carimbos, anotações e highlight

Armazenamento Óptico Magnético

Um drive que combina tecnologia laser e magnética para criar armazenamento apagável de alta capacidade.

Arquitetura Cliente-Servidor vs Compartilhamento de Arquivos

Duas arquiteturas comuns de aplicações de software encontradas em rede de computadores. No compartilhamento de arquivos, todas as buscas ocorrem na estação de trabalho, enquanto que o banco de dados de documentos reside no servidor. Na arquitetura cliente-servidor, processos intensivos de CPU (tais como buscas e indexação) são realizados no servidor, enquanto que a visualização da imagem e o OCR ocorrem no cliente. Aplicações de compartilhamento de arquivos são fáceis de desenvolver, mas tendem a gerar um tremendo tráfego de dados na rede. Além disso, há a exposição do banco de dados à corrupção devido às interrupções nas estações de trabalho. Aplicações cliente-servidor são mais difíceis de desenvolver, mas reduzem dramaticamente o tráfego de dados na rede e isolam o banco de dados de interrupções nas estações de trabalho.

ASCII - American Standard Computer Information Interchange

Usado para definir o texto para computador que foi criado como um conjunto de 255 caracteres. ASCII tem sido usado como um padrão, um formato de texto não-proprietário desde 1983.

Bitmap/Bitmapped

Veja Raster/Rasterized

BMP

Um formato nativo do Windows para armazenar imagens chamadas “bitmaps”.

Browser de Pastas

Um sistema usado para organizar documentos (geralmente hierárquico or “empilhado”). Por exemplo, o programa de Gerenciamento de Arquivos no Microsoft Windows é um tipo de Browser de Pastas que mostra os diretórios no seu disco.

Caching (de Imagens)

Armazenamento temporário de arquivos de imagem em um disco rígido para uma posterior transferência para um armazenamento permanente, como um óptico ou jukebox de CD.

Campo Chave

Campos de banco de dados usados para busca e recuperação de documentos. Sinônimo para campo de indexação.

Campos de Indexação

Campos de banco de dados usados para categorizar e organizar documentos. Frequentemente definidos pelo usuário, esses campos podem ser usados para buscas.

CD-R

Abreviação para CD Regravável. Significa um CD que pode ser escrito (ou gravado) apenas uma vez. Pode ser copiado para distribuição de um grande volume de dados. CD-Rs podem ser lidos em qualquer driver de CD, em um computador isolado ou em uma rede. Isso torna o intercâmbio entre os sistemas mais fácil.

CD-ROM

Compact Disc Read Only Memory. Gravado em grande escala e não em um gravador de CD (CD writer), é uma mídia de armazenamento em disco óptico popular para armazenar arquivos de computador, assim como gravação de músicas digitais.

Código de Barras

Um modelo de linhas verticais que é lido por um laser ou um scanner óptico, e que corresponde a um registro em um banco de dados. Como um componente adicional de um sistema GED, essa característica foi desenhada para aumentar a velocidade na qual um documento pode ser armazenado.

COLD

Computer Output to Laser Disk. Processo de programação que gera registros eletrônicos e relatórios impressos em disco laser ao invés de uma impressora. Pode ser usado no lugar do COM (Computer Output to Microfilm) ou impressão de relatórios em formulários contínuos.

COM

Computer Output to Microfilm. Um processo que gera registros eletrônicos e relatórios de computador em microfilmes.

CPU

Central Processing Unit (Unidade de Processamento Central). É o “cérebro” do computador..

De-shading

Remoção de áreas de sombras para reconhecimento da imagem mais fácil pelo OCR. Um software de-shading tipicamente procura por áreas com um padrão regular ou pequenos pontos.

De-skewing

O processo de endireitar imagens enviesadas (descentralizadas). De-skewing é uma das melhorias de imagem que pode aumentar a precisão do OCR. Documentos muitas vezes ficam enviesadas quando enviados por fax ou escaneizados.

De-speckling

Remoção de salpicos isolados de um arquivo de imagem. Geralmente ocorre quando um documento é enviado por fax ou escaneizado.

Discos Ópticos

Mídia de computador similar ao disco compacto que não pode ser escrito. Um disco óptico usa laser para ler o dado armazenado.

Dithering

O processo de conversão da cor cinza em duas densidades de pontos preto, geralmente com o propósito de impressão ou armazenamento de imagens coloridas ou acinzentadas em imagens branco e preto.

Drag-and-Drop

O movimento de objetos na tela do computador (monitor) quando arrastados com o mouse.

Drive de CD-ROM

O movimento de objetos na tela do computador
Um drive de computador que lê discos compactos.

Drive Óptico Apagável

Um tipo de drive óptico que usa discos ópticos apagáveis.

Drivers de Scanner ISIS e TWAIN

Digital Video Disc or Digital Versatile Disc. Um Aplicações especializadas usadas para comunicação entre escaners e computadores. Drivers TWAIN foram desenvolvidos para edição de fotos e publicação. Eles tratam imagens em cores e em escala em cinza muito bem, mas não suportam escanização em alta velocidade. Drivers ISIS foram desenvolvidos para escaneização de imagens em alta velocidade. Foram desenhados para rápida escanização de imagens em preto e branco através do ADF. Nos últimos anos, a diferença entre eles tem diminuído e drivers ISIS agora incluem suporte a cores e em escala em cinza, enquanto drivers TWAIN agora suportam ADF.

DVD

Digital Video Disc or Digital Versatile Disc. Um disco de plástico, como um CD, aonde o dado pode ser escrito e lido. DVDs são mais rápidos, podem armazenar mais informações e suportam mais formatos que os CDs.

EDM – Eletronic Document Management

Software de Imagem que gerencia documentos eletrônicos

Escala de cinza

Uma opção para visualizar um imagem em preto e branco de um modo aperfeiçoado, tor-

nando a visualização mais fácil. Uma visualização em escala de cinza utiliza sombreamento em cinza para preencher os vazios ou furos que ocorrem quando a imagem é visualizada em uma tela de computador. Também conhecido como escala acinzentada.

Escalabilidade

Capacidade de expansão de um sistema sem requerer maiores reconfigurações ou reentrada de dados. Múltiplos servidores ou armazenamento podem ser adicionados facilmente.

Fichas

Conjunto de campos de índices para documentos.

Formato ISO 9660 CD

O Formato da Organização de Padrões Internacionais para criação de CD-ROMs que podem ser lidos no mundo todo.

GED (Gerenciamento Eletrônico de Documentos)

Software usado para armazenar, gerenciar, recuperar e distribuir documentos rapidamente e facilmente no computador.

Gerenciamento Hierárquico de Armazenamento

Um software que transfere automaticamente arquivos on-line para mídia de armazenamento “near-line”, geralmente com base na frequência de uso dos arquivos.

GIF

Padrão CompuServer para arquivos de imagem.

Gigabyte

Um bilhão de bytes. Também expresso como mil megabytes. Em termos de capacidade de arquivamento de imagem, um gigabyte equivale aproximadamente a 17.000 páginas no formato de 8 1/2"x11" escanerizadas a 300 dpi, armazenadas como imagens TIFF Grupo IV.

Gravação (CDs ou DVDs)

Gravação ou escrita de dados em um CD ou DVD.

Habilitador de Imagem

Uma função do software que cria conexões entre aplicações existentes e imagens armazenadas.

ICR

Reconhecimento Inteligente de Caracteres. Um processo que reconhece textos escritos a mão e impressos como caracteres alfanuméricos.

Interface de Scanner SCSI

O dispositivo usado para conectar um scanner a um computador.

Interface de Video Scanner

Um tipo de dispositivo usado para conectar scanners com computadores. Scanners com essa interface requerem uma placa de controle desenhada pela Kofax, Xionics ou Dunord.

IPX/SPX

Protocolo de comunicação usado pela rede Novell.

JPEG

Um formato de compressão de imagem usado para armazenar fotografias coloridas e imagens.

Jukebox

Um dispositivo de armazenamento de discos ópticos e os carrega em um drive.

Jukebox Óptico

Veja "Jukebox."

Lógica Booleana

O uso dos termos "E", "OU" e "NÃO" durante buscas de palavras ou textos. Usado para ampliar ou restringir o escopo da busca.

Lógica "Fuzzy"

Um procedimento de procura para uma palavra exata ou similar ao critério de busca, compensando erros de escrita e interpretação da função OCR.

MAPI

Mail Application Program Interface. Esse software padrão de Windows tornou-se uma interface de e-mail popular e é usado pelo MS Exchange, GroupWise e outros pacotes de e-mail.

Módulo Carregável NetWare (NLM)

Aplicação que roda como parte de um sistema operacional de rede (NOS – network operating system) de um servidor Novell Netware.

Mudança de Fase

Método de armazenamento de informação em discos ópticos regraváveis.

Near-Line

Documentos armazenados em discos ópticos ou discos compactos que se encontram o jukebox ou no trocador de CD, e podem ser recuperados sem intervenção humana.

NT

Network Technology. Se refere ao servidor Microsoft Windows NT e software de estação de trabalho.

OCR

Optical Character Recognition. Um processo de software que reconhece textos impressos como caracteres alfanuméricos.

Off-Line

Arquivo de documentos armazenados em discos ópticos ou discos compactos que não estão conectados ou instalados no computador, e requerem intervenção humana para serem acessados.

On-Line

Documentos armazenados em um disco rígido ou disco magnético em um computador e que estão disponíveis imediatamente.

Pastas Transitórias

Um método para simplificar o transporte de um grupo de documentos de um computador para outro.

Pixel

Um único ponto em uma imagem fotográfica. Pode ser preto e branco, cinza ou colorido.

Placa de Processamento de Imagem (IPC)

Uma placa localizada no computador, scanner ou impressora que facilita a aquisição e visualização de imagens. A função primária da maioria dos IPCs é a rápida compressão de descompressão de arquivos de imagem.

Processador de Formulários

Uma aplicação de imagem especializada em formulários pré-impressos. Sistemas de processamento de formulários geralmente usam ferramentas OCR de ponta (ou várias) e elaboram rotinas de validação de dados para extração de formulários escritos a mão ou com qualidade muito ruim. Esse tipo de aplicação de imagem possui vários desafios, sendo que muitos documentos escaneizados não foram desenhados para sistemas GED ou OCR.

Processamento em lote

O nome da técnica usada para a entrada de uma enorme quantidade de informação de uma única vez, é processo oposto entrada individual de dados.

Procura e indexação de textos

Possibilita a recuperação de documentos através de uma palavra ou frase. Todas as palavras no documento são indexados em uma lista mestre de palavras com apontadores para os documentos e páginas onde cada ocorrência da palavra ou frase aparecem.

Publicação em CD

Uma alternativa para copiar um grande volume de documentos em papel. Esse método envolve acoplar imagens e textos de um documento em um software de visualização em CDs. Algumas vezes softwares de busca são incluídos nos CDs para melhorar as capacidades de busca.

Publicação na Internet

Software de imagem especializado que permite que um grande volume de documentos em papel seja publicado na Internet ou Intranet. Esses arquivos podem ser disponibilizados para outros departamentos, colegas de outras localidades ou público geral para busca, visualização ou impressão.

RAID

Redundant Array of Independent Disks. Uma coleção de discos rígidos que atua como uma unidade única. Arquivos em drives RAID podem ser duplicados (“espelhados”) para preservar os dados. Sistemas RAID podem variar em níveis de redundância, sendo: sem redundância - nível único, disco sem-espelhamento - nível 0, dois discos que se espelham um no outro - nível 1, até o nível 5, o mais comum.

Raster/Rasterized

Um método de representação de imagem com uma grade (ou “mapa”) de pontos ou pixels. Formatos típicos de arquivos raster são GIF, JPEG, TIFF, PCX, BMP, etc ..

Região (de uma imagem)

Uma área de um arquivo de imagem que é selecionado para um processamento específico. Também chamado de zona.

Scanner

Um dispositivo de entrada normalmente usado para converter documentos em papel em imagens de computador. Scanner também pode escanear microfilmes e microfichas.

Scanners Duplex v. Scanner Double-Sided

Scanner duplex automaticamente escanear ambos os lados de uma página de dois lados simultaneamente, produzindo duas imagens de uma vez. Scanner Double-Sided utilizam um único scanner para escanear páginas de dois lados, escanear uma pilha ordenada de papéis, depois virando e escanear o outro lado.

Scanner Flatbed

Scanner com superfície plana que permite escanear livros e outros documentos.

SCSI - Small Computer Systems Interface.

Um padrão para conectar periféricos a computadores. Um SCSI permite que até 7 dispositivos sejam conectados por cabos. O atual padrão SCSI é o SCSI II, também conhecido como SCSI rápido..

SQL

Structured Query Language. Um padrão popular para rodar buscas em banco de dados (queries) e relatórios.

Taxa de Compressão

É a relação entre o tamanho de um arquivo comprimido e o de um não comprimido. Por exemplo, uma taxa de compressão de 20:1 significa que um arquivo de 1MB não comprimido tem o tamanho de 50KB quando comprimido.

TCP/IP

Protocolo de comunicação de redes. Esse é o protocolo usado pela Internet.

Thumbnails

Versão pequena de uma imagem usada para uma vista geral rápida ou ter-se uma idéia do que a imagem aparenta.

TIFF

Tagged Image File Format. Um formato gráfico raster de imagem não-proprietário que possui vários formatos de compressão. TIFF tem sido usado desde 1981.

TIFF Grupo III (compressão)

Um formato de compressão unidimensional de armazenamento de imagens em branco e preto que é utilizado pela maioria de máquinas de fax.

TIFF Grupo IV (compressão)

Um formato de compressão bidimensional de armazenamento de imagens em branco e preto. Tipicamente compressa em uma taxa de 20-1 para documentos padrão.

Volumes Portáteis

Uma característica que facilita a movimentação de um grande volume de documentos sem a necessidade de copiar múltiplos arquivos. Volumes portáteis permitem que CDs individuais sejam facilmente reagrupados, destacados e reatados a diferentes bancos de dados para uma maior troca de informações.

Workflow, Ad Hoc

Um processo manual simples através do qual documentos podem ser transportados em um sistema GED de multi-usuários baseado nas necessidades reais.

Workflow, Baseado em regras

Uma série programada de passos automatizados que direcionam documentos para vários usuários em um sistema GED

WORM Disks

Write Once Read Many Disks. Uma mídia popular dos anos 80 para armazenamento de arquivos. Reconhecido como o primeiro disco óptico, é usado para armazenar arquivos de dados que não podem ser alterados. Discos WORM são criados por PCs isolados e não podem ser usados em rede, como os CD-Rs.

ZIP

Um formato padrão de compressão de arquivos que permite rápido e fácil armazenamento para transporte.

Zona OCR

Uma característica adicional de um sistema GED que preenche ficha de documentos através da leitura de certas regiões ou zonas de um documento, e depois coloca o texto no campo de indexação do documento.

Sobre o Autor

LaserFiche desenvolveu esse guia para prover uma idéia geral sobre sistemas de Gerenciamento Eletrônico de Documentos (GED). Nós identificamos a necessidade de um recurso que fosse certo para pessoas que estão aprendendo sobre sistemas de imagem, aqueles que estão se preparando para comprar um sistema, e para todas as pessoas que estiverem interessadas nesta tecnologia.

Desde 1987 LaserFiche tem sido pioneira em sistemas de armazenamento e recuperação de um alto volume de documentos. O sistema GED da LaserFiche tem ajudado a gerenciar documentos em mais de 15,000 escolas, escritórios de advocacia, empresas de seguro e outros tipos de negócios ao redor do mundo, incluindo milhares de municípios, estados e agências federais. O principal recurso da LaserFiche é o grupo de profissionais excepcionalmente competentes e experientes, com um vasto conhecimento tanto teórico como prático dos aspectos da automação de escritório. Nossos analistas e programadores já possuíam experiência prévia na indústria e nos negócios nas áreas de desenvolvimento de programas, consultoria e soluções customizadas. Para cada projeto, LaserFiche empenha-se para oferecer:

- Objetividade, competência e imparcialidade
- Experiência, através da análise e soluções similares de problemas para outras organizações.
- Entendimento, das complexidade e relacionamento entre seres humanos.

- Tempo, para concentrar sem interrupção até obter a solução de um problema.

A experiência de trabalho diretamente com profissionais de gerenciamento de arquivos ensinou muito a LaserFiche sobre o que funciona, o que funciona melhor, e o que funciona melhor ainda. Diferentemente de empresas que possuem conhecimento apenas na tecnologia, LaserFiche combina o seu conhecimento em sistemas com a capacidade da tecnologia para criar um ambiente de trabalho melhor.

LaserFiche é uma divisão da CompuLink Management Center, Inc. Compulink é certificada em WBE e MBE, e possui uma história de sucesso em assistir organizações a estabelecer sistemas de gerenciamento eletrônico de documentos.

LaserFiche agradece a oportunidade de responder em detalhes qualquer questões sobre GED e de demonstrar o sistema LaserFiche.

LaserFiche Document Imaging

3545 Long Beach Blvd.
Long Beach, CA 90807
USA.

(310) 793-1888
(800) 985-8533
(310) 793-8531 fax
www.laserfiche.com

- **Notas** -