

Investigação de uma Arquitetura para Gerenciamento de Documentos

Sérgio Francisco Tavares de Oliveira Mendonça¹, Edson Alves de Carvalho Júnior¹, João Ferreira da Silva Júnior¹

¹Unidade Acadêmica de Garanhuns, Universidade Federal Rural de Pernambuco (UFRPE)

Av. Bom Pastor, s/n – Boa Vista – 55.292-270 – Garanhuns – PE, Brasil

²Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco (UFRPE)

Rua Dom Manoel de Medeiros, s/n, Dois Irmãos – 52.171-900 – Recife – PE – Brasil

sftom@uag.ufrpe.br, edacjr@msn.com, joaoferreirape@gmail.com,

Resumo. *Um dos grandes desafios no gerenciamento de documentos consiste na identificação adequada, através da especificação e extração da informação, através de diversas abordagens e características. Definir e implementar técnicas para aquisição, pré-processamento, extração de características e reconhecimento de imagens de documentos. Comparar as técnicas implementadas e, dessa forma verificar a adequação das mesmas no ambiente proposto. O resultado prático, desse trabalho, servirá de protótipo para sistemas de análise de imagens de documentos.*

Abstract. *A major challenge in document management is the proper identification, through the specification and information extraction, using a variety of approaches and features. Define and implement techniques for acquisition, preprocessing, feature extraction and recognition of document images. Compare the techniques implemented and thus verify the adequacy of the same in the proposed environment. The practical result of this work will serve as a prototype for systems analysis of document images.*

1. Introdução

O atual sucesso comercial em toda empresa de sistemas de gestão de documentos é um indicativo da necessidade de adoção de técnicas e estratégias semânticas para a especificação e extração de informações em um conjunto de documentos. A fim de tratar documentos de forma mais eficaz, departamentos introduziram conceitos de redes locais (LANs) nos anos 1980, permitindo aos usuários compartilhar um espaço de arquivo comum. Produtos de *groupware*, como o Lotus Notes facilita o fluxo de informações nas empresas, mas a forma da informação é mais do que uma base de dados que apoia documentos. Outros produtos que trabalham com a gestão de documentos, como Excalibur e Autonomia começaram a meta-estruturar documentos, fornecendo a possibilidade de realizar pesquisa em todo o documento, através de espaços heterogêneos. Embora semanticamente adequado, sistemas de recuperação de tais informações negligenciam a importância do documento para a organização, sabendo pouco sobre a razão de existência do documento.

Uma solução efetiva adotada por grande número de empresas tem sido a gestão documental, que deve considerar uma série de abordagens diferentes que podem ser usadas em conjunto para satisfazer as demandas identificadas no início desta proposta. Tais abordagens são de fluxo de trabalho e esforços de colaboração, o aumento do uso de hipertexto e linguagens de marcação, e a aplicação de tecnologias *Web* voltados à resolução dos problemas de gerenciamento de documentos, que descrevemos a seguir. A presente proposta apresenta um ambiente de gestão de documentos que foi dividido em dois módulos: Especificação e Extração (Trabalhos Futuros). No módulo de Especificação, o usuário informa a estrutura do documento. O módulo é responsável por capturar informações físicas, semânticas e relativas ao processamento do documento. Essas informações serão usadas pelo módulo de Extração para o processamento de um lote de documentos da mesma classe do documento especificado.

O módulo de Extração possui duas fases principais: Geração do Lote de Documento e Reconhecimento. A primeira fase realiza as tarefas de aquisição, indexação e armazenagem das imagens dos documentos. A fase de Reconhecimento faz uso das informações do módulo de Especificação para pré-processar, extrair características e reconhecer os campos de cada documento do lote.

Adicionalmente, o ambiente oferece a possibilidade de correção de eventuais erros causados pelo classificador, várias formas de visualizar o resultado do reconhecimento, além de permitir a reconstrução do documento através de uma referência cruzada com as informações do módulo de Especificação.

1.1. Justificativa

Memórias organizacionais são motivadas pelo desejo de preservar e compartilhar o conhecimento e experiências que residem em uma organização. A memória organizacional é definida como uma explícita, representação, persistente do conhecimento e da informação em uma organização.

Vale a pena notar que as exigências descritas acima se aplicam aos documentos e informações mais estruturadas da mesma forma. Embora a relação entre potência e custo de *hardware* de computação tem multiplicado por um fator de um milhão nos últimos 30 anos, o custo de organizar as informações dentro desses sistemas tem aumentado [Goranson92].

Há uma necessidade evidente de melhorar a gestão da informação digital de todos os tipos. No entanto, de acordo com o Relatório Gilbane de Informação e de Sistemas de Documentos Abertos cerca de 80% da informação eletrônica corporativa está na forma de documentos, ao contrário dos estruturados registros de banco de dados. Esta estatística destaca a necessidade de foco, pelo menos, tanto na gestão de documentos sobre a gestão de informações mais estruturadas dentro de uma empresa, o que justifica a realização deste projeto de pesquisa.

Sistemas como o Excalibur, Autonomia e Lotus Notes são habilitados para *Internet*, mas não *internet-driven* [Bentley97]. É evidente que uma convergência de técnicas de gerenciamento de documentos e distribuição WWW seria vantajosa para facilitar uma visão de toda a empresa de documentos.

1.2. Objetivos

O principal objetivo desse trabalho é o desenvolver uma arquitetura para um ambiente computacional, que possa extrair informações dos mais diferentes tipos de documentos em papel, transformando-os em documentos eletrônicos estruturados.

Para se conseguir sucesso nessa empreitada, foi necessário a interação de diversos mecanismos de Processamento de Imagens e Sistema de Informação. Esses mecanismos realizarão tarefas que vão desde a captura da imagem e conversão da mesma para um formato digital, passando pelo reconhecimento de suas partes, até o armazenamento dos dados classificados.

O ambiente possuirá dois módulos principais: Especificação e Extração de Informações. O primeiro especifica um documento em papel, baseado em um modelo orientado a objetos. Partindo dessa especificação, o segundo módulo reconhecerá, de forma automática, o conteúdo de vários documentos do mesmo modelo, adquiridos através de algum dispositivo de entrada. Os conteúdos classificados serão armazenados em um banco de dados.

É possível desmembrar o objetivo principal desse trabalho, que é desenvolver uma metodologia para a especificação e extração de informações de imagens, em alguns objetivos secundários, entre eles, podem-se destacar os seguintes:

- Definir e implementar técnicas para aquisição, pré-processamento, extração de características e reconhecimento de imagens de documentos;
- Comparar as técnicas implementadas e, dessa forma verificar a adequação das mesmas no ambiente proposto;
- Confeccionar um banco de dados, com imagens de documentos, que será utilizado durante o estudo experimental;
- Desenvolver um sistema computacional, que valide a metodologia proposta.

O resultado prático, desse trabalho, servirá de protótipo para sistemas de análise de imagens de documentos.

2. A Gestão de Documentos

Uma solução efetiva adotada por grande número de empresas tem sido a gestão documental, pois considera uma série de abordagens diferentes que podem ser usadas em conjunto para satisfazer as demandas identificadas no início desta proposta. Tais abordagens são de fluxo de trabalho e esforços de colaboração, o aumento do uso de hipertexto e linguagens de marcação, e a aplicação de tecnologias *Web* voltados à resolução dos problemas de gerenciamento de documentos, que descrevemos a seguir.

2.1. Fluxo de Trabalho e Colaboração

Muitos sistemas de informação se concentram nos aspectos informativos do ambiente de trabalho em que eles são usados. Grande parte da literatura de gestão de documentos está preocupada com as maneiras em que o conteúdo informacional de documentos pode ser capturado e manipulado. No ambiente de trabalho muitos destes sistemas implicitamente apoiam os colaboradores em suas funções operacionais e de tomada de decisão. Abordagens como Computer Supported Co-operative Work (CSCW) fazem o papel explícito de que sistemas de informação realizam para seus usuários. [Swaby98] reconhece três requisitos gerais de motivação que a CSCW procura abordar:

- Articulação do trabalho cooperativo de Coordenação entre pessoas e recursos para contribuir para o desempenho de uma tarefa comum;
- Compartilhamento de um espaço de informação para assegurar que os membros do grupo possam compartilhar dados, informações, conceitos e heurísticas de uma forma estruturada;
- Adaptação da tecnologia para a organização criando um contexto organizacional no qual as atividades do grupo possam ser situadas, e permita interações apropriadas nesse cenário.

Este artigo não replica a revisão do CSCW que pode ser encontrada em [Swaby98], mas não reconhece e apoia a necessidade de uma abordagem de fluxo de trabalho para sistemas de informação. Produtos de *workflow*, como Lotus Notes pode auxiliar o fluxo de informações em torno de uma organização distribuída, mantendo um modelo de processos de trabalho dos usuários finais.

‘Colaboração’ o termo tradicionalmente sugere um nível de trabalho cooperativo com um objetivo comum. No contexto desta proposta o objetivo comum é a construção e manutenção de uma coleção de recursos voltados à especificação e extração de informações presentes em um documento que satisfaz um conjunto de metas semanticamente estabelecido.

A primeira classe de ferramentas, aquelas de curtas distâncias conhecidas entre os trabalhadores são exemplificadas por espaços de trabalho compartilhados e sistemas de conferência [Greenwood95]. Um requisito fundamental para essa colaboração é a integridade garantida de um recurso compartilhado em todos os momentos. Produtos como o Lotus Notes fornece esses mecanismos há alguns anos, mas em um formato proprietário e dentro de uma rede restrita. WebDAV [Whitehead98] é um desenvolvimento atual que amplia a *World Wide Web* (WWW) para fornecer uma infraestrutura padrão para criação colaborativa assíncrona através da *Internet*.

2.2. Hipertextos e Linguagens de Marcação

O campo de hipertexto teve início, em 1945, com o trabalho de Vannevar Bush que criou uma máquina que associava dois documentos, similar à maneira que a mente opera por associação [Bush45]. Como uma definição, um hipertexto é um conjunto de nós de informação com a máquina apoiando ligações entre esses nós e uma interface de usuário comum [Cunningham93]. Hipertexto expressa as relações entre os documentos. O primeiro protótipo de hipertexto eletrônico, NLS Aumentar, foi demonstrado em 1968. Muitos dos princípios fundamentais caracterizado por sistemas hipermídia de hoje já estavam presentes no protótipo inicial. O criador, Doug Engelbert, mesmo assim, viu o potencial para o computador de suporte aumentar o QI coletivo de comunidades ou a inteligência coletiva das organizações [Balasubramanian94].

A WWW tem popularizado a noção de hipertexto, com sua habilidade para saltar de uma parte de um documento para outra parte de outro documento ou para outra parte do mesmo documento. A sobrecarga cognitiva de seguir o *link* é menor do que o de realizar uma pesquisa através de um espaço de documento para encontrar os recursos adequados. [Golovchinsky97] sugere que as consultas de *link*-dirigido são mais eficazes do que as consultas específicas na recuperação de informações relevantes. A mecânica de hiperlinks é poderosa, mas a qualidade da ligação é dependente da pessoa que criou o *link*, ou seja pode ocorrer problemas adicionais com hiperlinks.

Em 1969 a IBM desenvolveu a linguagem de marcação generalizada (GML), destinada desde o início para separar forma e conteúdo na exibição de documentos [Berghe199]. Em 1986, o Padrão Generalizado de Marcação foi estabelecido como um padrão aceitável pela comunidade.

2.3. Tecnologias Web

A rede de computadores reconhecida pela primeira vez, a ARPANET, foi formada em dezembro de 1969. Ela consistia de quatro nós e tinha a intenção de descentralizar as informações militares e controle. Os usuários da ARPANET começaram a usar a rede para mais do que a intenção de computação de longa distância. Pesquisadores começaram a usar ARPANET para colaborar em projetos, ao comércio e para enviar mensagens de uma natureza mais social. Não houve dependência do tipo de computador usado para conectar a ARPANET, a única condição era o protocolo de rede usado para se comunicar. A ARPANET formou os primórdios da *Internet* como é hoje: um conjunto diversificado de computadores capazes de se comunicar com qualquer outra por meio de protocolos comuns e de uma rede de comunicação generalizada.

A computação cliente/servidor pode ser pensada como “processos de computador para a comunicação entre os computadores” em que um processo é o solicitante e os serviços de outro processo do pedido, normalmente através de uma rede, utilizando um protocolo comum a todos [Schwartz95]. ARPANET foi concebido desde o início para implementar uma arquitetura cliente-servidor, e a rede mundial moderna *Wide* é baseado no mesmo conceito.

A descentralização do poder computacional e informacional permitem a possibilidade de ultrapassar a organização de gestão de documentos: retenção local de ‘propriedade’ de documentos e acesso global a outros documentos dentro ou fora da organização. O sistema de gestão de documentos fornece controle central sobre os dados com acesso distribuído a esses dados através de soluções baseadas em padrões, fornecendo redes e protocolos abertos através da qual os servidores podem transportar dados a seus usuários distribuídos.

O sucesso sem precedentes da WWW levou as empresas a introduzir tecnologias *Web* para sua infraestrutura organizacional, denominando “*intranets*” as redes resultantes. O protocolo de rede padrão e a linguagem de marcação da WWW deve fornecer uma solução leve, voltada à resolução dos problemas de distribuição de muitas informações, com uma possibilidade de interrupção mínima em toda a organização.

Um navegador-cliente padrão é o requisito único para o *desktop*, com as complexidades de ser gerida de forma centralizada no servidor. No entanto, devido à falta de estrutura semântica dentro da linguagem de marcação atualmente utilizados em todo o WWW (HTML) e *intranets* têm tendência para Sistemas de Informação baseados na *Web* (WIS), em vez de baseado em Sistemas de Gestão de Documentos *Web*. Função que o servidor responsabiliza-se em codificar e separar as informações que fluem entre o sistema e o usuário, tratando assim do valor agregado. É evidente que as tecnologias da *Web* atual não podem apoiar plenamente em toda a empresa de gestão de documentos. [Balasubramanian98].

2.4. A Natureza da Documentação

Tal como acontece com muitos termos amplamente usados não existe uma definição axiomática de “documento”, mas a literatura aponta a pesquisa a alguns temas comuns. Ele também destaca que a atual e abrangente definição de “documento” nem sempre foram tão abertas. Definições mais tradicionais começam e terminam com um documento como “fornecer informações ou provas” um artefato de papel [Levy99].

Embora não seja construtivo para debater a questão em profundidade, um breve olhar sobre as atitudes diferentes aos documentos é útil.

O debate filosófico sobre a forma e a finalidade de documentos tem sido defendido ao longo dos anos, abrangendo muitas tecnologias. Em 1951 Suzanne Briet, um cientista da informação pioneiro, argumentou que o escopo de “documento” prolongado para além de texto serve para qualquer forma de material de provas. Embora útil como um marcador, a definição Briet é demasiadamente ampla para ser útil a este artigo (Briet foi tão longe para concluir que, sob certas circunstâncias, mesmo um antílope poderia ser um documento [Levy99]). Por outro lado, Sandy Ressler afirma que “*todo o conceito de um documento está se tornando cada vez mais obsoleto*” [Ressler95] como documentos e bases de dados convergem. Entre estes pontos de vista opostos são algumas outras definições importantes e relevantes.

2.5. Documentos como Objetos *Black Box*

Um documento pode ser descrito como uma unidade de “informação estruturada gravada para o consumo humano” citado em [Sprague95]. Esta definição implica uma distinção entre o método de armazenamento (informação registrada) e os meios pelos quais o conteúdo é entendido (consumo humano). Ele constrange o apoio da máquina potencial de pouco mais de um mecanismo de transporte para objetos *Black Box*. A noção de um documento como um objeto binário grande (BLOB): uma *Black Box* é armazenada e recuperada de acordo com alguma classificação externa.

Durante o desenvolvimento das redes de distribuição da indústria de computação houve uma preocupação em ajudar o usuário a criar, armazenar e recuperar documentos, mas não forneceu suporte explícito para a compreensão de conteúdo do documento fora do aplicativo de *software*. Ao considerar os documentos como objetos *Black Box* é confiar unicamente nos esquemas de classificação e métodos de gestão que a comunidade de usuários tem aplicado ao espaço do documento independentemente do conteúdo dos documentos armazenados.

Não são apenas as classificações subjetivas de documentos [Nurnberg97], impulsionado pelo contexto local e prática, mas quantidades crescentes de documentação exacerbam os problemas do uso de classificações aplicado apenas explicitamente como uma chave em uma loja de documento [Apte94]. A classificação única do documento só pode, em parte, descrever a finalidade ou função de um documento [McCarty97] e uma classificação mal aplicada pode processar um documento incoerente desde o início.

2.6. A Composição de Documentos

Ao considerar os documentos como objetos *Black-Box* ignora-se a intenção original e o propósito da criação do documento. Para compreender a relevância dos documentos

para uma organização é necessário para entender o que um documento contém. Um documento é um objeto instantâneo de um conjunto de informações que podem incorporar muitos tipos de informação, existem em vários lugares através de uma rede, dependem de outros documentos de informação, mudança de objetivos, e têm uma estrutura complexa [Sprague95]. Esta definição realça a natureza de composição de documentos.

Heeman discute dois componentes separados: a estrutura do documento e conteúdo do documento [Heeman92], com destaque para uma distinção entre o conteúdo físico de um documento e sua estrutura lógica. O documento descrito por suas marcações mais o conteúdo do documento separado de sua apresentação, criando três camadas amplas, mas distintas: estrutura, conteúdo e apresentação (ver Fig. 1).

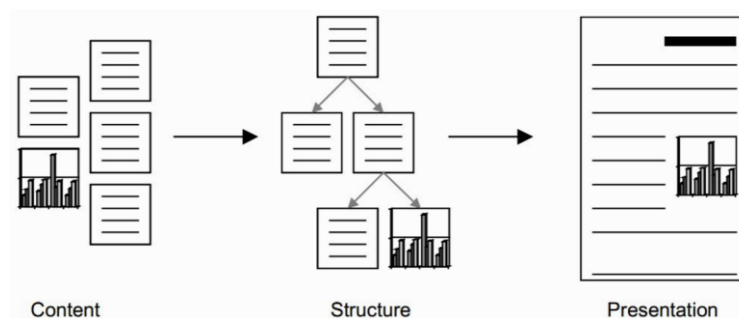


Figura 1: Estrutura básica de um documento (Heeman, 1992).

Na Fig. 1 descrevemos *Content*: os componentes (palavras, imagens, etc) que compõem um documento; *Structure*: a organização e os inter-relacionamentos entre os componentes; *Presentation*: como um documento parece e os processos que são aplicados a ele.

O conteúdo (tipo) foi reorganizado para formar uma estrutura (página) e depois apresentado em um meio adequado (papel ou meio digital). Métodos modernos facilitam a variância em todas as três camadas: a personalização da apresentação, validação da estrutura e da reutilização do conteúdo. Atualmente linguagens de marcação continuam a desenvolver as três camadas de separação e, sobretudo, aumentar a conscientização e aceitação das distinções entre as camadas. O reconhecimento da estrutura dentro de documentos é significativo em relação ao documento de discussão do banco de dados [Ressler95].

A maneira central no qual um tipo de informação difere da outra está em suas categorias semânticas: os objetos que são tidos como primitivos e as relações lógicas entre esses objetos. Convencionalmente o tipo de informação que precisa ser capturado determina a escolha do método e suporte de *software* subsequente [Fillion95]. Se ambos os dados tradicionais e os documentos podem ser descritos em termos de conteúdo e estrutura não deve haver distinção no método de seu tratamento. Dados e documentos estruturados são diferentes apenas na granularidade, não em sua natureza [Heeman92].

3. Metodologia

A Figura 2 mostra uma breve descrição do fluxo de informações no ambiente proposto. A partir dessa representação pictórica do sistema é possível identificar a ordem na qual as ações devem ser realizadas.

O primeiro passo é a criação de uma representação para o documento em questão – Especificação – Uma vez que o documento foi especificado será necessário reunir um conjunto de documentos, todos do mesmo tipo do documento especificado, e formar um lote – Geração do Lote de Documentos. Unindo as informações da especificação a cada documento do lote gerado, o sistema pode reconhecer os campos especificados – Reconhecimento.

Esses 3 (três) primeiros passos devem ser seguidos estritamente na ordem em que foram apresentados. Já que não é possível, nesse sistema, reconhecer um documento que não possua uma especificação. Entretanto, nada impede que a especificação de um documento possa ser mudada, ou que, sejam incluídos mais documentos em um determinado lote.

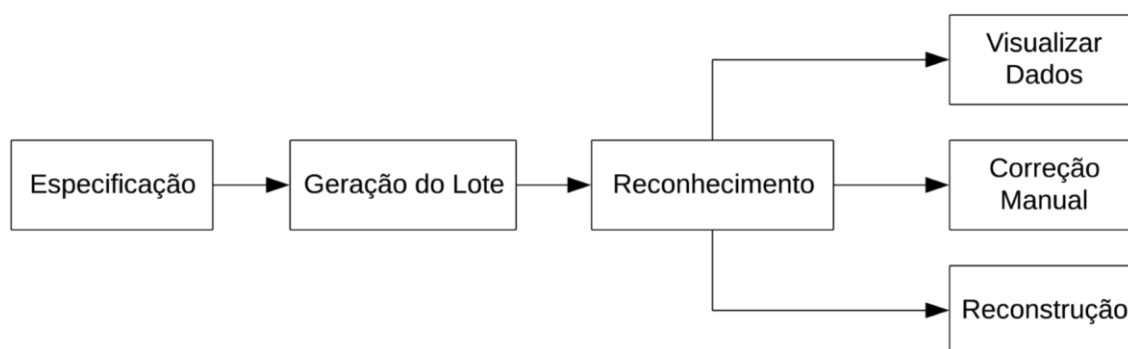


Figura 2: Fluxo de informações no ambiente.

Na Figura 2, nota-se, ainda, outros 3 (três) passos: Visualização dos Dados, Correção Manual e Reconstrução, que encontram-se desenhados com linhas pontilhadas. Após o Reconhecimento qualquer um desses passos pode ser acessado. No passo de Visualização dos Dados, como o próprio nome já diz, é possível verificar o resultado dos algoritmos de classificação. Já no passo de Correção Manual, o usuário pode corrigir eventuais erros gerados por uma má classificação. Além disso, é possível “recriar” o documento a partir dos dados classificados e da especificação do mesmo – Reconstrução.

4. Conclusões e Trabalhos Futuros

A arquitetura desenvolvida para um ambiente computacional conseguiu extrair informações dos mais diferentes tipos de documentos em papel, transformando-os em documentos eletrônicos estruturados, baseados em uma estrutura e metodologia pré-definidas, através uma especificação e extração de informações de imagens. Foram definidas e implementadas técnicas para aquisição, pré-processamento, extração de características e reconhecimento de imagens de documentos. Em seguida, identificamos comparamos as técnicas implementadas e, dessa forma foi verificada a adequação das mesmas no ambiente proposto.

Como Trabalhos Futuros, criamos alguns modelos relacionais de um banco de dados, para armazenamento das imagens de documentos, que serão utilizadas durante o estudo experimental. Resta ainda desenvolver o sistema computacional, que valide a metodologia proposta. Espera-se que resultado prático desse trabalho, sirva de protótipo para sistemas de análise de imagens de documentos.

Referências

- [Apte94] Chidanand Apte, Fred Damerau and Sholom Weiss. Automated Learning of Decision Rules for Text Categorisation. ACM Transactions on Information Systems, 12(3):233{251, July 1994.
- [Berghel99] Hal Berghel and Douglas Blank. The World Wide Web. Marvin Zelkowitz, editor, Advances in Computers Volume 49, pages 178{218. Academic Press, October 1999. [Balasubramanian94] V Balasubramanian and Helen Ashman. Long Distance Perspectives on Hypermedia: Keynote Address at the 1994 European Conference on Hypermedia Technology (ECHT94) by Douglas C. Engelbart and Christina Engelbart, Bootstrap Institute, September 1994.
- [Balasubramanian98] V Balasubramanian and Alf Bashian. Document Management and Web Technologies: Alice Marries the Mad Hatter. Communications of the ACM, 41(7):107{115, July 1998.
- [Bentley97] R Bentley, W Appelt, U Busbah, E Hinrihs, D Kerr, K Sikkell, J Trevor and G Woet-zel. Basi Support for Cooperative Work on the World Wide Web. International Journal of Human Computer Studies: Speial Issue on Novel Appliations of the World Wide Web, 46(6):827{846, June 1997.
- [Bush45] Vannevar Bush. As We May Think. The Atlanti Monthly, pages 101{108, July 1945.
- [Cunningham93] Donald J Cunningham, Thomas M Duy and Randy A Knuth. The Textbook of the Future. C M Knight, A Dillon and J Rihardson, editors, Hypertext A Psychologia Perspetive, pages 19{49. Ellis Horwood Series in Interative Information Systems, 1993.
- [Fillion95] Florene Fillion, Christopher Menzel, Thomas Blinn and Rihard Mayer. An Ontology-Based Environment for Enterprise Model Integration. Proceedings of the International JointConference on Arti ial Intel ligene (IJCAI), Montreal, Canada, August 1995.
- [Golovhinsky97] Gene Golovhinsky. What the Query Told the Link: The Integration of Hy-pertext and Information Retrieval. Proeedings of the 8th ACM International HypertextConferene (Hypertext97), Southampton, UK, pages 67{74, April 1997.
- [Goranson92] H T Goranson. Some Initial Results from Enterprise Integration Studies. Reports of the International Conference on Enterprise Integration Modeling Technology (ICEIMT) Special Interest Groups, Department of Defense, USA, August 1992.
- [Greenwood95] R M Greenwood, I Robertson, R A Snowdon and B C Warboys. Ative Models in Business. Proceedings of the 5th Conferene on Business Information Tehnology (CBIT95),Manhester, UK, November 1995.
- [Heeman92] Frans C Heeman. Granularity in Struted Douments. Eletronic Publishing, 5(3):143{155, September 1992.

- [Levy99] David M Levy. The Universe is Expanding: Reactions on the Social (and Cosmological) Significance of Documents in a Digital Age. Bulletin of the American Society for Information Science, 25(4):17-20, 1999.
- [MCarty97] Willard MCarty. The Shape of Things to Come is Continuous Change: Fundamental Problems in Electronic Publishing. Presented at Electronic Publishing A Day Conference, University of London, UK, January 1997.
- [Nurnberg97] Peter J Nurnberg, John J Leggett and Erich R Schneider. As We Should Have Thought. Proceedings of the 8th ACM Conference on Hypertext (Hypertext97), Southampton, UK, pages 96-101, April 1997
- [Ressler95] Sandy Ressler. Network Based Documents. OII Spectrum Management Handbook. January 1995. Available on-line at <http://www.itl.nist.gov/iaui/ovrt/people/sressler/netDocs/netDocs.fm.html> (Last verified 8th March 2000).
- [Shwartz95] Gary Shwartz. Everything You Ever Wanted to Know About Client/Servers. Information Technology Article, April 1995. Available on-line at <http://magi.com/mmlick/it95april.htm> (Last verified 8th March 2000).
- [Sprague95] Ralph Sprague. Electronic Document Management: Challenges and Opportunities for Information Systems Managers. Management Information Systems Quarterly, 19(1):29-49, March 1995.
- [Swaby98] Mike Swaby. A Model-Based Approach to Construction of Integrated Internet CSCW Systems. PhD thesis, School of Computer Studies, University of Leeds, UK, 1998.
- [Whitehead98] E James Whitehead Jr and Meredith Wiggins. WEBDAV: IETF Standard for Collaborative Authoring on the Web. IEEE Internet Computing, 2(5), September/October 1998.