LUIZ ANTONIO DO NASCIMENTO

**PROPOSTA DE UM SISTEMA DE RECUPERAÇÃO DE
INFORMAÇÃO PARA EXTRANET DE PROJETO**

Dissertação apresentada à Escola
Politécnica da Universidade de São
Paulo para obtenção do Título de
Mestre em Engenharia.

São Paulo
2004

LUIZ ANTONIO DO NASCIMENTO

**PROPOSTA DE UM SISTEMA DE RECUPERAÇÃO DE
INFORMAÇÃO PARA EXTRANET DE PROJETO**

Dissertação apresentada à Escola
Politécnica da Universidade de São
Paulo para obtenção do Título de
Mestre em Engenharia.

Área de Concentração:
Engenharia de Construção Civil e
Urbana

Orientador:
Prof. Dr. Eduardo Toledo Santos

São Paulo
2004

FICHA CATALOGRÁFICA

Nascimento, Luiz Antonio do
Proposta de um Sistema de Recuperação de
Informação para Extranet de Projeto. São Paulo,
2004.
123p.

Dissertação (Mestrado) – Escola Politécnica da
Universidade de São Paulo. Departamento de
Engenharia de Construção Civil.

1.Armazenamento e Recuperação de Informação
2.Sistema de Informações para Construção Civil
3.Tecnologia da Informação. I.Universidade de São
Paulo. Escola Politécnica. Departamento de
Engenharia de Construção Civil II.t

À minha avó Maria, *in memoriam*,
aos meus pais José Oscar e Lucia
pela minha educação, à minha
esposa Luiza pelo apoio e
compreensão e aos meus filhos
Luan e Luana pelo tempo que não
pude dispor a eles.

AGRADECIMENTOS

Ao meu orientador Prof. Dr. Eduardo Toledo Santos pelo apoio e incentivo ao longo do desenvolvimento deste trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos que possibilitou a minha dedicação ao desenvolvimento desta pesquisa.

Ao Prof. Dr. Sérgio Leusin da UFF pelas informações a respeito do projeto CDCON utilizado neste trabalho, ao Eng. Marcelo Picarelli diretor da empresa Sistrut que forneceu a extranet de projeto e ao Eng. Marco Antonio Manso da empresa Fortenge pelo apoio e interesse em contribuir com o trabalho fornecendo uma base de documentos para avaliação do sistema apresentado neste trabalho.

Aos colegas Rovilson Mafalda, Kemil Kangussu, Walter Dinis e Marco Yamamoto pela amizade, disponibilidade e apoio.

Aos Professores João Petreche e Silvio Melhado pelas contribuições e sugestões para aperfeiçoamento deste trabalho.

Aos professores Lucio Soibelman da Universidade de Illinois e Carlos Caldas da Universidade do Texas pelas indicações de literatura sobre recuperação de informação.

Aos colegas do Departamento de Engenharia de Construção Civil – PCC, especialmente aos membros do grupo de pesquisa SUP do laboratório LabCAD pela amizade e companheirismo.

A minha família pelo apoio, incentivo e compreensão.

A todos que, direta ou indiretamente, colaboraram na execução deste trabalho.

Agradeço a Deus, nosso criador, pois sem ele nada disso teria acontecido.

RESUMO

A indústria da construção civil tem adotado várias tecnologias nos últimos anos com o propósito de melhorar seu desempenho, entre elas a Tecnologia da Informação (TI). Alguns dos principais objetivos da TI no setor são o de integrar e agregar valor aos processos e ajudar no gerenciamento de empreendimentos. Neste sentido, com a consolidação da Internet, foram desenvolvidas as chamadas extranets de projeto para gestão dos processos e armazenamento de documentos. Com estes sistemas é possível integrar as várias equipes intervenientes nos processos da construção civil, oferecendo um grande volume de informações a cada usuário. A proposta deste trabalho é conceber uma ferramenta especializada que auxilie na tarefa de recuperar informações em documentos de construção civil armazenados em extranets de projeto. As ferramentas atualmente disponíveis nas extranets normalmente possuem apenas mecanismos convencionais, com recursos baseados em busca por palavras-chave. Estes mecanismos podem retornar grande quantidade de documentos irrelevantes frente a uma solicitação. Isso ocorre pelo generalismo e pela descontextualização dessas ferramentas. Numa extranet de projeto, o tema dos documentos armazenados é delimitado (a Construção Civil). A ferramenta proposta procura otimizar os argumentos das consultas e ordenar por relevância a lista de documentos recuperados. Os algoritmos de busca e ordenação de documentos apresentados neste sistema são baseados no conhecimento do domínio da construção civil objetivando maior eficácia na recuperação. Os resultados obtidos mostraram grande melhora nos parâmetros de precisão e revocação do sistema quando comparados a mecanismos convencionais.

ABSTRACT

The Building Industry has been adopting several technologies in the last years for improving its performance, including those from the Information Technology (IT). Some of the main objectives of IT in this sector are to integrate and to add value to processes, as well as to help in enterprise management. In this sense, along with the consolidation of the Internet, the so-called project extranets were developed for process management and document archival. With these systems it is possible to integrate all the different teams involved with the building construction processes, offering each user a large amount of information. The purpose of this work is to devise an expert tool to help in the information retrieval from building documents stored in project extranets. The searching engines available in current extranets are conventional, featuring only keyword search. These mechanisms can retrieve many irrelevant documents upon a request. This is due to the generic nature and the non-contextualization of these tools. The subject matter of the documents in a project extranet is restricted (Building Construction). The proposed tool aims to optimize the user query terms and to rank retrieved documents according to their relevance. The document search and ranking algorithms of this system are based on building construction domain knowledge for improving retrieval efficacy. The obtained results show large improvements on the system's precision and recall parameters when compared to a standard search mechanism.

ÍNDICE

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

RESUMO

ABSTRACT

1	INTRODUÇÃO	6
1.1	JUSTIFICATIVA.....	8
1.2	OBJETIVOS.....	10
1.3	ORGANIZAÇÃO DO TRABALHO	11
2	TECNOLOGIA DA INFORMAÇÃO NA CONSTRUÇÃO CIVIL	12
2.1	VISÃO GERAL	12
2.2	BARREIRAS PARA A TI NA INDÚSTRIA DA CONSTRUÇÃO CIVIL	14
2.3	EXTRANETS DE PROJETO	16
2.3.1	DEFINIÇÕES E CARACTERÍSTICAS	16
2.4	USUÁRIOS DO SISTEMA	19
2.5	BENEFÍCIOS DA TECNOLOGIA.....	21
2.6	PROBLEMAS DA TECNOLOGIA.....	23
2.7	TIPOS DE DOCUMENTOS ARMAZENADOS NAS EXTRANETS DE PROJETO	24
2.8	ACESSO ÀS INFORMAÇÕES	25
3	RECUPERAÇÃO DE INFORMAÇÃO.....	31
3.1	DEFINIÇÕES E CARACTERÍSTICAS	31
3.2	SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÃO.....	34
3.3	AVALIAÇÃO DE UM SISTEMA DE RECUPERAÇÃO DE INFORMAÇÃO	37
4	REPRESENTAÇÃO DO CONHECIMENTO	42
4.1	REPRESENTAÇÃO POR REGRAS.....	42
4.2	REDES SEMÂNTICAS	42
4.3	REPRESENTAÇÃO BASEADA EM FRAMES	43
5	ESTRUTURAS DE CLASSIFICAÇÃO PARA A CONSTRUÇÃO	46
5.1	PADRÕES ISO.....	46
5.1.1	ISO 14177	47
5.1.2	ISO 10303 - STEP.....	48
5.2	SISTEMAS DE CLASSIFICAÇÃO NO MUNDO	49
5.2.1	MASTER FORMAT	49
5.2.2	OCCS	50
5.3	SISTEMAS DE CLASSIFICAÇÃO NO BRASIL	51
5.3.1	CDCON	51
5.3.2	DIRETRIZES GERAIS PARA INTERCAMBIALIDADE DE PROJETOS EM CADD	54
5.4	OUTROS TIPOS DE CLASSIFICAÇÕES.....	55
5.4.1	IFC	56
5.4.2	AECXML	57
5.4.3	CONSIDERAÇÕES SOBRE SISTEMAS DE CLASSIFICAÇÕES PARA A CONSTRUÇÃO CIVIL	59

6	PROPOSTA DE SISTEMA DE RECUPERAÇÃO DE INFORMAÇÃO	61
6.1	VISÃO GERAL DO SISTEMA RICON.....	61
6.2	OPERAÇÃO DA FERRAMENTA	62
6.3	PRÉ-PROCESSAMENTO	63
6.4	EXPANSÃO DE TERMOS	64
6.4.1	EXPANSÃO COM RELACIONAMENTO DE EQUIVALÊNCIA	64
6.4.2	EXPANSÃO COM RELACIONAMENTO HIERÁRQUICO	65
6.5	BUSCA CONTEXTUAL	66
6.6	EXEMPLO DE UMA BUSCA NO SISTEMA RICON	68
7	AVALIAÇÃO DO SISTEMA	74
7.1	ESCOLHA DA EXTRANET DE PROJETO	75
7.1.1	CARACTERÍSTICAS DA EXTRANET DE PROJETO DO EXPERIMENTO.....	75
7.2	COLEÇÃO DE DOCUMENTOS	77
7.3	MÉTODO DE AVALIAÇÃO DO SISTEMA.....	78
7.4	RESULTADOS OBTIDOS	80
8	CONCLUSÕES E TRABALHOS FUTUROS	84
8.1	CONCLUSÕES	84
8.2	TRABALHOS FUTUROS	87
	ANEXO A - DETALHAMENTO DE UMA BUSCA.....	89
	ANEXO B – CONSULTAS EXECUTADAS NO EXPERIMENTO	98
	REFERÊNCIAS	116

LISTA DE FIGURAS

Figura 1 – Produtividade da Construção Civil (EUA).....	13
Figura 2: Página de entrada de uma extranet de projeto	16
Figura 3: Comunicação típica de um empreendimento sem extranet de projeto	19
Figura 4: Comunicação em um empreendimento com extranet de projeto	20
Figura 5: Página de uma extranet de projeto com foto de uma obra.....	22
Figura 6 – Formulário transferência e descrição de arquivo de uma extranet de projeto.....	27
Figura 7 – Formulário de busca de uma extranet de projeto utilizada no Brasil..	28
Figura 8 – Formulário de busca de uma grande extranet de projeto americana.	28
Figura 9 – Resultado de uma busca em uma extranet de projeto	29
Figura 10 – Sistema de Recuperação de Informações.....	35
Figura 11 – Avaliação de sistemas de recuperação de informações	38
Figura 12: Exemplo de Representação por <i>Frames</i>	44
Figura 13: Expansão de Termos por Relacionamento de Equivalência	65
Figura 14: Expansão de Termos por Relacionamento Hierárquico.....	66
Figura 15: Estrutura hierárquica do frame bloco de alvenaria	68
Figura 16: Simulação do resultado de uma busca típica do sistema RIcon.....	73
Figura 17: Fluxograma de processamento de uma busca no sistema RIcon	73
Figura 18: Opções de configuração da extranet do experimento.....	76
Figura 19: Opções de configuração da extranet do experimento.....	77
Figura 20: Precisão média para os 10 primeiros documentos recuperados.....	81
Figura 21: Revocação média para os 10 primeiros documentos recuperados.....	82
Figura 22: Curva Precisão x Revocação da média dos 10 primeiros documentos recuperados.....	82
Figura 23: Formulário de consulta do sistema RIcon	89
Figura 24: Representação do frame fundação na base de conhecimento através do Protégé-2000	94
Figura 25: Precisão média para os 10 primeiros documentos	94
Figura 26: Revocação média para os 10 primeiros documentos	95
Figura 27: Curva Precisão x Revocação para todos os documentos recuperados	95
Figura 28: Precisão média para os 10 primeiros documentos – Pesquisa 1	101
Figura 29: Revocação média para os 10 primeiros documentos – Pesquisa 1....	101
Figura 30: Curva Precisão x Revocação para todos os documentos recuperados – Pesquisa 1	102
Figura 31: Precisão média para os 10 primeiros documentos – Pesquisa 2	105
Figura 32: Revocação média para os 10 primeiros documentos – Pesquisa 2....	105
Figura 33: Curva Precisão x Revocação para todos os documentos recuperados – Pesquisa 2	106
Figura 34: Precisão média para os 10 primeiros documentos – Pesquisa 3	107
Figura 35: Revocação média para os 10 primeiros documentos – Pesquisa 3....	108
Figura 36: Curva Precisão x Revocação para todos os documentos recuperados – Pesquisa 3	108
Figura 37: Precisão média para os 10 primeiros documentos – Pesquisa 4	110
Figura 38: Revocação média para os 10 primeiros documentos – Pesquisa 4....	110
Figura 39: Curva Precisão x Revocação para todos os documentos recuperados – Pesquisa 4	111
Figura 40: Precisão média para os 10 primeiros documentos – Pesquisa 5	114
Figura 41: Revocação média para os 10 primeiros documentos – Pesquisa 5....	114
Figura 42: Curva Precisão x Revocação para todos os documentos recuperados – Pesquisa 5	115

LISTA DE TABELAS

Tabela 1. Tabela de Contingência	39
Tabela 2: Classes do Technical Report 14177 da ISO.	47
Tabela 3 - Grupos de documentos do STEP	48
Tabela 4 – Divisões do MasterFormat	50
Tabela 5 – Tabelas do OCCS	50
Tabela 6 – Facetas do CDCON	52
Tabela 7 – Abreviação para nomenclatura de nomes de arquivos (AsBEA)	55
Tabela 8 – Classificação do padrão IFC (IAI, 2000)	57
Tabela 9 – Classificação do aecXML (versão 1.0)	58
Tabela 10 – Tabela de documentos x termos com pesos	71
Tabela 11. Listagem típica dos resultados do sistema do experimento.	78
Tabela 12 – Ganhos de precisão e revocação do sistema proposto em relação ao existente.	83
Tabela 13 – Ordenação, documentos relevantes, precisão e revocação para consulta “Maracaju”	93
Tabela 14 – Ordenação, documentos relevantes, precisão e revocação para consulta “Fundação”	93
Tabela 15 – Ordenação, documentos relevantes, precisão e revocação no sistema RIcon	96
Tabela 16 – Tabela com pesos e notas dos 10 primeiros documentos de uma consulta	97
Tabela 17 – Consultas válidas	98
Tabela 18 – Consultas rejeitadas	98
Tabela 19 – Ordenação, precisão e revocação da Pesquisa 1 - SADP	100
Tabela 20 – Ordenação, precisão e revocação da Pesquisa 1 RIcon	100
Tabela 21 – Ordenação, precisão e revocação da Pesquisa 2 – 2º opção SADP	103
Tabela 22 – Ordenação, precisão e revocação da Pesquisa 2 – 1º opção SADP	103
Tabela 23 – Ordenação, precisão e revocação da Pesquisa 2 – RIcon	104
Tabela 24 – Ordenação, precisão e revocação da Pesquisa 3 – SADP	107
Tabela 25 – Ordenação, precisão e revocação da Pesquisa 3 – RIcon	107
Tabela 26 – Ordenação, precisão e revocação da Pesquisa 4 – SADP	109
Tabela 27 – Ordenação, precisão e revocação da Pesquisa 4 – RIcon	109
Tabela 28 – Ordenação, precisão e revocação da Pesquisa 5 – SADP	112
Tabela 29 – Ordenação, precisão e revocação da Pesquisa 5 – RIcon	113

LISTA DE ABREVIATURAS E SIGLAS

AEC	Arquitetura, Engenharia e Construção
ASP	<i>Application Service Provider</i>
CAD	Projeto Auxiliado por Computador (<i>Computer Aided Design</i>)
CADD	Projeto e Desenho Auxiliados por Computador (<i>Computer Aided Design and Drafting</i>)
CAE	Engenharia Auxiliada por Computador (<i>Computer Aided Engineering</i>)
CAM	Fabricação Auxiliada por Computador (<i>Computer Aided Manufacturing</i>)
GC	Gestão do Conhecimento
CNC	Controle Numérico Computadorizado
HTML	<i>HyperText Markup Language</i>
IFC	<i>Industry Foundation Classes.</i>
IR	Recuperação de Informação (<i>Information Retrieval</i>). Parte da Ciência da Computação que estuda a recuperação de informações relevantes de uma coleção de documentos através de uma consulta de um usuário em uma base de dados.
ISO	<i>International Organization for Standardization</i>
SGML	<i>Standard Generalized Markup Language</i>
SI	Sistema de Informação.
TI	Tecnologia da Informação
XML	<i>eXtensible Markup Language</i> . Uma metalinguagem, subconjunto da SGML, que permite que uma marcação específica seja criada para especificar semântica e definir o conteúdo de um documento compartilhado em uma rede.

1 INTRODUÇÃO

A indústria da construção civil é um dos setores mais tradicionais da indústria brasileira e tem um papel sócio-econômico importante no Brasil, com uma participação de todo o macrossetor em torno de 19% do PIB brasileiro (CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO, 2002). No mundo, representa um mercado de U\$3.5 trilhões (DRI-WEFA; DODGE, 2000 apud DRI-WEFA; DODGE, 2002). Apesar disso, é um dos setores tecnologicamente menos desenvolvidos da indústria e tem se caracterizado por baixos níveis de produtividade (SCHWEGLER et al., 2001) e de utilização de Tecnologia da Informação (STEWART; MOHAMED; DAET, 2002). Outras indústrias, como a de transformação, por exemplo, que utilizam intensivamente a Tecnologia da Informação (TI) no processo de projeto (CAD, CAE, simulação, etc.) e na automação da produção (CNC, CAM, robotização, etc.), estão mais desenvolvidas que a indústria da construção civil. Segundo Newton (1998), as oportunidades para aumento da eficiência desta indústria pela aplicação de TI são muitas, proporcionando impacto significativo no crescimento nacional, provavelmente maior que na maioria dos demais setores. Este aumento de eficiência é necessário pois a globalização e o novo panorama mundial, bem como o atual cenário nacional desestatizado e com escassez de financiamento, requerem da construção civil urgente melhora de produtividade e competitividade. Porém, não se consegue implementar com sucesso algumas experiências adotadas em outros setores. Um dos motivos é que a indústria da construção civil tem algumas características diferenciadas em relação a outras indústrias:

- as diversas empresas e profissionais que atuam na indústria da construção civil estão normalmente dispostos em locais

diferentes (canteiro de obra, escritório de arquitetura, engenharia, etc.) e se reúnem em forma de um consórcio virtual e temporário para realizarem um empreendimento (MICALI, 2000);

- ocorrência de graves problemas de qualidade de produtos intermediários e finais da cadeia produtiva e elevados custos de correções e manutenção pós-entrega (CARDOSO, et al. 2002);
- desestímulo ao uso mais intensivo de componentes industrializados devido à alta incidência de impostos e conseqüente encarecimento dos mesmos (CARDOSO, et al. 2002);
- a fragmentação do processo produtivo dificulta a adoção de medidas de padronização (MICALI, 2000).

Atualmente, com o grande crescimento no uso de documentos no formato digital, diversos sistemas computadorizados para armazenamento de documentos e comunicação são utilizados na indústria da construção civil. Estes sistemas auxiliam a comunicação, a coordenação e as tomadas de decisão dos agentes, a partir de diversos instrumentos e tecnologias que suprem o indivíduo de informações para que possa avaliar melhor os riscos e as alternativas. Muitas dessas tecnologias atualmente estão ligadas à Internet. Uma das mais promissoras tecnologias ligadas à Internet para a indústria da construção civil são os *web sites* para gerenciamento dos processos de projeto, chamados de extranets de projetos. Estes sistemas são baseados em tecnologias que interligam empresas através da Internet. Com estas tecnologias é possível trocar informações estratégicas e a substituição de práticas como as de tirar fotocópias, envio de fax, reuniões presenciais e uso de correio. As extranets armazenam toda a

documentação em formato eletrônico que é produzida num empreendimento e as disponibiliza aos seus usuários.

1.1 Justificativa

As extranets de projeto gerenciam muitas informações já que em um projeto típico podem ser produzidos milhares de documentos (NEWTON, 1998). A localização de documentos nestes repositórios pode, em muitos casos, se tornar problemática, caso a organização ou o sistema de busca sejam deficientes. Este trabalho visa tornar as extranets de projeto mais eficientes e eficazes através da concepção de um mecanismo para recuperação de documentos de sua base de dados. Este mecanismo, chamado de RIcon (Recuperação de Informações da Construção), pretende melhorar o processo de localização de documentos relevantes na base de dados. A melhora na recuperação de documentos nas extranets contribuirá para que seus usuários tomem decisões mais rápidas através da localização de documentos relevantes armazenados no sistema, evitando perda de tempo e estresse com a procura de informações.

Atualmente, a busca de informações em extranets de projetos ainda não é o principal problema para os usuários destes sistemas porque esta tecnologia é razoavelmente nova e seus usuários ainda não utilizam totalmente seus recursos. Em muitos casos, a figura do coordenador de projetos é responsável por organizar e disponibilizar os arquivos de projeto de forma de cada agente só tenha contato direto com os documentos que precisa para realizar as tarefas que lhe cabem em certo momento.

Vislumbra-se no futuro, com a maturação da tecnologia e, principalmente, de seus usuários, que as extranets passem a ser

utilizadas como uma efetiva ferramenta de colaboração (Engenharia Simultânea), onde a atuação do coordenador de projeto não seja estritamente necessária. Neste cenário, as informações de projeto devem estar disponíveis a todos os intervenientes dos processos da construção civil para que possam colaborar e tomar decisões simultaneamente. A disponibilidade de um instrumento eficaz de busca de documentos relevantes, como o proposto neste trabalho, se torna indispensável neste contexto.

Atualmente no contexto da “Era da Informação” o conhecimento é um diferencial estratégico para as empresas e por isso são necessários métodos e ferramentas para tratá-los. Este assunto está despertando interesse nas empresas dos mais variados setores, inclusive nas empresas do setor da construção (SILVA, HEINECK, 2001). Para tratar deste assunto equipes de Gestão do Conhecimento (GC) procuram identificar quais conhecimentos são estratégicos para a organização. Uma das maneiras de conseguir estes objetivos é através de fontes de informações apropriadas para a GC. Uma extranet é uma boa fonte de informações para estas equipes, pois elas armazenam todo o histórico de um empreendimento. Este trabalho propõe uma ferramenta especializada que auxilie equipes de GC na recuperação de informações neste repositório de documentos de projeto. Segundo Davenport e Prusak, (1999) a mais valiosa função da tecnologia na gestão do conhecimento é estender o alcance e aumentar a velocidade da transferência de conhecimento.

1.2 Objetivos

Os objetivos gerais deste estudo são:

- propor uma ferramenta informatizada para rápido acesso a informações relevantes aos intervenientes do processo de projeto da indústria da construção civil;
- reduzir a sobrecarga de informações gerada pelos sistemas de busca em extranets de projetos através de algoritmos de recuperação de informações que levem em conta o conhecimento do domínio da indústria da construção civil;
- melhorar a busca de informações relevantes e oportunas em extranets de projeto, tornando-as mais eficazes;
- propiciar que os agentes participantes nos processos da indústria da construção civil obtenham o conhecimento de experiências passadas através da rápida extração de informações em repositório de documentos do histórico de empreendimentos.

Os objetivos específicos deste estudo são:

- propor um novo sistema para recuperação de informações oportunas e relevantes em extranets de projetos com desempenho superior àqueles normalmente disponíveis nestes sistemas;
- comparar o sistema proposto com um em uso pelo mercado;
- proporcionar maior seletividade em relação à relevância dos resultados nos mecanismos de busca das extranets de projeto (melhor precisão);
- melhorar a precisão do resultado das buscas em extranets de projeto através da contextualização semântica dos termos utilizados pelos usuários nos mecanismos de busca.

1.3 Organização do trabalho

O texto desta dissertação foi estruturado em 8 capítulos, da forma a seguir:

- Capítulo 1 – Contém a introdução e a justificativa para o desenvolvimento deste trabalho, seus objetivos e sua forma de organização;
- Capítulo 2 – Apresenta uma visão geral do uso atual da Tecnologia da Informação e de extranets de projeto na indústria da construção civil, seus problemas, benefícios e características;
- Capítulo 3 – Introduz conceitos e definições da área de Recuperação de Informação e métodos de avaliação de sistemas de recuperação de informação;
- Capítulo 4 – Mostra alguns conceitos e métodos de representação do conhecimento;
- Capítulo 5 – Apresenta várias estruturas de classificação e padronização utilizadas na indústria da construção civil no Brasil e no mundo, usadas como fonte do conhecimento utilizado pelo sistema proposto;
- Capítulo 6 – Propõe um sistema de recuperação de informação para extranets de projeto baseado no conhecimento da indústria da construção civil;
- Capítulo 7 – Apresenta a metodologia e os resultados obtidos em uma avaliação do sistema de recuperação proposto em relação ao existente em uma extranet de projeto utilizada para avaliação com dados de um empreendimento real;
- Capítulo 8 – Contém as conclusões e propostas futuras.

2 TECNOLOGIA DA INFORMAÇÃO NA CONSTRUÇÃO CIVIL

2.1 Visão geral

Tecnologia da Informação (TI) é a tecnologia que envolve a introdução, armazenamento, processamento e distribuição da informação por meios eletrônicos. A TI compreende eletrônica, automação, computação (*hardware* e *software*) e telecomunicações. Sua finalidade na indústria da construção civil é principalmente facilitar a troca e a gerência da informação ao longo dos processos do setor (RIVARD, 2000).

As mudanças ocorridas nos últimos anos no cenário mundial com a globalização e os avanços tecnológicos na informática e nas telecomunicações têm levado a tradicional e conservadora indústria da construção civil a adotar inovações. Algumas podem ser consideradas como estratégias competitivas para as organizações do setor. Porém, já que os riscos e incertezas inerentes às inovações tecnológicas não são aceitáveis para boa parte do setor da construção (TOLEDO et al., 2000), apenas depois de consolidada é que uma tecnologia passa a ser adotada por um número razoável de empresas. Segundo Andresen et al. (2000), a indústria da construção civil tem investido pouco em TI em relação a outros setores da indústria. Este autor afirma também que a falta de casos demonstrados do uso efetivo de TI no setor cria dificuldades para os gerentes justificarem investimentos futuros, apostando nos benefícios desta tecnologia. Além disso, o problema é agravado pela estrutura fragmentada da Construção Civil e pelo limitado capital reservado para investimento em TI. Outro fator importante é que as inovações no setor se dão ao longo de anos (TOLEDO et al., 2000).

O aumento da produtividade é um dos principais objetivos na adoção de TI. É muito difícil medir o aumento de produtividade com o uso destas tecnologias no Brasil. Na indústria da construção brasileira não há índices históricos de produtividade, pois somente na década de 90 surgem estudos importantes para definição de tais índices (LANTELME, 1994). Porém, nos Estados Unidos, a produtividade da indústria da construção é mensurada e este índice mostra que lá este setor é menos produtivo que as outras indústrias (figura 1). Considerando que as tecnologias e os processos construtivos adotados naquele país favorecem a produtividade, é de se esperar que a situação brasileira seja igual ou pior àquela e que aqui o setor também continue mais ineficiente que outros da indústria como um todo.

Figura 1 – Produtividade da Construção Civil (EUA) Adaptado de (TEICHOLZ, 2000)

Nestes últimos anos, a indústria da construção civil brasileira sofreu várias influências políticas e econômicas como as privatizações das empresas estatais, os efeitos da globalização, a diminuição da intervenção do Estado no mercado, a variação da moeda, o aumento da taxa de juros, a exigência do governo aos programas de qualidade, o aumento da competitividade,

diminuição dos riscos nos investimentos e a diminuição das margens de lucro das companhias. Muitas destas mudanças impediram um maior crescimento das empresas e exigiram que adotassem medidas para se adequarem ao mercado. Algumas companhias utilizaram a TI para aumentar seus níveis de produtividade. Acredita-se que, de maneira geral, a utilização da TI no setor da Construção Civil é muito semelhante em todo o mundo. Porém, pela instabilidade política e recessão econômica que inibe os investidores no Brasil, não há o uso tão intenso como em alguns países desenvolvidos. As tecnologias mais usadas no setor são aquelas mais específicas como CAD (*Computer Aided Design*) e sistemas para cálculo de estruturas ou ferramentas para automação de escritórios (programas para edição de textos, elaboração de planilhas eletrônicas, criação de banco de dados e gerenciamento de correio eletrônico). As tecnologias mais genéricas como EDI (*Electronic Data Interchange*), ERP (*Enterprise Resource Planning*), PDM (*Product Data Management*), EDMS (*Electronic Document Management System*), WF (*Workflow*) e aplicações de *e-commerce* ainda são usadas em pequena escala.

2.2 Barreiras para a TI na Indústria da Construção Civil

A indústria da construção civil enfrenta alguns obstáculos quanto à adoção de TI. Por causa dos diferenciais do setor em relação a outras indústrias, não se pode adotar as mesmas soluções. Porém, paradoxalmente, a pequena penetração da TI em relação a outros setores se dá, em grande parte, devido a um conjunto de barreiras ligadas aos profissionais que atuam na área, aos seus processos longamente estabelecidos, às características do próprio setor e às deficiências da tecnologia (NASCIMENTO e SANTOS, 2002b):

- O pessoal dos níveis superiores das companhias normalmente não possui desenvoltura com a aplicação de TI e não está preparado para especificar nem avaliar ferramentas;
- Ainda há falta de padronização na comunicação, apesar dos esforços de organismos como a IAI (IFC e aecXML), ANTAC (CDCON) e AsBEA (padronização de desenhos CAD). O desenvolvimento de padrões e classificações para promover a interoperabilidade e integração para a indústria da construção tem demonstrado ser uma atividade complexa (ROE; REINA, 2001);
- Na maioria das empresas não existe a área de TI e, mesmo quando existe, não tem orçamento significativo que permita investimento adequado para alavancar resultados positivos (HASSEL et al., 2000);
- O impacto das poucas empresas que investem em TI é pequeno, pois a indústria da construção civil é muito grande, diversificada e fragmentada (ZEGARRA et al., 1999). Em consequência disto, menores serão os benefícios destes investimentos quanto menos agentes usarem a TI (HASSEL et al., 2000);
- Problemas com custos de aquisição e manutenção de equipamentos e software: apesar do barateamento do custo, este fator foi identificado como a barreira mais significativa em pesquisa do professor Rivard (2000), realizada no Canadá. Provavelmente é significativa também no Brasil;
- As estruturas curriculares das universidades não dão uma maior ênfase às aplicações de TI para os futuros profissionais, de forma que tenham uma visão mais ampla e prática das Tecnologias de Informação existentes para sua área;

- Até há pouco tempo atrás, praticamente não existia na formação dos profissionais de AEC a utilização de ferramentas informatizadas. Assim, mesmo os jovens engenheiros e arquitetos que se formam hoje, terão por muitos anos um impacto ainda limitado no setor (HASSELL et al., 2000);
- Muitos profissionais da área de nível estratégico das companhias possuem algum conhecimento de TI, porém de nível operacional, não compatível com suas funções.

2.3 Extranets de Projeto

2.3.1 Definições e características

As chamadas extranets de projeto são sistemas de informação baseados na Web (figura 2) para gerenciamento de documentos gerados em um empreendimento. Segundo Soibelman e Caldas (2000), as extranets de projeto, de forma genérica, podem ser definidas como uma rede de computadores que utiliza a Internet com o propósito de conectar empresas com seus parceiros que compartilham objetivos comuns.

Figura 2: Página de entrada de uma extranet de projeto

As extranets de projeto podem ser classificadas em extranets de armazenamento e de gerenciamento de projetos (NASCIMENTO; SANTOS, 2002a). As extranets de armazenamento apenas arquivam os documentos em um servidor na *Web* e possuem alguns recursos de correio eletrônico enquanto que as de gerenciamento possuem recursos como o de monitoramento do fluxo de documentos e processos (*workflow*), sistema de comunicação com notificações de novas atividades de projeto, reuniões virtuais, mensagens de novos documentos, circulares, recomendações, atualizações de arquivos, registro de operações e visualização de arquivos de diversos aplicativos.

Estes sistemas permitem acesso controlado aos documentos para os agentes participantes dos processos de um empreendimento. Esta tecnologia, que vem ganhando muitos adeptos, surgiu na segunda metade da década de 1990 através de empreendimentos conjuntos (*joint ventures*) de grandes companhias (ANDRESEN; CHRISTENSEN; HOWARD, 2003). Estes sistemas são razoavelmente baratos em relação a outros sistemas, como por exemplo, sistemas de gestão integrada ou ERP (*Enterprise Resource Planning*). Isto se deve em parte ao baixo custo (pelo menos para grandes empreendimentos) dos serviços dos provedores ASP (*Application Service Provider*), ao barateamento dos equipamentos de rede e ao aumento da velocidade de transmissão de dados na Internet.

As principais funcionalidades das extranets de projeto são:

- armazenamento de documentos (arquivos CAD, figuras, memorandos, planilhas, etc.);
- visualização de documentos (há visualizadores para vários tipos de arquivos);

- comunicação (correio eletrônico, grupos de discussão, vídeo conferência, etc.);
- automação dos processos e do fluxo de trabalho (*workflow*).

As extranets de projeto são o resultado de esforços ao redor do mundo para se conseguir maior produtividade e eficiência na indústria da construção civil (ANDRESEN; CHRISTENSEN; HOWARD, 2003). As extranets permitem melhorar a comunicação entre companhias que participam de um empreendimento (construtoras, projetistas, consultorias, fornecedores, incorporadoras, etc.) através de ferramentas que permitem a troca de informações digitais entre seus profissionais (engenheiros, arquitetos, tecnólogos, projetistas, consultores, etc.). Há uma melhoria na comunicação com estes sistemas porque a qualidade e a velocidade do fluxo das informações pode ser aumentada. Além disso, estes sistemas possuem ferramentas que possibilitam aos agentes participantes de um empreendimento armazenar, gerenciar e auditar todos os documentos gerados desde as etapas iniciais de estudo de viabilidade até o término da obra, disponibilizando informações vinte e quatro horas por dia, sete dias por semana. São registradas todas as inclusões, alterações e exclusões em todo o ciclo de vida dos documentos armazenados. A segurança das informações é mantida através de mecanismos de *hardware* e *software* normalmente utilizados em servidores Web (*firewall*, criptografia, *backups* e antivírus) e através do controle de acesso dos usuários (acesso, armazenamento ou alteração de documentos) definidos pelos coordenadores do empreendimento.

2.4 Usuários do sistema

Todos os participantes de um empreendimento da indústria da construção civil são potencialmente usuários das extranets de projeto (consultores, projetistas, técnicos, empreiteiros, fornecedores, pessoal administrativo, pessoal de *bureau* de plotagens e de empresas prestadoras de serviços, etc).

Com a adoção de uma extranet de projetos, o fluxo desordenado de comunicação entre os intervenientes de um empreendimento é mudado. Normalmente, o fluxo de comunicação entre os diversos intervenientes de um empreendimento é do tipo “todos para todos” (figura 3).

Figura 3: Comunicação típica de um empreendimento sem extranet de projeto

Neste modelo de comunicação há uma grande dificuldade do coordenador de projetos em controlar e gerenciar as informações geradas e trocadas entre os participantes de um empreendimento. Os intervenientes comunicam-se desordenadamente entre si,

muitas vezes definindo opções de projeto sem nenhum registro dessas decisões e sem o seu conhecimento ou de outro interveniente interessado e afetado por estas decisões. Estes fatos podem levar a problemas no desenvolvimento de um empreendimento e causar dificuldades para o coordenador compatibilizar os projetos com suas equipes. Além disso, fica muito mais difícil rastrear informações e compartilhá-las com todos os intervenientes de um empreendimento. A utilização de sistemas de informação na indústria da construção civil tem como um de seus principais objetivos a solução destes problemas de comunicação.

Figura 4: Comunicação em um empreendimento com extranet de projeto

Quando se adota uma extranet de projeto em um empreendimento, o fluxo de informações fica centralizado e controlado pela extranet de projeto (figura 4). Deve-se evitar comunicação entre os intervenientes através de mecanismos externos ao sistema (telefone, fax, reuniões presenciais sem

registro, etc.) ou documentar claramente tais comunicações, inserindo-as prontamente no sistema.

O sucesso de uma extranet de projeto está ligada à sua intensidade de utilização pelos participantes de um empreendimento. Em pesquisas feitas com empreendimentos que utilizam extranets de projeto no Brasil (PICORAL; SOLANO, 2001) e nos Estados Unidos (SOIBELMAN; CALDAS, 2000) percebeu-se que é proporcionalmente pequena a parcela de intervenientes que não utilizam o sistema. Dawood, Akinsola e Hobbs (2002) consideram a efetiva comunicação entre toda a cadeia produtiva como um pré-requisito para se conseguir maior produtividade na indústria da construção civil. As extranets de projetos viabilizam um meio de comunicação onde empreendedores, projetistas e construtores podem harmonizar suas relações em prol dos interesses do desenvolvimento de um empreendimento.

2.5 Benefícios da Tecnologia

Existe uma grande expectativa na comunidade do setor da construção civil de que as extranets permitam aumentar a eficiência de seus processos podendo assim contribuir para um aumento de sua produtividade. Muitos dos benefícios obtidos com a TI estão ligados aos recursos disponíveis pela Internet, como notificações automáticas do andamento de um empreendimento por e-mail, reuniões virtuais das equipes de um projeto em salas de bate papo e fotos digitais que mostram o andamento da obra (figura 5).

Figura 5: Página de uma extranet de projeto com foto de uma obra

As empresas brasileiras que mantêm estes serviços (entre outras, SADP¹, Neogera² e Construtivo³) divulgam muitos dos benefícios que os sistemas podem trazer:

- maior eficiência na transferência de documentações de projeto entre empresas do que os meios eletrônicos tradicionais, como *e-mail* e disquete, que têm capacidade limitada;
- melhor documentação do envio e recebimento de arquivos;
- eliminação de erros de planejamento e execução devido à falta de comunicação referente às mudanças de projeto;
- economia de tempo pelo aumento da velocidade do fluxo de informações trocadas entre os intervenientes;
- redução nos custos com impressões e plotagens já que os sistemas permitem a visualização direta dos arquivos digitais;

¹ <http://www.sistrut.com.br/sadp>

² <http://www.neogera.com.br>

³ <http://www.construtivo.com>

- redução do tempo de ciclo de projeto e do custo com transporte pela facilidade de comunicação de agentes dispersos fisicamente.

As grandes empresas do exterior que oferecem o mesmo serviço, como a Citadon⁴, Viecon⁵ e Buzzsaw⁶, também divulgam os mesmos benefícios com ênfase no aumento do lucro e redução de riscos, tempo gasto com gerenciamento de informações (em torno de 30%) e do volume da papelada gerada ao longo do desenvolvimento de um projeto. Ao final de um empreendimento, normalmente toda a documentação armazenada na base de dados da extranet de projeto é arquivada permanentemente em CD-R (Compact Disc) para a empresa contratante do sistema.

2.6 Problemas da Tecnologia

Apesar da grande expectativa gerada com o surgimento das extranets de projeto, existem alguns problemas ainda não solucionados no uso desta tecnologia. Soibelman e Caldas (2000) citam como principais problemas a falta de adequação ao fluxo de informações, a ausência de tecnologias eficientes para indexar e buscar informações e o acúmulo de informação desnecessária pela falta de conhecimento e de critérios para se avaliar a qualidade da informação. Para Andressen, Christensen e Howard (2003), as principais barreiras encontradas no uso de uma extranet de projeto são a falta de planejamento e coordenação sem a adoção de objetivos comuns (reduzir custo, tempo, precisão em desenhos CAD, reduzir o número de plotagens, etc.) para todas as empresas participantes do processo, a dificuldade em elaborar contratos especificando direitos e responsabilidades de cada agente

⁴ <http://www.citadon.com>

⁵ <http://www.viecon.com>

⁶ <http://www.buzzsaw.com>

participante, a legalidade dos documentos eletrônicos (em substituição aos de papel), e o alto custo de implementação (pelo menos na Dinamarca, em relação aos poucos benefícios com este investimento em tecnologia da informação). Além disso, como as extranets ainda são consideradas inovações tecnológicas que requerem estratégias organizacionais e mudanças de processos, poderá ocorrer resistência cultural e organizacional à sua implantação (CRUZ, 1998).

2.7 Tipos de documentos armazenados nas extranets de projeto

Nos processos da indústria da construção civil há uma intensa troca de informações entre os intervenientes por meio de uma grande variedade de documentos eletrônicos. Tais documentos são armazenados no sistema pelos usuários e podem se apresentar na forma de diferentes tipos de dados (MAHER; SIMOFF, 1998):

- **dados estruturados** – tabelas de bancos de dados relacionais ou estruturas de dados orientados a objeto;
- **dados fracamente estruturados** – textos livres ou em formato de tabelas, gráficos vetoriais como desenhos CAD;
- **dados brutos** – imagens *raster* de fotografias ou desenhos, imagens animadas, dados de áudio e vídeo;
- **dados de links** – *links* e *hiperlinks* entre dados (utilizados em derivados do SGML como HTML e XML).

Os tipos de dados estruturados, apesar de serem encontrados dentro do sistema de uma extranet de projeto, não são normalmente armazenados por seus usuários, pois são processados por aplicativos específicos de cada cliente para controle de estoques, compras, finanças, etc. Dados brutos costumam se apresentar em arquivos muito grandes, portanto, é

evitada sua utilização em sistemas que utilizam tecnologia Internet como as extranets de projetos. Os *links*, por fazerem parte de uma estrutura de uma página da *Web*, também fazem parte do sistema de uma extranet. A maior parte dos documentos armazenados em uma extranet normalmente são do tipo fracamente estruturados, constituindo em torno de 80% dos dados de uma extranet de projeto (SOIBELMAN; CALDAS, 2001). Estes dados podem estar contidos em arquivos gerados por programas editores de texto (memorial descritivo, memorando, etc.), planilhas eletrônicas (orçamento, lista de material, etc.) e aplicativos CAD (plantas, cortes, detalhes, etc.). Em um breve levantamento feito pelo autor com extranets de projeto em São Paulo, verificou-se que em média mais de 90% dos arquivos armazenados eram arquivos CAD ou arquivos para plotagem (.plt).

O sistema apresentado neste trabalho recupera informações em dados fracamente estruturados ou em registros de texto que descrevem todos os outros tipos de documentos armazenáveis em uma extranet de projeto (meta-dados), como aqueles dos formulários recomendados por Picoral e Solano (2001).

2.8 Acesso às informações

As extranets de projetos têm como um de seus principais objetivos o de solucionar o problema da falta de informação para elaboração de projetos da indústria da construção civil. Este problema tem sido detectado em vários estudos nos últimos anos (SCHMITT, 1993), (ANDERY; CARVALHO JR; HELMAN, 1998), (DAWOOD; AKINSOLA; HOBBS, 2002) e sua solução é considerada como um dos meios pelos quais se pretende alcançar maior produtividade e qualidade na indústria da construção civil. O problema é antigo e pode resultar em patologias nas construções. Lichtenstein (1985)

destaca que aproximadamente 40% dos problemas patológicos nas construções são originados por falhas em projetos. O autor cita ainda a falta de clareza nas informações como um dos fatores para estas falhas. Para Tzortzopoulos (1999), ao longo dos processos de um empreendimento, muitas decisões essenciais não são tomadas adequadamente por falta de tempo ou pressões do mercado. Para a mesma autora, estas decisões são tomadas sem considerar todas as informações necessárias podendo gerar perdas como, por exemplo, retrabalhos, indefinições do produto e consideração inadequada ou insuficiente das necessidades dos clientes no projeto e no desenvolvimento de estudos de viabilidade econômica. Para Melhado (2001), em um ambiente de gestão da qualidade, o processo de projeto deve estar voltado ao atendimento das necessidades de informação de todos os clientes internos que atuam no ciclo de produção do empreendimento.

Em um empreendimento típico, verifica-se que são gerados milhares de documentos. Quando um grande volume de informação está circulando entre os membros de um projeto, é fundamental garantir-se que estas informações sejam precisas e relevantes (CALDAS; SOIBELMAN, 2001). Picoral e Solano (2001), em pesquisa com extranets de projeto no Brasil, recomendam o preenchimento de formulários eletrônicos para a descrição dos arquivos transferidos para o sistema. Estes formulários auxiliam na identificação dos arquivos evitando que tenham que ser abertos para se verificar seu conteúdo. Além disso, o autor evidencia também a importância de anotar em formulários as alterações nas versões dos arquivos. A figura 6 mostra um formulário para transferência e descrição de arquivos em uma extranet de projeto nacional.

CADASTRAMENTO DE ARQUIVOS

Cadastramento ou Atualização de **novo** arquivo.
No cadastramento do arquivo, deverá ser escolhido um NOME ainda não existente entre os arquivos do CLIENTE.

- **Selecione arquivo na área temporária**
Arq-32-rev0.dwg
☐ **SIM**, o arquivo selecionado é MÚLTIPLO CADASTRAMENTO
(VEJA [COMO CADASTRAR VÁRIOS ARQUIVOS SIMULTANEAMENTE](#))
- **Nome completo que o arquivo terá no sistema para DownLoad**
Arq-32-rev0.dwg (Padrões da [ASBEA](#))
não preencher para manter o nome original
- **Projeto o qual este arquivo pertence**
Ricon
- **Nível de acesso para o arquivo**
Arquitetura
- **Pasta do Projeto**
Alterações de Obra
- **Comentário sobre o arquivo**
Levantamento as built da edícula do sobrado B14 do Condomínio Sol Nascente com indicação da locação do Quadro de Luz (QDL).

Figura 6 – Formulário transferência e descrição de arquivo de uma extranet de projeto.

Algumas extranets de projeto dispõem de ferramentas para localização de informações em documentos de suas bases de dados. A maioria desses sistemas dispõe apenas de uma ferramenta de busca de arquivos baseada em pesquisa direta de palavras-chave. As buscas por palavras-chave apenas localizam e recuperam documentos que apresentam em seu conteúdo os termos especificados na consulta. As figuras 7 e 8 ilustram o formulário de busca de duas extranets de projeto. A primeira de origem americana, mas traduzida para o português e a segunda de uma grande extranet de projeto americana com atuação em várias partes do mundo. A figura 9 mostra a página de resultados da primeira extranet de projeto. Alguns mecanismos de busca podem procurar documentos por palavras chave nos nomes dos arquivos. Esta não é uma boa solução, pois algumas empresas padronizam a criação de nome de arquivos através de códigos. Estes códigos são formados por letras e números de acordo com informações como a disciplina do projetista e o tipo do desenho. Existem vários destes padrões ao redor do mundo sendo que no Brasil é muito conhecido

o desenvolvido pela ASBEA que é descrito no capítulo 5 deste trabalho.

NEO GERA

Página Inicial | Aplicativos | Ajuda | Sair | Portuguese

prof. eduardo toledo

Neogera > Edifício Neogera > Buscar

Tipo de Busca: Documento

Busca Avançada de Documentos:

Nome:

Descrição:

limpar cancelar localizar documento

©2001 EPC América Latina S.A. - Todos os direitos reservados

Figura 7 – Formulário de busca de uma extranet de projeto utilizada no Brasil.

ProjectNet - Search RFI - Microsoft Internet Explorer

Search RFI

Keyword: contains

Match case: ☐

Search on: ☒ Subject ☐ Requested Information ☐ Proposed Solution ☐ Reference # ☐ Co-Author

Status: ☐ Open ☐ Closed ☒ Open/Closed

Created By: All Barry Milliken Carina Nicklas Charlie Brown

Assigned To: All Barry Milliken Carina Nicklas Charlie Brown

Note: *Italicized fields are optional* search cancel

Figura 8 – Formulário de busca de uma grande extranet de projeto americana.

Figura 9 – Resultado de uma busca em uma extranet de projeto

Estes mecanismos de busca podem localizar informações a partir dos textos contidos nos arquivos ou nos descritores dos mesmos. Grande parte dos documentos armazenados nestes sistemas é criada por ferramentas CAD ou estão agrupados em arquivos comprimidos (para reduzir espaço ocupado na base de dados e diminuir o tempo de transferência para o sistema pela Internet). Por estes motivos é comum a utilização da varredura do conteúdo de descritores o que torna importante o detalhamento da descrição dos documentos transferidos no sistema para sua localização.

Os mecanismos de busca das extranets de projeto podem retornar dezenas de arquivos de documentos em uma consulta, dificultando a localização da informação desejada. Dessa forma, as extranets de projeto podem, por um lado, ajudar na obtenção rápida de dados atualizados (WATSON; DAVOODI, 2002) mas, por outro, causar a chamada “sobrecarga de informações” (*information overload*) (FARHOOMAND; DRURY, 2002). A sobrecarga de informações ocorre quando indivíduos recebem mais informação relevante do que conseguem absorver (FARHOOMAND; DRURY, 2002), ou quando é enviada grande quantidade de informação não solicitada, parte dela potencialmente relevante (EDMUNDS; MORRIS, 2000).

O problema de encontrar informações previamente armazenadas é estudado pela área denominada “Recuperação de Informações”, uma parte da Ciência da Computação. Este tema, central neste trabalho, será detalhado no capítulo seguinte.

3 RECUPERAÇÃO DE INFORMAÇÃO

3.1 Definições e características

Recuperação de Informação (*Information Retrieval*) é uma parte da Ciência da Computação que estuda a representação, armazenamento, organização e acesso à informação a partir de coleções de documentos. Porém, o problema de recuperar informações relevantes e oportunas existe desde antes da invenção dos computadores. Segundo Rijsbergen (1979), o problema teve um aumento de interesse desde a década de 1940. Posteriormente, com a utilização de computadores, este problema foi parcialmente solucionado, mas ainda existem dificuldades relacionadas à complexidade e subjetividade da relação homem-máquina. Para o homem existe a limitação de selecionar informações relevantes, em tempo satisfatório, em um grande volume de dados e, em relação à máquina, existe a dificuldade em criar-se algoritmos que julguem a relevância das informações nos documentos e interpretem totalmente todos os tipos de arquivos em uma base de dados.

Nos últimos anos têm ocorrido mudanças nas pesquisas e trabalhos com Recuperação de Informações (RI). Atualmente há interesse apenas em estudar o desenvolvimento de sistemas de recuperação de informação automáticos. *Automático* como “oposto” a *manual* e *informação* como “oposto” a *dado* (RIJSBERGEN, 1979). Os principais motivos são a proliferação de computadores pessoais, a expansão da Web, o barateamento de dispositivos de armazenamento de massa de dados, a utilização de sistemas com interface gráfica e as novas aplicações multimídia.

Os principais objetivos da Recuperação de Informações são construir índices eficientes, processar consultas de usuários com bom desempenho e desenvolver algoritmos de ordenação (*ranking*)

que aumentem a qualidade das respostas. Atualmente, pesquisas em Recuperação de Informação incluem modelagem, classificação, arquitetura de sistemas, interface de usuários, visualização de dados, filtragem e linguagens (Baeza-Yates e Ribeiro-Neto; 1999).

A *consulta* em RI é a forma de comunicação do usuário para introduzir sua necessidade de informação em um sistema de informação. A necessidade de informação, segundo Mizzaro (1996), pode ser definida como a falta de conhecimento que o usuário tem para realizar determinada tarefa. Para solucionar determinado problema, o usuário necessita de determinado conhecimento. Se ele não possui esse conhecimento ele passa a ter, então, uma necessidade de informação. Essa necessidade de informação deve ser satisfeita de alguma forma. Geralmente isso é obtido através de uma consulta em um Sistema de Recuperação de Informações.

A necessidade de informação normalmente não é especificada na interface do mecanismo de recuperação em linguagem natural e sim através de palavras-chave (Baeza-Yates e Ribeiro-Neto; 1999). A introdução dos dados de uma consulta por palavras-chave usualmente é feita através de um formulário eletrônico. Após a introdução dos dados de pesquisa, o sistema de RI inicia uma busca para localizar documentos que provavelmente possam ser relevantes para o usuário. Um sistema de busca não tem a responsabilidade de responder precisamente a uma consulta mas sim de identificar documentos que permitam que o usuário satisfaça sua necessidade de informação.

Um aspecto importante em RI que influencia no resultado de uma busca é a maneira que as informações contidas em documentos são representadas. Para recuperar informações através de técnicas de RI é necessário varrer uma grande quantidade de documentos para tentar encontrar a informação desejada. Estes documentos

usualmente estão armazenados em um sistema de banco de dados. Atualmente pode-se representar um documento por todas as palavras do seu texto. Este tipo de representação apresenta uma boa visão do conteúdo de um documento, mas requer grande esforço computacional. Em grandes massas de dados, pode-se reduzir esta representação a algumas palavras-chave. As palavras-chave podem ser extraídas automaticamente do texto dos documentos ou podem ser especificadas por um humano. Frequentemente é criado um índice de termos com as palavras-chave contidas nos documentos. Este índice pode ser formado automaticamente através da eliminação de *stopwords* (palavras muito comuns, com baixo potencial para influenciar o resultado de uma busca como preposições, artigos, conjunções, etc. - correspondem a aproximadamente 40% das palavras em um texto (FUSARO, 1988)) e o uso de lematização (*stemming*). Esta técnica captura apenas as raízes das palavras, eliminando prefixos e sufixos, reduzindo a quantidade de termos diferentes mas relacionados. Assim, permite a captura do núcleo significativo de uma palavra, identificando termos da mesma família gramatical independente de sua flexão, como tempo, modo, gênero, número, etc. Neste trabalho serão utilizadas 70 *stopwords* em português propostas por Fusaro (1988) e um algoritmo de lematização adaptado de Ribeiro (2003).

A eliminação de *stopwords* e a lematização são chamadas de operações de texto. As operações de texto reduzem a complexidade da representação de documentos e permitem a mudança da visão de todo o conteúdo dos documentos para uma simplificada com as principais palavras-chaves que representam os documentos. Segundo Fusaro (1988), a eficácia destes índices de termos (palavras-chave) depende da exaustividade da indexação

(se foram considerados os vários assuntos de um documento) e da especificidade dos termos de indexação (o grau de detalhe dos termos de indexação).

3.2 Sistemas de Recuperação de Informação

Um Sistema de Recuperação de Informações (SRI) é formado por máquinas e programas de computadores utilizados para recuperar informações em bases de dados onde são armazenados documentos (Baeza-Yates e Ribeiro-Neto; 1999). Estes sistemas executam operações que envolvem processos de pesquisa, indexação e estratégias de busca, relevância e apresentação de documentos de um sistema de informação através de mecanismos de busca de documentos.

Apesar de se usar os termos Recuperação de Informação e Sistema de Recuperação de Informação, ambos, na verdade, se referem ao estudo ou técnica de melhor recuperar *documentos* (onde possivelmente contenham informações relevantes), sendo assim poderiam se chamar Recuperação de Documentos e Sistemas de Recuperação de Documentos. A função mais básica destes sistemas é procurar palavras-chave nos documentos de uma base de dados para determinar quais deles são possivelmente relevantes para uma consulta. Outra função importante é escolher a melhor forma de apresentar os documentos recuperados da coleção, dando prioridade na exibição para os documentos com maior probabilidade de relevância. A eficácia de um SRI está relacionada com a melhor forma de representar os documentos e com a melhor maneira de recuperá-los de acordo com os argumentos disponíveis na consulta (RIJSBERGEN, 1979). Um SRI pode ser representado conforme a figura 10.

Figura 10 – Sistema de Recuperação de Informações

As operações de entrada no SRI são representadas pelas atividades de introdução de dados e consultas dos usuários no sistema. Os dados criados ou transferidos para um sistema de informação são processados pelo SRI e armazenados em uma base de dados. O SRI normalmente armazena a representação do documento por meio de operações de texto que geram uma visão lógica dele através de palavras-chave que o descrevem para que possa ser recuperado pelo usuário em uma consulta. Dependendo do formato do arquivo, o SRI pode armazenar todo o texto do documento. Porém, em alguns formatos de documentos (gráficos e multimídia) não é possível a identificação e o registro de todo o seu conteúdo. A eficácia do SRI nestes casos está relacionada com a melhor forma de combinar (*matching*) a representação dos documentos com a representação de sua necessidade de informação na consulta. Após capturar as informações do documento, o gerenciador de banco de dados do sistema cria um índice (lista de termos com ponteiros para as posições dos documentos) com as palavras-chave de modo que permita pesquisa rápida em uma

grande quantidade de documentos. A estrutura de índice mais utilizada é a de *arquivos invertidos*.

As consultas normalmente são feitas preenchendo-se formulários onde os usuários especificam as suas necessidades de informação. Estes formulários devem ter condições e opções para que o usuário especifique argumentos para localizar uma informação em um documento, normalmente com palavras-chave que são procuradas nos descritores dos documentos armazenados. Em seguida, os documentos possivelmente relevantes são ordenados e apresentados para o usuário que os examina para localizar a informação desejada.

Para processar uma consulta em um SRI, os algoritmos de busca selecionam os documentos possivelmente relevantes através de seus descritores e descartam todos os outros. Alguns SRIs podem controlar o tempo máximo de resposta de uma busca e a quantidade máxima de documentos que serão apresentados por vez nos resultados de uma consulta. Algoritmos de busca processam as regras para quantificar e determinar a relevância dos documentos. Estas regras identificam as similaridades da representação da informação feita pelo usuário na consulta e as representações ou descrições (palavras-chave) dos documentos armazenados pelo SRI. Em representações feitas através de palavras-chave, deve-se considerar a semelhança do vocabulário utilizado na consulta e nas representações dos documentos. Um usuário pode não receber alguns documentos relevantes caso especifique um sinônimo de uma palavra ou receber documentos com informações irrelevantes caso especifique palavras homônimas.

O resultado de uma consulta em um SRI apresenta uma lista de documentos que possivelmente satisfazem à necessidade do

usuário. Geralmente, os algoritmos de busca tentam apresentar primeiro os documentos mais relevantes, normalmente determinados de acordo com fatores probabilísticos (número de palavras chave, número de vezes em que uma palavra chave é encontrada em um documento, documentos mais consultados, mais recentes, etc.) e algoritmos de ordenamento (*ranking*). Estes documentos podem ser apresentados em forma de lista com um pequeno resumo com sua descrição. Alguns sistemas podem ser retroalimentados através de novas buscas apenas nos documentos recuperados.

3.3 Avaliação de um sistema de recuperação de informação

Para avaliação dos resultados de um sistema de recuperação de informação existem diversas técnicas. Segundo Fusaro (1988), um sistema de recuperação de informações pode ser avaliado sob o ponto de vista de seus operadores e gerentes (medidas da relação custo / benefício) e sob o ponto de vista de seus usuários (medidas relacionadas à capacidade do sistema em recuperar documentos úteis). Neste trabalho serão estudadas medidas relacionadas ao ponto de vista dos usuários (profissionais da indústria da construção civil). As duas medidas de desempenho mais utilizadas são as de precisão e revocação (BAEZA-YATES; RIBEIRO NETO, 1999), ilustradas na figura 11. A precisão (*precision*), também chamada de pertinência, é a porcentagem dos itens recuperados que são relevantes. A revocação (*recall*), também chamada de abrangência, é a porcentagem de itens relevantes que foram recuperados.

Figura 11 – Avaliação de sistemas de recuperação de informações

Para aumentar a precisão de uma busca, o sistema de recuperação de informações deve ser capaz de descartar itens recuperados que são irrelevantes. Para isso, deve saber contextualizar os dados recuperados, de forma a identificar falsos relacionamentos que levaram à inclusão inicial de um documento no conjunto de itens recuperados. Para melhorar a revocação, o algoritmo de busca utilizado deve perceber a relevância de documentos de forma não explícita. Em ambos os casos, o conhecimento do domínio da aplicação pode ajudar.

Segundo a Lei de Cleverdon, a precisão varia em função inversa da revocação (BARSOTTI, 1990). Em uma grande coleção, é preferível a precisão para evitar uma grande quantidade de documentos apresentados que pode inviabilizar a localização da informação desejada nos documentos. Em uma coleção pequena, é preferível a revocação, já que mesmo que haja alguns documentos irrelevantes se terá uma maior garantia de que os documentos relevantes foram recuperados.

Os cálculos da precisão e da revocação de um sistema são feitos através da determinação de alguns fatores de desempenho do sistema de recuperação (VICKERY, 1970) tais como os obtidos na Tabela de Contingência (tabela 1).

	Relevantes	Irrelevantes	
Recuperados	rr	ri	TR = rr + ri
Não recuperados	nr	ni	TN = nr + ni
	TC = rr + nr	TI = ri + ni	N = rr + nr + ri + ni

Tabela 1. Tabela de Contingência

Onde:

- rr = número de documentos relevantes recuperados;
- nr = número de documentos relevantes não recuperados pela pesquisa;
- TC = núm. total de documentos armazenados, relevantes para uma pesquisa;
- ri = número de documentos recuperados, porém irrelevantes para a pesquisa;
- ni = número de documentos não recuperados e irrelevantes para a pesquisa;
- TI = número total de documentos irrelevantes para uma pesquisa;
- TR = número total de documentos recuperados em uma pesquisa;
- TN = número total de documentos não recuperados em uma pesquisa;
- N = número total de documentos armazenados no sistema.

Para calcular a medida de revocação (**R**), determina-se, em porcentagem, o quociente do número de documentos relevantes para uma pesquisa que foram recuperados pelo sistema, pela quantidade total de documentos relevantes para a consulta, conforme a eq. (1).

$$R = 100 \times \frac{rr}{TC} \quad (1)$$

A precisão (**P**) é calculada, em porcentagem, através do quociente do número de documentos relevantes para uma pesquisa que foram recuperados pelo sistema, pela quantidade total de documentos recuperados, conforme a eq. (2).

$$P = 100 \times \frac{rr}{TR} \quad (2)$$

A revocação mede o sucesso de um sistema em evitar perda de documentos relevantes e a precisão mede o sucesso em evitar a recuperação de documentos irrelevantes (VICKERY, 1970).

A partir da década de 90, foram padronizadas algumas metodologias para avaliação de sistemas de recuperação de informação em uma conferência anual internacional promovida pelo NIST (*National Institute of Standards and Technology*) e pela DARPA (*Defence Advance Research Projects Agency*) chamada TREC - *Text REtrieval Conference* (<http://trec.nist.gov>). Para cada conferência, os participantes testam seus sistemas de forma padronizada em uma mesma coleção de documentos de forma a poder comparar seu desempenho. A coleção de documentos normalmente é composta de diversas fontes de informação, tais como, *The Federal Register*, *Abstratcs*, *The San Jose Mercury*, *The Wall Street Journal*, *The Foreign Broadcast Information Service*, entre outros. Até este momento, a grande maioria dos documentos da base de testes do TREC é em inglês, existindo alguns documentos em chinês, espanhol, alemão, francês e italiano. Por essa razão, a base de dados do TREC não pode ser usada neste trabalho. No entanto, a metodologia utilizada nos testes de avaliação da ferramenta de busca apresentada neste trabalho é baseada nas metodologias do TREC (capítulo 7).

Outra razão que impediu o uso da base de documentos do TREC é o fato de que trata-se de uma base genérica, não especializada no

tema “Construção Civil”. A ferramenta concebida neste trabalho faz uso do conhecimento específico desta área para melhorar o desempenho da busca. O capítulo seguinte trata da representação deste conhecimento para uso do sistema de recuperação de informação.

4 REPRESENTAÇÃO DO CONHECIMENTO

A representação do conhecimento é um método utilizado na área da inteligência artificial para codificar o conhecimento de forma simbólica, normalmente para uso em um sistema especialista (DURKIN, 1994). Conhecimento e sistema especialista são definidos pelo mesmo autor, respectivamente, como a compreensão de uma disciplina e um programa de computador projetado para modelar a habilidade de resolver problemas de um especialista humano.

As principais técnicas empregadas são a *representação por regras*, as *redes semânticas* e a *representação por frames*. Este último termo freqüentemente é confundido com o termo Sistema Baseado em Conhecimento, que é mais abrangente e cobre grande número de técnicas de inteligência artificial (KARP, 1993).

4.1 Representação por regras

A representação por regras é uma forma de conhecimento procedimental. É uma estrutura de conhecimento que relaciona uma informação conhecida a uma outra que dela pode ser inferida ou concluída (DURKIN, 1994). As regras conectam uma ou mais premissas a uma ou mais conclusões. Regras executam operações e associam uma informação a uma ação.

4.2 Redes Semânticas

Rede semântica é um método de representação do conhecimento que utiliza grafos constituídos por nós e arcos onde os nós representam objetos e os arcos representam as relações entre os objetos (MULLIER; DIXON, 2000). As relações em redes semânticas podem ser, entre outras:

- **ako** [a-kind-of] (um tipo de) – é uma relação entre duas classes através de herança especificando que uma é um subconjunto da outra, mais geral;
- **is-a** (é um) – é uma relação entre classe e instâncias, isto é, indica que um objeto (instância) pertence a uma certa classe;
- **has-a** (tem um) – é uma relação onde são identificados atributos ou características de uma entidade;
- **part-of** (parte de) – representa a relação entre uma entidade e seus componentes ou atributos.

Uma das grandes vantagens da utilização de redes semânticas é que elas são uma representação de fácil visualização e possuem grande flexibilidade na manipulação de seus elementos.

4.3 Representação Baseada em Frames

A estrutura adotada neste trabalho é a da Representação Baseada em *Frames*, capaz de representar conceitos e objetos (KARP, 1993). A representação baseada em *frames* foi proposta originalmente por Marvin Minsky em 1975 (DURKIN, 1994). Um *frame* é definido como uma estrutura de dados para representar um conhecimento estereotipado de algum conceito ou objeto (DURKIN, 1994). Os quadros desta estrutura possuem propriedades ou atributos chamados de *slots*. Os *frames* integram conhecimento declarativo sobre objetos, eventos e conhecimento procedimental sobre como recuperar informações ou calcular valores. As vantagens da Representação por *Frames*, segundo Sulzbach (2001), é que esta é uma representação próxima a dos objetos, possibilita fácil visualização e compreensão do conhecimento, possui modularidade na modelagem e sintaxe

legível. Os *frames* possuem vantagens em relação às redes semânticas já que procedimentos podem ser anexados em sua estrutura. Os *frames* permitem herança múltipla. Cada atributo do *frame* pode apontar para um outro *frame* ou para um tipo de dado básico (como uma cadeia de caracteres).

Os *frames* são arranjados em uma hierarquia taxonômica em que cada *frame* é ligado a um ou mais *frames*-pai (herança múltipla). Um *frame*-pai representa um conceito mais geral e um *frame*-filho um conceito mais específico. A coleção de *frames* que formam esta hierarquia de heranças forma uma base de conhecimento (KARP, 1993). Uma hierarquia taxonômica é uma hierarquia de generalização-especialização com classes e subclasses. Um *frame* criado a partir de um *frame*-pai herda seus *slots* de forma que estes *slots* guiam o novo conceito formado pelo *frame*-filho. A figura 12 apresenta um exemplo de representação por *frames*.

Figura 12: Exemplo de Representação por *Frames*

Para a construção de uma base de conhecimento é necessário, além da definição da forma de representação do conhecimento, tratada aqui, dispor-se ou capturar-se o conhecimento que será

armazenado nela. O capítulo seguinte trata das estruturas de classificação para a construção civil que serão usadas como a fonte do conhecimento utilizado no sistema de recuperação de informação proposto neste trabalho.

5 ESTRUTURAS DE CLASSIFICAÇÃO PARA A CONSTRUÇÃO

Existe um grande esforço em todo o mundo para integrar as informações de diversas indústrias. Na indústria da construção civil não é diferente. Algumas entidades (ISO, IAI, CSI, CSC, AsBEA, ANTAC, etc.) procuram soluções através de padronizações e classificações. Para isso é necessário que se adotem padrões que especifiquem produtos, serviços e processos. Para estas entidades a padronização de informações na indústria da construção civil é necessária e tem como objetivos:

- melhorar a integração entre os agentes através de informações padronizadas;
- aumentar a produtividade com a melhora na logística através da padronização;
- reduzir a variedade de nomenclaturas utilizadas na comunicação do setor para produtos, serviços e processos;
- permitir sistemas de informações e de comunicação mais eficazes;
- possibilitar o aumento do capital intelectual das companhias através da melhoria da qualidade das informações por meio de métodos, técnicas e ferramentas computacionais que utilizem informações padronizadas para a gestão do conhecimento;
- permitir a criação de sistemas interoperáveis através do desenvolvimento de ferramentas computacionais.

5.1 Padrões ISO

A maioria das entidades no mundo que estão estudando e propondo modelos para criar padrões de classificações para o setor da construção civil estão se baseando em normas da ISO (*International Organization for Standardization*). A classificação no

setor da construção é baseada em conceitos que representam as propriedades que são de interesse no processo de projeto, no processo de construção e no processo de gerenciamento. Cada um desses processos contém um conjunto de fases e atividades que podem ser conceituados terminologicamente.

5.1.1 ISO 14177

A ISO desenvolve padrões adotados em todo o mundo em vários segmentos, sendo que, em relação à classificação de informações na indústria da construção civil, foi desenvolvido pelo grupo do comitê de trabalho ISO/TC59/SC13 o ISO Technical Report (1994) “*Classification in the construction industry*”. Em linhas gerais, o relatório classifica a indústria da construção em classes e tabelas como mostra a tabela 2:

Tipo	Classes	Exemplos de Elementos	Exemplos de Usos
Recursos físicos e resultados da construção	<i>Facilities</i> *Facilidades	Enfermaria, estação de força, rede de fornecimento de água e restaurante.	Dados históricos de projeto e custos.
	<i>Spaces</i> *Espaços	Recreação, armazém, circulação, escritório, aeroporto e residencial.	Especificações e desenhos de projeto.
	<i>Elements</i> *Elementos	Fundação, coluna, pavimento, linha de transmissão e paredes externas.	Bibliotecas e <i>layers</i> em sistemas CAD.
	<i>Work sections</i> *Etapas de trabalho	Escavação e aterramento, drenagem, fechamento e instalação de elevador.	Cálculos de custos e transações EDI.
	<i>Construction products</i> *Produtos da construção	Cimento, concreto, tijolo, porta, janela, tubo e cabo elétrico.	Banco de dados de produtos.
	<i>Construction aids</i> *Apoio para construção	Máquina e equipamentos provisórios para construção (andaime, grua, etc).	Banco de dados de equipamentos.
Informação contextual	<i>Management</i> *Administração	Contratos, legislação, treinamento, análise e controle de custos.	Produção de documentos.
	<i>Attributes</i> *Atributos	Material, formato, tamanho, precisão, propriedades acústicas e térmicas.	Estruturação de banco de dados de produtos.

Tabela 2: Classes do Technical Report 14177 da ISO.

A principal finalidade desta norma é permitir uma maior qualidade do fluxo de informação durante a criação e uso da construção e dar uma linha de direção para a organização da informação na indústria da construção. As recomendações desta norma visam melhorar o fluxo de informação entre países ou em um país específico.

5.1.2 ISO 10303 - STEP

Além do *Technical Report* 14177, a ISO também criou um modelo de dados dos produtos da construção através de diversas normas e padrões chamados de ISO 10303, desenvolvidos pelo grupo STEP AEC do comitê ISO TC184/SC4. Esta norma também é conhecida como STEP (*Standard for the Exchange of Product Model Data*). O STEP é descrito por uma grande quantidade de documentos onde cada documento descreve um assunto específico do padrão. Por exemplo, a ISO 10303-11, descreve a linguagem Express, que é uma linguagem formal (*DDL - Data Definition Language*) para descrever o modelo do produto ou a ISO 10303-21, que descreve um formato de arquivo de intercâmbio de dados. A documentação do STEP é dividida em grupos de documentos numerados de acordo com a tabela 3:

Documentos STEP	Número dos documentos
Introdução	0 a 9
Métodos de descrição (Linguagem Express)	11 a 13
Métodos de implementação (<i>Standard Data Access Interface</i>)	21 a 26
Metodologia de teste e conformidade	31 a 35
Recursos genéricos integrados	41 a 49
Recursos integrados de aplicação	101 a 106
Protocolos de aplicação	201 a 233
Suítes de testes abstratos	301 a 332
Construções de aplicações interpretadas	501 a 518

Tabela 3 - Grupos de documentos do STEP

O padrão STEP está sendo desenvolvido com a finalidade de descrever e controlar os dados do produto durante todo o seu ciclo de vida.

5.2 Sistemas de classificação no mundo

Outras entidades formadas por grupos de companhias ligadas ao setor da construção civil e de tecnologia da informação também estão desenvolvendo projetos para padronizar e classificar informações da construção. Alguns dos principais projetos ao redor do mundo são o OCCS (*OmniClass™ Construction Classification System*) e o Master Format™.

5.2.1 Master Format

O projeto mais antigo é o Master Format que teve sua estrutura introduzida em 1963, quando foi aceita como prática padrão no Canadá e Estados Unidos. Sua primeira edição saiu em 1978 com a união dos institutos *Construction Specifications Canada* (CSC) e *Construction Specifications Institute* (CSI) dos dois países, respectivamente. Sua estrutura apresenta 16 divisões criadas para organizar a informação de produtos e atividades da construção. O sistema pode ser usado para padronizar, organizar e arquivar dados de custo, de projeto, de produto e de outras características técnicas. Em sua última versão (1995) possibilita acomodar todas as fases da construção e permite expansões futuras, garantindo compatibilidade com outros padrões como o OCCS. Esta última versão apresenta, de acordo com a tabela 4, as seguintes divisões:

Nº	Divisão	Nº	Divisão
1	<i>General Requirements</i>	9	<i>Finishes</i>
2	<i>Site construction</i>	10	<i>Specialties</i>
3	<i>Concrete</i>	11	<i>Equipment</i>
4	<i>Masonry</i>	12	<i>Furnishings</i>
5	<i>Metals</i>	13	<i>Special Construction</i>
6	<i>Wood and Plastics</i>	14	<i>Conveying Systems</i>
7	<i>Thermal and Moisture Protection</i>	15	<i>Mechanical</i>
8	<i>Doors and Windows</i>	16	<i>Electrical</i>

Tabela 4 – Divisões do MasterFormat

Segundo Amorim (2003), o sistema possui alguns problemas em sua estrutura de classificação, pois foi construída através de uma estrutura hierarquizada de acordo com a lógica dos processos da construção, mas de forma empírica, sem um padrão lógico bem definido. Um mesmo produto ou atividade pode ser classificado em mais de uma divisão. O sistema possui também uma listagem de palavras chave em ordem alfabética de produtos e atividades.

5.2.2 OCCS

O *OmniClass™ Construction Classification System* (OCCS) é um novo sistema que está sendo desenvolvido pelo *Construction Specifications Institute* (CSI), pela *International Alliance for Interoperability* (IAI) e mais 45 entidades. O sistema é o resultado de um grande esforço de padronização da indústria da construção que abrange todo o processo de construção, da concepção à demolição. O OCCS é focado nos processos da indústria norte-americana e em sua versão preliminar (2001) possui 12 tabelas de classificação conforme a tabela 5. A versão 1.0 tem lançamento previsto para o 2º trimestre de 2004 e conterá 15 tabelas.

Nº	Tabela	Nº	Tabela
1	<i>Facilities</i>	7	<i>Process Phases</i>
2	<i>Constructed Entities</i>	8	<i>Process Services</i>
3	<i>Spaces</i>	9	<i>Process Participants</i>
4	<i>Elements</i>	10	<i>Process Aids</i>
5	<i>Work Results</i>	11	<i>Process Information</i>
6	<i>Products</i>	12	<i>Attributes</i>

Tabela 5 – Tabelas do OCCS

O OCCS é um padrão aberto que pretende ser compatível com outras iniciativas internacionais. O sistema visa ser usado para facilitar o armazenamento de informações e a documentação de todo o ciclo da construção, além de possibilitar a definição de uma terminologia para sistemas computacionais. O objetivo do projeto é o de assegurar meios mais eficientes para o armazenamento e distribuição de informação e o de permitir comunicação simultânea mais eficaz em todas as etapas da construção através de um sistema de classificação hierárquico e um vocabulário controlado (*tesauro*).

5.3 Sistemas de classificação no Brasil

No Brasil, um dos principais sistemas de classificação para a construção é aquele sendo desenvolvido no projeto CDCON (AMORIM et al., 2002) da ANTAC (Associação Nacional de Tecnologia do Ambiente Construído) que desenvolve a Terminologia e Codificação de Materiais e Serviços para Construção. Outro esforço importante são as Diretrizes Gerais para Intercambialidade de Projetos em CADD da AsBEA (Associação Brasileira de Escritórios de Arquitetura) (CAMBIAGHI et al., 2000).

5.3.1 CDCON

O CDCON - Terminologia e Codificação de Materiais e Serviços para Construção é o maior projeto para classificação de informações da indústria da construção civil brasileira. Com o CDCON pretende-se criar uma codificação de materiais e serviços para construção e um vocabulário controlado através da classificação, categorização, indexação, associação e relação lógica entre os termos utilizados nos processos, serviços e materiais utilizados na indústria da construção. O projeto apresenta um padrão com termos da língua

portuguesa falada no Brasil. Leva em conta a cultura e técnicas da indústria da construção civil brasileira, que possui vários aspectos diferentes nas várias regiões do território nacional, e são distintos daqueles praticados em outros lugares do mundo. O projeto CDCON atualmente propõe uma classificação baseada em 7 facetas (ANTAC, 2003) conforme a tabela 6:

Faceta	Descrição	Exemplos
Função	Tipos de Edificações que podem ser definidas por seus usos específicos.	Central telefônica, usina termelétrica, estação elevatória, estação ferroviária, autódromo, ginásio esportivo e hotel.
Espaços	Partes de uma edificação com uso específico, delimitado não somente por limites físicos, bem como por fronteiras abstratas.	Ático, ala, compartimento, subsolo, apartamento, quarto, sala, quintal, corredor, saguão, escada, rampa, casa de máquinas e casa de bombas.
Elementos	Parte física de um sistema ou de uma entidade construída, a qual sozinha ou combinada com outros elementos preenche uma função.	Sapata, estaca, pilar, viga, cobertura de duas águas, parede, janela basculante, porta corta fogo, escada caracol, chapisco, chaminé e piscina.
Agentes	Intervenientes (equipamentos, pessoas e entidades) no processo da construção.	Arquiteto, engenheiro de instalações, desenhista cadista, consultor, alicate, bate-estaca, consórcio e construtora.
Produtos	Tipos de produtos utilizados na construção.	Telhas cerâmicas, resinas epóxicas, tubo de cobre, tanque, pregos, asfalto e tinta anticorrosiva.
Informação	Tipos de informação e documentos utilizados na construção. Inclui os atributos de processos, produtos, espaços, elementos e agentes da construção.	Legislação municipal, normas, estudo de viabilidade, cronograma físico, ordem de serviço, requisição de material, anteprojeto de arquitetura, contrato, dimensão nominal e textura.
Processos	Processos ao longo de todo o ciclo de vida do espaço construído, desde sua concepção até a demolição ou reciclagem.	Concepção, analisar necessidade, estudar alternativas, revisão, planejamento da produção, conferir as dimensões do terreno e pintura.

Tabela 6 – Facetas do CDCON

A vantagem da adoção da classificação facetada é permitir maior nível de detalhes da classificação de produtos e serviços. As facetas

e subclasses da classificação são definidas por conceitos que são relacionados através de uma estrutura hierárquica. Nas subclasses das facetas há uma relação partitiva (entidade inteira e suas partes).

O projeto do CDCON é dividido em dois objetivos principais: desenvolver um sistema de classificação de informações para a indústria da construção civil e criar uma codificação padronizada de materiais e serviços para o setor. Pretende-se criar com o projeto uma base para a interoperabilidade no setor da construção, criação de catálogos eletrônicos, adequar as diferenças de terminologia utilizadas em diversas regiões, criar sistemas de cotação e compras pela Internet e propiciar uma referência para implantar programas de gestão do conhecimento (ANTAC, 2002). Os objetivos específicos do CDCON são (ANTAC, 2003):

- proposta de projeto, ou conjunto de projetos, de Norma Brasileira de terminologia para materiais e serviços para construção, segmento edificações;
- proposta de projeto de Norma Brasileira para codificação de materiais e serviços para construção, segmento edificações;
- *software* para consulta das regras e terminologia para materiais e serviços para construção, a ser disponibilizado em meio físico e através de site na rede pública de computadores Internet;
- suporte para complementação das regras de definição de códigos de barras e similares para compatibilização com o sistema de codificação e terminologia propostas;
- disseminação da proposta de codificação, através de seminários, páginas na Internet e distribuição de *software* específico.

5.3.2 Diretrizes Gerais para Intercambialidade de Projetos em CADD

As Diretrizes Gerais para Intercambialidade de Projetos em CADD da Associação Brasileira de Escritórios de Arquitetura (AsBEA), desenvolvida no Brasil desde 1999, apresenta uma proposta para padronização de nomes de diretórios e arquivos CAD e seus *layers*. O objetivo do projeto é criar uma linguagem única através da padronização de bases de dados e sistemas de nomenclaturas para integrar projetistas, construtoras e clientes, e rastrear todo tipo de informação relacionada a projetos. Na classificação da AsBEA, a nomenclatura para nomes de arquivos é baseada nas disciplinas e tipo de desenho (tabela 7). Caso os participantes de um empreendimento sigam estas recomendações, é possível a localização mais imediata de informações contidas em arquivos CAD.

Disciplinas		Desenhos	
TO	Topografia	LV	Condições existentes, levantamentos
AR	Arquitetura	DE	Demolições
ES	Estrutura	IM	Implantações
EL	Instalações elétricas	PL	Plantas baixas
HI	Instalações hidráulicas	FO	Plantas de forro
IN	Combate ao fogo	EL	Elevações
AC	Instalações de ar-condicionado	EI	Elevações internas
TE	Telecomunicações e dados	CB	Coberturas
PA	Paisagismo	CT	Cortes
AI	Interiores e decoração	DH	Detalhes horizontais
CO	Instalações de cozinha	DV	Detalhes verticais
LU	Luminotécnica	DT	Detalhes gerais
AU	Acústica	EQ	Equipamentos
VD	Vedações	LE	Acabamentos/legendas
O*	Outras consultorias	3D	Desenhos 3d/ isométricos
		DG	Diagramas
		AM	Ampliações específicas
		...	...
		FU	Fundações
		FR	Formas
		...	...
		LU	Iluminação
		QD	Quadros
		...	...
		AF	Água fria
		AQ	Água quente
		...	...
		HD	Hidrantes
		DU	Dutos e bocas de ar
		TL	Telefonia

Tabela 7 – Abreviação para nomenclatura de nomes de arquivos (AsBEA)

5.4 Outros tipos de classificações

Além dos sistemas de classificação existem também os modelos de projeto baseados em esquemas computacionais padronizados como o *Industry Foundation Classes* (IFC) e o *aecXML* desenvolvidos pela IAI – *International Alliance for Interoperability*. O objetivo desses sistemas é criar padrões para a troca de informações que sejam implementados comercialmente em diversos sistemas computacionais possibilitando a interoperabilidade entre diferentes programas para AEC. A IAI desenvolve seus projetos com a colaboração de várias organizações e entidades ao redor do mundo

que representam mais de 600 membros em 24 países. O IFC e o aecXML podem ser integrados aos padrões da ISO 10303 (STEP). O IFC é mais apropriado para intercâmbio de projetos e operações da indústria AEC enquanto que o aecXML é mais apropriado para intercâmbio de documentos em comércio eletrônico (ordens de compra, contratos, dados de contabilidade e dados gerais das empresas do setor AEC).

5.4.1 IFC

O IFC (*Industry Foundation Classes*) é um modelo de dados aberto e não proprietário para ser utilizado pela comunidade AEC. O objetivo principal do IFC é de padronizar as classes utilizadas em sistemas orientados a objetos baseados em modelos do projeto para que diferentes aplicativos (inclusive com diferentes funções como CAD, orçamentação, planejamento, programação, cálculo estrutural, conforto térmico, avaliação de consumo de energia, entre outros) possam compartilhar os mesmos dados. O IFC (2000) é desenvolvido pela IAI (*International Alliance for Interoperability*), entidade formada por diversas companhias (principalmente de software e de AEC). Inicialmente, o IAI foi formado por companhias parceiras da empresa Autodesk em 1995 e significava Industry Alliance for Interoperability. A entidade passou a ser International Alliance for Interoperability em 1996 e se abriu para novas companhias. A primeira versão do IFC surgiu em 1997 (IFC 1.0) e sua última versão foi desenvolvida em 2000 (IFC 2X). Com o IFC, a intenção do IAI é a de especificar como são os elementos de uma construção (reais como portas, paredes, janelas, etc. ou abstratos como espaço, organização, processo, etc.) e como devem ser representados eletronicamente. Estas especificações são chamadas de classes e representam uma estrutura de dados que apóiam um modelo eletrônico de projeto compartilhando dados através de

aplicações (IAI, 2002). O IFC é classificado através de camadas conforme a tabela 8.

Layers	Schemas
<i>Resource Layer</i>	<i>IfcActorResource</i>
	<i>IfcDateTimeResource</i>
	<i>IfcGeometricModelResource</i>
	<i>... (mais 23 schemas)</i>
<i>Core Layer</i>	<i>IfcKernel</i>
	<i>IfcControlExtension</i>
	<i>IfcProcessExtension</i>
	<i>IfcProductExtension</i>
<i>Interoperability Layer</i>	<i>IfcSharedBldgElements</i>
	<i>IfcSharedBldgServiceElements</i>
	<i>IfcSharedComponentElements</i>
	<i>IfcSharedFacilitiesElements</i>
	<i>IfcSharedMgmtElements</i>
<i>Domain Layer</i>	<i>IfcArchitectureDomain</i>
	<i>IfcBuildingControlsDomain</i>
	<i>IfcConstructionMgmtDomain</i>
	<i>IfcElectricalDomain</i>
	<i>IfcFacilitiesMgmtDomain</i>
	<i>IfcHvacDomain</i>
	<i>IfcPlumbingFireProtectionDomain</i>
	<i>IfcStructuralElementsDomain</i>
	<i>IfcStructuralAnalysisDomain</i>

Tabela 8 – Classificação do padrão IFC (IAI, 2000)

5.4.2 aecXML

O aecXML (IAI, 2003) foi proposto inicialmente em 1999 pela Bentley Systems como uma especificação (padronização para os dados não gráficos envolvidos na indústria da construção) para revisão, comentários e formação de um grupo independente (aecXML Working Group) para criação de um padrão neutro que suporte serviços de todos os segmentos da comunidade AEC (HARROD, 2003). Atualmente o projeto está sendo desenvolvido sob orientação do IAI em acordo com o IFC e o ISO STEP, e objetiva criar padrões para trocas de dados específicos da AEC, envolvendo especialmente atividades como comércio eletrônico e

troca de dados entre bancos de dados que possuem estruturas diferentes. Com isso, através da adoção de sistemas padronizados, torna-se possível a interoperabilidade, a integração dos sistemas de comunicação e a criação de ferramentas para negócios eletrônicos (extranets de projeto, e-commerce, business to business, e-procurement, etc.) para o setor. O aecXML em sua versão 1.0 é classificado através de 12 tipos de elementos (tabela 9).

Tipo de elementos	Tags
<i>Project</i>	<i>Name, Description, Owner, Participant, CurrentPhase</i>
<i>DocumentSet</i>	<i>DocumentSetId, Description, DocumentType</i>
<i>BuildingComponent</i>	<i>Name, Description, EstimationData, MaintenanceData</i>
<i>Communication</i>	<i>CommunicationId</i>
<i>ProfessionalService</i>	<i>Participant, Expertise, AdditionalInformation, Project</i>
<i>Organization</i>	<i>Description, Participant OrganizationType</i>
<i>Software</i>	<i>Vendor, Product, Version, Description</i>
<i>Estimate</i>	<i>EstimateId, Cost, UnitCost</i>
<i>Schedule</i>	<i>ScheduleId, Task, Equipment, Resource</i>
<i>Maintenance</i>	<i>MaintenanceId</i>
<i>Supplier</i>	<i>Mixed</i>
<i>Manufacturer</i>	<i>Mixed</i>

Tabela 9 – Classificação do aecXML (versão 1.0)

O aecXML fornece à comunidade AEC uma especificação que determina regras para a criação de documentos genéricos para a indústria da construção. Esta especificação utiliza a linguagem XML (*eXtensible Markup Language*) que é uma meta linguagem de marcação derivada do SGML (*Standard Generalized Markup Language*) assim como o HTML (*Hypertext Markup Language*). O XML é uma linguagem de marcação criada em 1996, pelo comitê W3C (*World Wide Web Consortium*), formada por regras para definir *tags* (etiquetas) semânticas que dividem e identificam um documento em partes.

5.4.3 Considerações sobre sistemas de classificações para a construção civil

Existem muitos sistemas de classificações para a indústria da construção civil ao redor do mundo que estão em desenvolvimento. Este trabalho tem se mostrado complexo e demorado, pois os sistemas existentes chegam a tomar mais de dez anos de desenvolvimento, além das atualizações necessárias (MICALI, 2000). A utilização de sistemas de classificações de informações da indústria da construção civil pode contribuir com o aumento da integração do macro setor e elevar seus ganhos de produtividade (AMORIM, et al.; 2002). Segundo os mesmos autores, tem sido uma constante a busca pela compatibilidade entre os diversos sistemas utilizados no setor e que uma terminologia na língua portuguesa (já que a maioria é na língua inglesa) teria um impacto mais significativo no Brasil. Isto porque, além das diferenças dos significados dos termos com outras línguas, existem as diferenças da organização e da cultura da indústria brasileira com a de outros países. Por este motivo neste trabalho foi adotada a classificação em língua portuguesa do CDCON. O tesauro do CDCON tem como objetivos a indexação e a recuperação de documentos em sistemas de informação (ANTAC, 2002). Além disso, como neste trabalho o tesauro pretende ser utilizado em uma ferramenta de recuperação de informação para extranet de projeto (detalhada no próximo capítulo) onde várias empresas utilizam o sistema, um tesauro padronizado pode contribuir com a gestão do conhecimento do empreendimento, sendo assim mais útil do que a construção ou utilização de um tesauro baseado em indexação automática como o Extractor (KOSOVAC; VANIER; FROESE, 2000). Conforme os mesmos autores, um tesauro pode ser usado para mapear a terminologia utilizada em diferentes aplicações para a indústria da

construção civil e para definir termos para cada conceito sem ambigüidade. Neste trabalho será utilizado para resolver problemas de ambigüidades através do processo de busca contextual apresentado no capítulo 6.

6 PROPOSTA DE SISTEMA DE RECUPERAÇÃO DE INFORMAÇÃO

6.1 Visão geral do sistema RIcon

Este trabalho apresenta o projeto de desenvolvimento de uma ferramenta chamada RIcon (Recuperação de Informações da Construção) que foi projetada para ser integrada a um sistema de recuperação de informação de uma extranet de projeto da indústria da construção civil.

É proposta uma metodologia para recuperação de informações com heurística baseada em um sistema de classificação das informações da construção civil. Estas informações abrangem todos os processos do ciclo de vida de uma construção com classificações relacionadas aos processos, fases, materiais, produtos e espaços. O objetivo desta ferramenta é melhorar o resultado dos mecanismos de busca convencionais das extranets de projeto. Para isso são propostas a expansão dos argumentos de busca e a ordenação da apresentação dos resultados por ordem de relevância dos documentos recuperados (*ranking*) através da adoção de pesos. Estes documentos serão apresentados através de uma lista com nomes e descrições de documentos que apontam para o documento original (arquivo eletrônico) no banco de dados do sistema. A expansão dos argumentos de busca será por meio de operações de expansão de termos por relacionamento de equivalência e por relacionamento hierárquico. Estas operações do mecanismo de busca levam em conta as especificidades da indústria da construção civil o que tornando estes mecanismos mais eficazes do que aqueles obtidos com técnicas convencionais genéricas. Enquanto os sistemas de busca tradicionais recuperam todos os documentos que contêm um ou mais termos

especificados, no RIcon são recuperados documentos que apresentam variados graus de similaridade com a consulta. A base de conhecimento da ferramenta utiliza representação do conhecimento baseada em *frames*, conforme mencionado no capítulo anterior. Pretende-se localizar documentos em que não necessariamente constam as palavras-chave utilizadas nos argumentos de busca. Com isso os agentes dos processos da indústria da construção civil poderão localizar rapidamente informações relevantes e oportunas e melhorar seus processos de tomada de decisões além de reduzir a sobrecarga de informação geradas nas extranets de projeto.

6.2 Operação da ferramenta

O funcionamento do sistema consiste em processar uma consulta em um mecanismo de busca para que um usuário obtenha uma resposta à sua necessidade de informação, obtendo documentos armazenados na extranet de projeto. A consulta é feita por meio de um formulário onde o usuário pode usar linguagem natural ou palavras-chave. A busca não é feita no conteúdo dos arquivos no banco de dados do sistema mas sim nos descritores dos documentos. Os descritores são informações sobre o conteúdo dos arquivos de documentos, introduzidas através de formulários da extranet de projeto no momento em que os usuários transferem os arquivos para o sistema.

A eficácia do sistema de recuperação de informação depende em grande medida da qualidade dos argumentos de busca especificados pelo usuário. Idealmente devem ser específicos e em boa quantidade. O pequeno número de argumentos de busca digitados pelo usuário talvez seja um dos maiores problemas para obter-se uma maior eficácia do sistema. De acordo com Andreassen

(2003) a maioria dos usuários utilizam poucos argumentos em sistemas de busca, pois a falta de boas informações recuperadas fez com que os usuários adquirissem o hábito de serem minimalistas. O número médio de argumentos utilizados em um mecanismo de busca na Internet, em pesquisa de 2003, por exemplo, era menor que dois (ANDREASEN, 2003). Outra pesquisa recente mostra que o número de argumentos de busca aumentou nos últimos anos (ONESTAT, 2004) e que atualmente a maior porcentagem das buscas são feitas utilizando-se dois (33%), três (26%) ou apenas um argumento (19%). Os usuários muitas vezes não conhecem a maneira e os recursos como podem especificar suas necessidades de informação e não possuem tempo ou vontade de aprender.

Os termos das consultas serão processados e expandidos. Após a recuperação, os documentos são ordenados e apresentados por ordem de relevância (*ranking*).

6.3 Pré-processamento

Os termos dos argumentos das consultas precisam ser preparados antes de serem utilizados pelos algoritmos da ferramenta de busca. Os primeiros passos deste pré-processamento são as operações sobre texto. Inicialmente, o texto da consulta é carregado para a ferramenta e separado em termos para ser montada uma lista. Em seguida, os caracteres retirados dos argumentos da consulta sofrem os seguintes processos:

- Análise Léxica – Converte uma cadeia de caracteres em uma cadeia de termos. Nesta operação as marcas de pontuação, acentos, hífen, espaços e as variações de caixas de letras são eliminados;

- Eliminação de *Stopwords* – as palavras consideradas irrelevantes são desconsideradas (preposições, artigos, conjunções, etc);
- Lematização (*Stemming*) – Redução de palavras ao seu radical com eliminação de sufixos (PORTER, 1997);
- Criação de lista de termos – Criação de uma lista com os termos que resultaram das operações acima.

6.4 Expansão de Termos

Após as operações de pré-processamento, a lista de termos da consulta é ampliada utilizando a técnica de Expansão de Termos (ZAZO, et al. 2002). Com essa técnica, a informação consultada pode ser expressa por diferentes palavras. A expansão é feita através do acréscimo de termos nos argumentos da consulta por intermédio de um tesouro (sinônimos e palavras derivadas) da construção civil. A expansão dos termos de uma consulta é feita através da combinação (*matching*) com os *frames* da base de conhecimento. A base de conhecimento foi criada a partir das informações do tesouro do CDCON. Os termos são pesquisados na estrutura de *frames* para encontrar *frames* cujos nomes coincidam com eles. O objetivo da expansão de termos é a de melhorar a revocação do sistema. Neste trabalho são executados dois tipos de operações de expansão de termos: a Expansão de Termos com Relacionamento Equivalente e a Expansão de Termos com Relacionamento Hierárquico.

6.4.1 Expansão com Relacionamento de Equivalência

A expansão baseada em relacionamento de equivalência consiste, conforme a figura 13, na obtenção dos termos indicados pelos conteúdos dos *slots* "*s_sinônimos*" dos *frames* cujos nomes

coincidem com os termos de consulta do usuário. Os termos obtidos pelo processo de expansão terão peso menor do que os termos originais de consulta. Sinônimos de sinônimos também podem ser considerados para ampliação dos termos, embora apresentem pesos menores. No exemplo da figura 13, o termo “Fundação” é expandido pelos sinônimos “Alicerce”, “Base” e “Sustentáculo”. A presença de quaisquer destes termos na descrição de um documento da extranet indicará que o documento é potencialmente relevante para a consulta feita, ainda que em diferentes graus.

Figura 13: Expansão de Termos por Relacionamento de Equivalência

6.4.2 Expansão com Relacionamento Hierárquico

A expansão de termos baseada em relacionamento hierárquico (figura 14) considera a relação entre *frames*. Assim, um termo é expandido pela adição de outro(s) mais genérico(s) e/ou mais específico(s). No exemplo da figura 14, o termo fundação é expandido com o termo mais genérico “Estrutura de Concreto” e com os termos mais específicos “Bloco”, “Sapata” e “Radier”.

Figura 14: Expansão de Termos por Relacionamento Hierárquico

Os pesos dos termos da superclasse e subclasse são sempre menores do que o do *frame* de origem, preservando a preferência do usuário na ordenação por relevância.

6.5 Busca contextual

Alguns termos utilizados na construção civil são ambíguos. Este fato prejudica a precisão do resultado de uma busca por informações, podendo retornar documentos irrelevantes ou não recuperar documentos interessantes para o usuário. Uma solução possível para este problema é contextualizar a consulta. Um exemplo onde é necessário contextualizar, conforme ANTAC (2002), seria o caso de um usuário digitar uma palavra como “bloco” em uma consulta. O sistema retornaria documentos relacionados ao termo “bloco” nos vários contextos, como um “tipo de fundação”, um “elemento de vedação” e um “conjunto de edifícios”. Para auxiliar o processo de busca quando os termos

especificados em uma consulta forem ambíguos é utilizada uma técnica de busca contextual. Na proposta deste trabalho, o termo é inicialmente relacionado com todos os seus significados (são utilizadas palavras compostas, já que na base de conhecimento não existem *frames* com nomes iguais e significados diferentes), pois conforme o projeto CDCON o termo “bloco” se encontra em diversas definições:

- **Bloco** – sistema de fundação rasa através de elementos estruturais que distribuem a carga conforme a capacidade de sustentação do solo;
- **Bloco** de edifícios – cada um dos edifícios que formam um conjunto de prédios;
- **Bloco** de vedação – tipo de elemento constituinte da vedação de uma construção.

Segundo Kent (1978) apud Loh; Wives; Frainer (1997), o problema da ambigüidade das palavras é resolvida pelos humanos através do entendimento do contexto. Neste sentido, o problema de contextualização no sistema RICon é resolvido através de um *feedback* ao usuário para que ele defina o contexto de sua necessidade de informação já que normalmente, devido à concisão das consultas, é impossível ao sistema deduzir a intenção do usuário. O sistema retornará, em caso de ambigüidades, uma pergunta para o usuário (através de caixa de diálogo) para que ele escolha o contexto da consulta. Para isso, a ferramenta de busca pesquisará, em toda a base de conhecimento, nomes de *frames* que possuam termos iguais aos dos termos das buscas. Encontrando mais de um *frame* com o termo, o sistema indagará o usuário a qual contexto ele se refere, mostrando-lhe o termo mais genérico correspondente na estrutura hierárquica de *frames*, como no exemplo de bloco de alvenaria (figura 15).

Figura 15: Estrutura hierárquica do frame bloco de alvenaria

Assim, no exemplo da palavra “bloco” as perguntas originadas de generalização para identificar qual o contexto seria:

- **Bloco** (Fundação)?
- **Bloco** (Alvenaria)?
- **Bloco** (Elemento)?

Se o usuário escolher, por exemplo, o contexto para bloco de alvenaria, o sistema pesquisará documentos que contenham além do termo especificado (bloco), um ou mais termos obtidos pelo processo de expansão de termos (equivalente e hierárquica) do frame do contexto escolhido.

6.6 Exemplo de uma busca no sistema RIcon

O sistema RIcon é iniciado quando um usuário faz uma consulta no sistema. A consulta é feita através do preenchimento de sua necessidade de informação em um formulário de consulta. O formulário pode ser preenchido com palavras-chaves ou linguagem natural. Neste exemplo, o usuário especificou sua necessidade de informação em linguagem natural:

“Qual o fck da fundação?”

O sistema, para retornar uma resposta, primeiro faz um pré-processamento da consulta iniciando pelo processo de Análise Léxica dos termos da frase da consulta onde é eliminado os acentos, a pontuação e é colocado todo o texto em caixa baixa. No exemplo são eliminados a cedilha, o acento til e o ponto de interrogação, além de ser convertido todo o texto para minúsculas. Os termos que seguem no processo são:

- qual
- o
- fck
- da
- fundacao

Em seguida, são eliminadas a palavra com baixo potencial para influenciar o resultado da busca (*stopwords*), que no caso são as palavras "qual", "o" e "da". Então os termos que seguem no processo são:

- fck
- fundacao

Na última atividade do pré-processamento, é feito o processo de lematização onde os termos são reduzidos à sua raiz. No caso é eliminado o sufixo "acao" da palavra fundacao. Assim, os termos remanescentes deste processo formam então a Lista de Termos:

- fck
- fundac

Na próxima etapa ocorre o processo de pós-processamento iniciando-se a busca contextual quando houver termos homógrafos na base de conhecimento e em seguida é iniciado o processo de expansão de termos com a expansão por equivalência e por hierarquia. Na expansão de termos por equivalência são agregados os termos capturados nos *frames* da base de conhecimento que

são sinônimos. Em relação ao termo fundac (fundação) são encontrados de acordo com a figura 13:

- alicerce;
- base;
- sustentáculo.

Na expansão hierárquica são agregados os termos através da hierarquia de *frames* da base de conhecimento (um nível acima e um nível abaixo) que, no caso, são de acordo com a figura 14:

- estrutura de concreto (termo mais genérico);
- bloco;
- sapata;
- radier.

A princípio serão estimados pesos (ω_i) para cada termo da busca, de acordo com sua origem (termo original, sinônimo, termos mais genéricos, termos mais específicos). Assim, será atribuído peso 1.0 para os termos originados do pré-processamento; para os termos obtidos por expansão são dados pesos menores, sendo 0.6 para os termos expandidos por equivalência e 0.5 para os termos expandidos por hierarquia. Estes pesos foram inicialmente atribuídos de forma estimada, mas seus valores poderão ser ajustados após testes com o sistema. A lista de termos final é composta pelos 2 termos originais (fck, fundac) mais os 3 sinônimos encontrados (alicerce, base, sustentáculo), mais 1 termo genérico (estrutura de concreto) e 3 específicos (bloco, sapata e radier), totalizando 9 termos de busca. Assim, no exemplo, são executadas 9 operações individuais de busca no sistema. Os documentos apresentados no resultado das buscas serão ordenados por relevância de acordo com suas notas (ver tabela 10) em uma lista de documentos. Se o resultado apresentar mais de 10 documentos relevantes, serão ordenados os dez primeiros

documentos em uma página e, caso o usuário deseje, serão apresentados os outros documentos provavelmente menos relevantes de dez em dez.

No exemplo do resultado da busca acima foram ordenados os documentos de acordo com a tabela 10. A tabela foi construída através de uma matriz de termos x arquivos com o valor relativo do peso pd_i de cada termo e o a nota η_{ij} calculada para cada arquivo de documento j em relação ao termo i (1 em caso do documento j conter o termo i , e 0 caso contrário)

Arquivos	Fck	fund	alicerce	base	Susten- táculo	Estrutura de concreto	bloco	sapata	radier	Nota Total (nt_j)
Peso (ω_i)	1.0	0,6	0.6	0.6	0.6	0.5	0.5	0.5	0.5	
1.doc	0	0	0	0	0	1	0	0	0	0.5
2.dxf	0	0	0	1	0	0	0	0	0	0.6
3.docxls	1	1	0	1	0	1	0	1	0	3.6
4.doc	0	0	0	0	0	0	0	0	0	0
...										
n.txt	0	0	0	1	0	0	1	0	0	1.1

Tabela 10 – Tabela de documentos x termos com pesos

A nota total de cada documento nt_j é calculada através da função de ponderação da importância relativa do termo i em um documento j pela fórmula da equação 3.

$$nt_j = \sum_{i=1}^n pd_i \times \eta_{ij} \quad (3)$$

O peso de um termo depende da frequência com que ele aparece nos documentos da base. Termos que aparecem em muitos documentos não são tão bons discriminadores quanto os que aparecem apenas em alguns documentos (FUSARO, 1988). Assim, o peso final (pd_i) de um termo i é calculado pela medida do peso em função de sua origem (ω_i) e por uma função proporcional ao inverso do número de documentos que possuem o termo (nd_i), como no exemplo da eq. (4) (SALTON; BUCKLEY, 1988).

$$pd_i = \omega_i \times \left(1 + \frac{1}{nd_i} \right) \quad (4)$$

No caso particular da base de documentos disponíveis a esta pesquisa, cujos documentos (na verdade, descritores de documentos) são extremamente curtos, consistindo de apenas algumas linhas de texto, a ponderação pela frequência inversa de termos (eq. 4) não produz o efeito desejado, conforme verificaram Salton e Buckley (1987) em seus experimentos com a base NPL (National Physical Laboratory) de características semelhantes à usada aqui. Esta base contém, em média, apenas 19.96 termos por documento. O esquema de melhor desempenho para buscas nesta base foi aquele que leva em conta a frequência dos termos de busca no documento, mas que despreza a frequência dos termos na base. Por essa razão, também desprezou-se este último componente. Por outro lado, considerando que no caso da base utilizada nesta pesquisa, praticamente nunca haverá palavras repetidas da descrição do documento, foi adotada a frequência de termos binária, sem prejuízo no desempenho do sistema.

De acordo com a tabela 10 a apresentação dos documentos para o usuário seria ordenada conforme a figura 16.

Figura 16: Simulação do resultado de uma busca típica do sistema RICON

Assim, de acordo com a ordenação dos arquivos, o documento mais provável da base de dados da extranet de projeto que contém a informação consultada (Fck da fundação?) é o documento 3.doc com nota 3,6. Todos os processos do sistema RICON são representados na figura 17:

Figura 17: Fluxograma de processamento de uma busca no sistema RICON

7 AVALIAÇÃO DO SISTEMA

Neste trabalho foram executadas quinze consultas em uma extranet de projeto para avaliar o sistema proposto. As medidas avaliadas foram as de precisão e revocação do sistema RIcon comparadas com as mesmas medidas do mecanismo de busca existente na extranet de projeto selecionada. Para isso foi feito um estudo de caso em um empreendimento de uma construtora que utiliza a extranet de projeto através de consultas no sistema. Os dados para avaliação foram obtidos através de buscas elaboradas pelo coordenador de projeto do empreendimento, do desenvolvedor da extranet de projeto e pelo autor deste trabalho.

Até a etapa de avaliação, o sistema RIcon não havia sido implementado totalmente. Por isso, as buscas no sistema RIcon foram feitas manualmente (através de cálculos para ordenação e consultas no sistema existente, simulando o processo de busca do sistema proposto). No sistema RIcon foram elaboradas consultas com mais de um termo e no mecanismo existente na extranet, como só é possível recuperar frases exatas, algumas consultas foram separadas em diversas consultas individuais (Ex.: A consulta sobre o tipo de Fundação do Edifício Maracaju no mecanismo existente na extranet foi feita em duas consultas, uma por “Fundação” e outra por “Maracaju”).

As quinze consultas foram igualmente divididas entre o coordenador do empreendimento, o desenvolvedor da extranet de projeto e o autor deste trabalho, isto é, houve 5 propostas de cada colaborador. Porém, muitas buscas não resultaram em uma quantidade satisfatória de documentos (poucos ou nenhum documento) que pudesse avaliar os sistemas. Estas buscas não foram analisadas, sendo descartadas.

7.1 Escolha da extranet de projeto

A extranet de projeto escolhida para fazer os testes do sistema RIcon foi o SADP (www.sistrut.com.br/sadp) da empresa Sistrut Software e Tecnologia Ltda. A empresa possui vinte e dois anos no mercado de projetos para a construção civil e a extranet está funcionando há cinco anos, sendo uma das mais antigas e mais utilizadas extranets de projeto brasileiras. A empresa criadora já era conhecida no mercado da construção civil por produzir um programa de cálculo de estruturas de concreto armado. Esta extranet de projeto já serviu para estudo em outros trabalhos acadêmicos no Brasil como (PICORAL e SOLANO, 2001). Outras importantes características da extranet que determinaram sua escolha foram o fato de ser 100% nacional (o que facilitou a negociação de seu uso para este estudo) e o fato de possuir um recurso de identificação dos documentos por comentários, já que a maioria dos clientes de provedores de extranets não autoriza a visualização do conteúdo dos documentos de projeto por questões de segurança e privacidade.

7.1.1 Características da extranet de projeto do experimento

O sistema de recuperação de informação da extranet de projeto escolhida (SADP) para avaliar o sistema RIcon possui alguns recursos que permite ao usuário:

- selecionar um projeto dentre os que tiver cadastrado no sistema;
- selecionar os níveis de habilitação do sistema e a pasta que deseja listar;
- formatar a listagem dos resultados da busca (figura 18).

The screenshot shows a configuration window for the extranet. At the top, there is a dropdown menu labeled 'Projeto' with the text 'Selecione um Projeto'. Below this, there are two list boxes. The left one is titled 'Níveis que deseja listar' and contains the following items: 'Coordenação', 'Todos que está habilitado', 'Agente Financeiro', 'Alvenaria Estrutural', and 'Alvenaria/Vedação de Fechamento'. The right one is titled 'Pasta do Projeto que deseja listar' and contains: 'Todas', '3D', 'Alterações de Obra', 'AnteProjeto', and 'Atas de Reunião'. Below each list box is a note: '(Para selecionar vários acione CTRL+mouse-click)'. At the bottom, there is a dropdown menu labeled 'Formatar a Listagem pelo Padrão' with the text 'Padrão SADP'.

Figura 18: Opções de configuração da extranet do experimento

O sistema possui ainda algumas opções de pesquisa (figura 19) onde é possível mostrar arquivos bloqueados, aprovados e apagados. Possui uma opção de ordem de apresentação dos documentos onde a opção padrão é a de data de cadastramento dos documentos. É possível ainda ordenar pela ordem invertida de cadastramento e por ordem alfabética. As opções de busca permitem que sejam procurados documentos pelo texto no nome do arquivo ou pelo conteúdo dos comentários do arquivo em seus descritores. O mecanismo de busca não aceita operações booleanas e por isso só pode ser utilizada mais de uma palavra quando estiver pesquisando a frase exata nos documentos (ex: "estrutura de concreto" não localiza documentos que possuem as palavras estrutura e concreto isoladamente).

Para realizar a avaliação dos sistemas de recuperação de informação foi escolhido um projeto (item 7.2) utilizando as opções de listar todos os arquivos (usando nível e coordenador para "Todos que está habilitado") em todas as pastas. As opções de pesquisa foram para apresentação por ordem de cadastramento, todos os arquivos (bloqueados, aprovados e apagados) e pelo texto dos comentários dos descritores. Os nomes dos arquivos no empreendimento escolhido possuem códigos que não descrevem claramente os arquivos.

☒ NÃO mostrar arquivos BLOQUEADOS
☐ Mostrar SOMENTE arquivos APROVADOS

☐ Mostrar SOMENTE arquivos JÁ DELETADOS

Ordem de Apresentação
☒ Ordem de Cadastramento
☐ Ordem Invertida de Cadastramento
☐ Ordem Alfabética

Somente os Arquivos que (☐ não) tenham no
☒ Nome do Arquivo
☐ Comentários do Arquivo
 (comentário aceita data dd/mm)
 a palavra

Lista Arquivos do Projeto Selecionado

Figura 19: Opções de configuração da extranet do experimento

7.2 Coleção de documentos

A escolha da coleção de documentos para os experimentos do sistema foi feita através da seleção do empreendimento para fazer a avaliação do sistema RIcon. A partir de uma divulgação da extranet de projeto selecionada, foi enviado um e-mail para todos os coordenadores usuários da extranet convidando-os a participar do trabalho. A empresa selecionada foi a Fortenge Construções Ltda. que possui 6 empreendimentos em andamento na extranet de projeto. O empreendimento escolhido para fazer a avaliação possui as seguintes características:

- Nome do Projeto: 139-Condomínio Costa do Sol;
- Tempo de cadastramento na extranet: 3 anos;
- Quantidade de documentos armazenados: 1049 arquivos;
- Porcentagem de documentos CAD ou de plotagem: 96%
- Espaço ocupado no sistema: 282 Mbytes;
- Características gerais do empreendimento: condomínio composto por 7 edifícios de 13 pavimentos tipo mais térreo com centro de lazer coberto e portaria, totalizando 567

unidades com área total construída de 38.800,00 m² em terreno com 20.232,13 m².

A maioria dos documentos armazenados estava em formato de arquivo comprimido (às vezes com mais de um arquivo de projeto no mesmo documento comprimido), sendo que a grande maioria eram arquivos CAD e de plotagem. Nos descritores eram resumidos seus conteúdos de forma bem sucinta conforme a tabela 11.

Nº	COMENTÁRIO	DATA	NOME
1	LOCAÇÃO DE ESTACAS E BROCA - EDIF. 2 - ITACARÉ (FL.02 - REV.1) OBS.: PLOTAR COLORIDO	03/07/03	139-FUN-02-PE-R1-DWG.zip
2	LOCAÇÃO DE ESTACAS E BROCA - EDIF. 2 - ITACARÉ (FL.02 - REV.1) OBS.: PLOTAR COLORIDO	03/07/03	139-FUN-02-PE-R1-PLT.zip
3	Locação de Estacas e Broca - Edifício 1 - Itapuã OBS.: Plotar Colorido (FL.01 - Rev.03)	11/07/03	139-FUN-01-PE-R3-DWG.zip
4	Locação de Estacas e Broca - Edifício 1 - Itapuã OBS.: Plotar Colorido (FL.01 - Rev.03)	11/07/03	139-FUN-01-PE-R3-PLT.zip
5	LOCAÇÃO DE TUBULÕES - EDIF. JACUMÃ (ED.3) - EDIF. BÚZIOS (ED.6)- FL.03 - REV.1	30/07/03	139-FUN-03-PE-R1-DWG.zip
6	LOCAÇÃO DE TUBULÕES - EDIF. JACUMÃ (ED.3) - EDIF. BÚZIOS (ED.6)- FL.03 - REV.1	30/07/03	139-FUN-03-PE-R1-PLT.zip
7	LOCAÇÃO DE ESTACAS E TUBULÕES - PÓRTICOS 1 e 2, PORTARIA E ADMINISTRAÇÃO - FL.05 - PROJETO	25/08/03	139-FUN-05-PE-R0-DWG.zip
8	LOCAÇÃO DE ESTACAS E TUBULÕES - PÓRTICOS 1 e 2, PORTARIA E ADMINISTRAÇÃO - FL.05 - PROJETO	25/08/03	139-FUN-05-PE-R0-PLT.zip
9	Locação de Sapatas e Tubulões - FL.04 - Rev.2 - Edif. Maracajú (ED.4), Edif. Malembar (ED.5) e Edif. Camburí (ED.7)	20/10/03	139-FUN-04-PE-R2-DWG.zip
10	Locação de Sapatas e Tubulões - FL.04 - Rev.2 - Edif. Maracajú (ED.4), Edif. Malembar (ED.5) e Edif. Camburí (ED.7)	20/10/03	139-FUN-04-PE-R2-PLT.zip

Tabela 11. Listagem típica dos resultados do sistema do experimento.

7.3 Método de avaliação do sistema

O método de avaliação do sistema foi baseado nas técnicas utilizadas nas conferencias TREC (HARMAN, 1997). Inicialmente, as necessidades de informação definidas pelo coordenador do empreendimento, pelo desenvolvedor da extranet e pelo autor deste trabalho foram transformadas em consultas. Após as consultas serem executadas foram examinados manualmente todos os documentos recuperados em cada consulta, identificando-os como relevantes e não relevantes. Para identificar o total

aproximado de documentos relevantes armazenados na extranet de projeto (para calcular a revocação do sistema em cada consulta), foram realizadas outras consultas mais abrangentes e mais específicas, além de ser verificado todo o conteúdo da pasta da extranet relacionada ao assunto da consulta. Procurou-se identificar como documentos relevantes aqueles que claramente atendiam à necessidade de informação expressa na consulta pois, de acordo com Rijsbergen (1979), os critérios para julgamento de relevância são subjetivos e podem diferir entre diferentes indivíduos para um documento em relação a uma consulta. A partir das necessidades de informações expressas em linguagem natural, foram elaboradas consultas para avaliar e comparar o sistema RIcon e o sistema de busca nativo da extranet. As consultas que obtiveram menos de 10 documentos relevantes recuperados para cada um dos dois mecanismos de busca foram descartadas. Desta forma, conseguiu-se cinco consultas satisfatórias dentre quinze tentativas.

Para cada consulta válida, foram calculadas:

- a quantidade de documentos recuperados (TR);
- a quantidade de documentos relevantes (TC);
- a quantidade de documentos relevantes recuperados (rr);
- a quantidade de documentos irrelevantes recuperados (ri);
- a medida de precisão para os 10 primeiros documentos recuperados (P_{10});
- a medida de revocação para os 10 primeiros documentos recuperados (R_{10});
- o gráfico com a curva de Precisão x Revocação para cada consulta em cada mecanismo de busca.

A partir dos resultados de cada consulta, foi calculado para análise a precisão média para os 10 primeiros documentos, a revocação

média para os 10 primeiros documentos e foi construída uma curva de Precisão x Revocação média através da média de precisão e revocação dos 10 primeiros documentos recuperados para todas as consultas.

7.4 Resultados obtidos

Os resultados obtidos na avaliação servem para comparar a eficácia do sistema de recuperação de informações proposto (RIcon) e o sistema existente (SADP). As consultas de teste versam sobre temas variados e têm como objetivo cobrir diferentes áreas da indústria da construção civil (Vide Anexo B). Os resultados obtidos oferecem um forte indicativo da comparação da efetividade dos dois sistemas.

A figura 20 ilustra a comparação da qualidade de recuperação de informação em relação à precisão média para os 10 primeiros documentos recuperados nas consultas executadas no RIcon e no mecanismo existente na extranet de projeto (SADP). Foi calculada a média dos 10 primeiros documentos recuperados, seguindo a mesma metodologia adotada por Zazo et al. (2002), que justifica esta escolha pelo fato de que este é normalmente o número de documentos mostrados em uma tela da interface de usuário de sistemas de busca. De acordo com o gráfico da precisão média, percebe-se sempre um ganho significativo de precisão do sistema RIcon em relação ao SADP, em toda a faixa dos 10 primeiros documentos recuperados. Isto se deve ao cálculo de pesos para ordenamento do sistema RIcon que privilegia os documentos potencialmente mais relevantes, enquanto que o SADP apresenta listagem em ordem de cadastramento.

Figura 20: Precisão média para os 10 primeiros documentos recuperados.

Na figura 21 é observado a revocação média dos dois mecanismos de busca em relação aos 10 primeiros documentos recuperados nos dois mecanismos. Da mesma forma que a precisão, a revocação do sistema RIcon é sistematicamente melhor que no sistema SADP. A diferença tende a aumentar conforme a lista de documentos apresentados vai aumentando já que com o processo de expansão de termos o sistema RIcon recuperou na média mais documentos relevantes que o SADP.

Figura 21: Revocação média para os 10 primeiros documentos recuperados.

A figura 22 ilustra o desempenho dos dois mecanismos de busca através das curvas de Precisão x Revocação. Conforme os gráficos anteriores, é também evidente neste gráfico o melhor desempenho do sistema RIcon sobre o SADP e que ambos apresentam menor precisão com o aumento da revocação (Lei de Cleverdon).

Figura 22: Curva Precisão x Revocação da média dos 10 primeiros documentos recuperados.

Os dados para elaboração dos gráficos anteriores são detalhados na tabela 12 onde são mostrados os ganhos do sistema RIcon em cada nível de ordenação dos documentos recuperados. A tabela mostra ainda que o ganho médio de precisão do sistema RIcon é de 40% e o de revocação 55% em relação ao sistema existente na extranet de projeto analisando o resultado dos primeiros dez documentos apresentados na buscas.

Ord.	SADP	RIcon			SADP	RIcon		
	Precisão	Precisão	Diferença	Ganho	Revocação	Revocação	Diferença	Ganho
1	67	100	33	49%	8	16	8	100%
2	75	100	25	33%	19	32	13	68%
3	61	93	32	52%	19	35	16	84%
4	58	90	32	55%	22	38	16	73%
5	60	84	24	40%	27	41	14	52%
6	58	80	22	38%	30	44	14	47%
7	60	80	20	33%	34	47	13	38%
8	58	78	20	34%	37	47	10	27%
9	57	76	19	33%	39	49	10	26%
10	55	74	19	35%	40	53	13	33%
		Média	25	40%		Média	13	55%

Tabela 12 – Ganhos de precisão e revocação do sistema proposto em relação ao existente.

Os maiores ganhos de precisão observados nos testes ocorrem para o quarto nível de ordenação (55%) e o maior ganho de revocação está no primeiro nível de ordenação (100%). Através dos dados da tabela 12, pode-se verificar também o melhor ganho de precisão e de revocação nos primeiros níveis de ordenação. Estes fatos confirmam que a contextualização e a expansão de termos nos algoritmos de busca em sistemas de recuperação de informação melhoram os resultados de precisão e revocação do sistema.

8 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho foi descrito um sistema de recuperação de informação para a indústria da construção civil. Os principais processos que tornam o sistema específico para o setor da construção civil são os relacionados à busca contextual e à expansão de termos. A principal fonte de informações para construção dos algoritmos do mecanismo de busca foi o tesouro do CDCON. Outro processo importante do sistema RIcon é o cálculo de ordenação dos documentos recuperados para apresentação ao usuário. O sistema proposto foi avaliado através de uma extranet de projeto com documentos de um empreendimento real.

8.1 Conclusões

Este trabalho apresenta uma ferramenta (RIcon) que pretende contribuir para a solução de problemas para a falta de informações para elaboração de projetos da indústria da construção civil. A solução deste problema pode trazer benefícios para os profissionais que trabalham com projetos, pois estes necessitam de informações oportunas para tomadas de decisões durante o processo de projeto. A informação é, para estes profissionais, uma “matéria prima” essencial para suas atividades (geração de documentos, tomadas de decisão, execução de tarefas e geração de novos conhecimentos). Esta iniciativa possibilita que profissionais do setor que utilizam extranets para gerenciar as informações de projetos possam localizar rapidamente tais informações em documentos armazenados no sistema.

O sistema RIcon pretende solucionar parte do problema que os atuais mecanismos de busca das extranets possuem em relação à sua eficácia devido ao generalismo e à descontextualização dessas ferramentas, além de evitar a sobrecarga de informações a seus

usuários. Com isso, acredita-se que as extranets sejam utilizadas mais eficazmente, com menos rejeição e, portanto haja maior interesse pela tecnologia. Com os benefícios de sua maior utilização acredita-se também que o setor possa conseguir maior produtividade. As Extranets de Projeto só se tornarão um recurso realmente eficaz para o melhor gerenciamento de projetos se os usuários puderem obter informações de forma rápida e confiável.

Através de experimentos feitos com o sistema proposto e o sistema nativo da extranet foram obtidos resultados de melhoria nas duas principais medidas de avaliação de sistemas de recuperação de informação: a precisão e a revocação.

A pequena quantidade de documentos armazenados na extranet de projeto comparado com a quantidade de documentos existentes em bases de documentos tradicionais para avaliação de sistemas de recuperações de informações como o TIPSTER das conferencias TREC dificulta a realização de experimentos nestes sistemas. Outro fato que dificulta experimentos para avaliação de sistemas de recuperação de informações em extranets de projeto é a proibição de grande parte das empresas do setor em permitir a verificação dos documentos armazenados em sua base de dados. Neste trabalho foi possível verificar apenas arquivos descritores que contêm informações resumidas dos documentos originais armazenados na extranet.

Apesar destes problemas, o resultado deste trabalho indica que o sistema proposto melhora a eficácia da recuperação de informação de projeto para os usuários das extranets de projeto. Esta eficácia é traduzida quantitativamente em melhora de 40% de precisão média e 55% de revocação média do sistema de recuperação de informação proposto em relação ao sistema existente na extranet.

O cálculo dos pesos para ordenação dos documentos e o processo de busca contextual mostraram-se importantes para a melhoria da precisão. A expansão dos termos de busca através de generalização e especialização e o processo de lematização mostraram-se eficazes para melhorar a revocação do sistema. Para conseguir recuperar mais eficientemente documentos nas extranets de projeto é necessário descrever detalhadamente os documentos armazenados nos descritores, principalmente arquivos de documentos em formato não texto como os arquivos comprimidos e os arquivos de desenhos em sistemas CAD, que são a maioria nestes sistemas.

A eficácia do sistema RIcon depende da qualidade do tesouro adotado para o processo de expansão de termos. O tesouro (CDCON) utilizado neste trabalho ainda não estava terminado, por isso foram utilizadas outras fontes para desenvolvimento do sistema.

Um eficiente sistema de recuperação de informação pode apresentar melhora significativa no valor das extranets de projeto como ferramenta para contribuir com o gerenciamento do conhecimento nas empresas do setor da construção permitindo o armazenamento e o acesso às informações de projeto e o compartilhamento de conhecimento entre os agentes. Quanto maior for uma base de dados mais necessária é uma ferramenta que possibilite filtragem das informações para localizar o conhecimento desejado.

Ao longo do desenvolvimento deste trabalho chegou-se a algumas conclusões que não faziam parte do escopo principal da pesquisa, tais como:

- A padronização e utilização de classificações podem ser utilizadas na construção de ferramentas computacionais para melhoria da eficiência do desenvolvimento de projetos da indústria da construção civil;
- As extranets de projeto devem ser melhor utilizadas para trazer maiores benefícios a seus usuários. Os recursos são sub-utilizados (muitos utilizam somente o recurso de armazenamento de arquivos) e de forma inadequada (algumas vezes os documentos são armazenados por secretárias e o profissional da construção civil não tem acesso a extranet). Em extranets que utilizam descritores de documentos, os usuários devem descrever claramente estes documentos.

8.2 Trabalhos futuros

Pretende-se implementar futuramente totalmente o sistema proposto. Através do desenvolvimento deste trabalho foram iniciadas algumas pesquisas com propostas de evolução do sistema RIcon (SANTOS; NASCIMENTO, 2002), (SANTOS; NASCIMENTO, 2003) e (NASCIMENTO; SANTOS, 2003). Nestas pesquisas é proposto um sistema baseado em conhecimento com busca contextual automática e sem o *feedback* com o usuário e sem a necessidade do uso do mecanismo de busca existente da extranet de projeto.

Alguns métodos mais sofisticados de recuperação de informação que poderiam trazer alguma melhoria nos resultados das buscas não foram utilizados neste trabalho devido à sua complexidade e ao limitado tempo de desenvolvimento no contexto doo programa de mestrado da Escola Politécnica da Universidade de São Paulo.

Porém estes métodos podem ser desenvolvidos em outros trabalhos que visem a otimização dos algoritmos do sistema RIcon.

ANEXO A - DETALHAMENTO DE UMA BUSCA

Detalhamento da busca: Tipo de fundação do edifício Maracaju?

A partir da necessidade de informação (Tipo de fundação do edifício Maracaju) foram elaboradas três consultas. No sistema SADP como não é aceito consultas com mais de um termo (a não ser a palavra exata) foram feitas duas consultas com uma palavra chave (uma com o termo “fundação” e outra com o termo “Maracaju”) e no sistema RIcon foi realizada a consulta em linguagem natural (“Tipo de fundação do edifício Maracaju?”). A consulta é iniciada com o preenchimento do formulário de consulta do sistema (figura 23).

Figura 23: Formulário de consulta do sistema RIcon

No sistema de busca existente na extranet de projeto com a consulta “Fundação” retornaram 56 documentos. Para a consulta “Maracaju” retornaram 24 documentos. A ordem de apresentação dos documentos no sistema é de acordo com a data de arquivamento. Todos os descritores dos documentos de todas as páginas de retorno das consultas foram analisados pelo autor deste

trabalho para encontrar documentos claramente relevantes. A partir desta análise, em cada busca foram criadas tabelas (tabelas 13 e 14) indicando a ordenação dos resultados das buscas, as posições dos documentos relevantes e a precisão e revocação para cada nível de ordenação.

No sistema RIcon foi executada a consulta “Tipo de fundação do edifício Maracaju?” que retornou 105 documentos. Neste experimento como o sistema RIcon não estava totalmente pronto o processamento das buscas foi executado “manualmente” de acordo com as seguintes etapas:

- Pré-processamento da consulta.
 - Análise Léxica⁷.
 - Eliminação do “?”.
 - Retirada das *stopwords*.
 - Eliminação do “de” e do “do”.
 - Lematização.
 - Eliminação sufixos “o”, “ação” e “io”.

Após o processo de lematização foram feitas consultas com o sistema existente na extranet com os termos restantes:

- Tipo.
- Fundação.
- Edifício.
- Maracaju

Para cada termo das consultas foram executados os processos de busca contextual e de expansão de termos. Para estes termos não foram encontrados termos homógrafos na base de conhecimento e por isso não foi realizada a busca contextual.

⁷ Não são feitas a eliminação dos acentos e a colocação do texto em caixa baixa, pois o sistema da extranet apenas faz a varredura do conteúdo dos arquivos descritores para localização de termos iguais e não faz diferenciação de caixa baixa e caixa alta.

- Pré-processamento da consulta.
 - Expansão de termos por equivalência.
 - Termos “alicerce”, “base” e “sustentáculo” para “fundação”.
 - Termos “casa”, “prédio” e “imóvel” para “edifício”.
 - Não houve expansão para o termo “tipo” e “Maracaju”.
 - Expansão de termos por hierarquia.
 - Termos “estrutura de concreto”, “bloco”, “sapata”, “broca”, “tubulão” e “radier” para “fundação”.
 - Termos “construção”, “bloco” e “apartamento” para “edifício”.
 - Não houve expansão para o termo “tipo” e “Maracaju”.

Com o processo de expansão de termos foram executadas mais 15 buscas além das 4 iniciais. Estas buscas foram criadas a partir da hierarquia e sinônimos de frames armazenadas na base de conhecimento através do programa Protégé-2000 (figura 24). O resultado das 19 buscas foi analisado e eliminado os documentos duplicados para o processo de cálculo de pesos e ordenação.

- Pós-processamento da consulta.
 - Ordenação dos resultados.
 - Cálculo de notas e pesos (tabela 16).

Através da ordenação e eliminação de documentos duplicados nas várias buscas foi formada a tabela 15 com um total de 105 documentos ordenados por relevância de acordo com o cálculos e pesos, indicação da posição dos relevantes e as medidas de precisão e revocação para cada nível de ordenação. Com os valores de precisão e revocação das tabelas foi construído o gráfico de precisão (figuras 25) e revocação (figuras 26) para os dez

primeiros documentos recuperados em cada consulta, além da curva Precisão x Revocação (figuras 27) de todos os documentos recuperados.

Consulta: Tipo fundação do Ed. Maracaju			
Primeira opção SADP: Fundação			
Número	Relevante	Precisão	Revocação
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0
19	0	0	0
20	0	0	0
21	0	0	0
22	0	0	0
23	0	0	0
24	0	0	0
25	0	0	0
26	0	0	0
27	0	0	0
28	0	0	0
29	0	0	0
30	0	0	0
31	0	0	0
32	1	3	10
33	0	3	10
34	0	3	10
35	0	3	10
36	1	6	20
37	0	5	20
38	0	5	20
39	0	5	20
40	0	5	20
41	0	5	20
42	0	5	20
43	0	5	20
44	1	7	30
45	1	9	40
46	1	11	50
47	0	11	50
48	0	10	50
49	0	10	50
50	0	10	50
51	0	10	50
52	0	10	50
53	0	9	50
54	1	11	60
55	1	13	70
56	1	14	80

Tabela 14 – Ordenação, documentos relevantes, precisão e revocação para consulta “Fundação”

Consulta: Tipo fundação do Ed. Maracaju			
Segunda opção SADP: Maracaju			
Número	Relevante	Precisão	Revocação
1	0	0	0
2	1	50	10
3	0	33	10
4	0	25	10
5	1	40	20
6	0	33	20
7	1	43	30
8	1	50	40
9	1	56	50
10	0	50	50
11	0	45	50
12	0	42	50
13	0	38	50
14	0	36	50
15	0	33	50
16	1	38	60
17	1	41	70
18	1	44	80
19	0	42	80
20	0	40	80
21	0	38	80
22	0	36	80
23	0	35	80
24	0	33	80

Tabela 13 – Ordenação, documentos relevantes, precisão e revocação para consulta “Maracaju”

Figura 24: Representação do frame fundação na base de conhecimento através do Protégé-2000

Figura 25: Precisão média para os 10 primeiros documentos

Figura 26: Revocação média para os 10 primeiros documentos

Figura 27: Curva Precisão x Revocação para todos os documentos recuperados

Consulta: Tipo fundação do Ed. Maracatu			
R1Con - Fundação Maracatu			
Número	Relevante	Precisão	Revocação
1	1	100	10
2	1	100	20
3	1	100	30
4	1	100	40
5	1	100	50
6	1	100	60
7	1	100	70
8	0	88	70
9	1	89	80
10	1	90	90
11	0	82	90
12	0	75	90
13	0	69	90
14	0	64	90
15	0	60	90
16	0	56	90
17	0	53	90
18	0	50	90
19	0	47	90
20	0	45	90
21	0	43	90
22	1	45	100
23	0	43	100
24	0	42	100
25	0	40	100
26	0	38	100
27	0	37	100
28	0	36	100
29	0	34	100
30	0	33	100
31	0	32	100
32	0	31	100
33	0	30	100
34	0	29	100
35	0	29	100
36	0	28	100
37	0	27	100
38	0	26	100
39	0	26	100
40	0	25	100
41	0	24	100
42	0	24	100
43	0	23	100
44	0	23	100
45	0	22	100
46	0	22	100
47	0	21	100
48	0	21	100
49	0	20	100
50	0	20	100
51	0	20	100
52	0	19	100
53	0	19	100
54	0	19	100
55	0	18	100
56	0	18	100
57	0	18	100
58	0	17	100
59	0	17	100
60	0	17	100
61	0	16	100
62	0	16	100
63	0	16	100
64	0	16	100
65	0	15	100
66	0	15	100
67	0	15	100
68	0	15	100
69	0	14	100
70	0	14	100
71	0	14	100
72	0	14	100
73	0	14	100
74	0	14	100
75	0	13	100
76	0	13	100
77	0	13	100
78	0	13	100
79	0	13	100
80	0	13	100
81	0	12	100
82	0	12	100
83	0	12	100
84	0	12	100
85	0	12	100
86	0	12	100
87	0	11	100
88	0	11	100
89	0	11	100
90	0	11	100
91	0	11	100
92	0	11	100
93	0	11	100
94	0	11	100
95	0	11	100
96	0	10	100
97	0	10	100
98	0	10	100
99	0	10	100
100	0	10	100
101	0	10	100
102	0	10	100
103	0	10	100
104	0	10	100
105	0	10	100

Tabela 15 – Ordenação, documentos relevantes, precisão e revocação no sistema R1Con

	Arquivos	Tipo	Fundação	Edifício	Maracaju	alicerce	base	sustentáculo	casa	prédio	imóvel	Estrutura de concreto	bloco	sapata	estaca	broca	tubulão	radier	construção	apartamento	Nota Total
n	Peso ($\square_i$)	1,0	1,0	1,0	1,0	0,6	0,6	0,6	0,6	0,6	0,6	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	nt_i
1	139-ES-63-PE-R0.ZIP	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2,5
2	139-ES-64-PE-R0.ZIP	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2,5
3	139-ES-63-PE-R0-PLT.ZIP	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2,5
4	139-ES-64-PE-R0-PLT.ZIP	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2,5
5	139-FUN-01-PE-R3-DWG.ZIP	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2,0
6	139-FUN-01-PE-R3-PLT.ZIP	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2,0
7	139-ES-08-PE-R2.ZIP	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2,0
8	139-ES-08-PE-R2-PLT.ZIP	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2,0
9	139-ES-65-PE-R0.ZIP	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	2,0
#	139-ES-68-PE-R0.ZIP	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2,0

Tabela 16 – Tabela com pesos e notas dos 10 primeiros documentos de uma consulta

A nota total de cada documento neste experimento pelo fato dos descritores serem muito resumidos não foi calculada a função de frequência do termo na base de dados.

ANEXO B – CONSULTAS EXECUTADAS NO EXPERIMENTO

Quantidade de buscas: 15.

Buscas válidas: 5.

No total de 15 buscas elaboradas pelo autor, do usuário coordenador do empreendimento e do desenvolvedor da extranet de projeto algumas foram rejeitadas por não retornarem documentos ou não retornarem documentos suficientes (pelo menos dez em cada mecanismo). Foi pedido para cada agente que enviasse para o autor cinco consultas para avaliação do sistema. As consultas foram refinadas pelo autor deste trabalho de acordo com necessidades de informações especificadas enviadas por e-mail. As consultas resultantes válidas deste processo foram as descritas na tabela 16 e as rejeitadas na tabela 17.

Nº	Consulta
1	Tipo de fundação da portaria
2	Tipo fundação do Ed. Maracaju
3	Projeto de Instalações elétricas
4	Projeto de Instalação Hidráulica
5	Projeto de estrutura da Portaria

Tabela 17 – Consultas válidas

Nº	Consulta
1	Desenhos da caixa d'água
2	Potência das luminárias
3	Sistema de captação de água pluvial
4	Disposição de rede de gás
5	Prumada elétrica
6	Diâmetro e carga das estacas
7	Projeto do pavimento térreo
8	Projeto do subsolo
9	Detalhe das rampas das garagens
10	Projeto da cobertura

Tabela 18 – Consultas rejeitadas

Dados Obtidos

A seguir são apresentados todos os dados obtidos nos experimentos para avaliação do sistema. Os dados estão organizados da seguinte maneira, pela ordem:

- Tabelas com a especificação das consultas válidas no sistema SADP e no RICon constando o número de ordenação e as medidas de precisão e revocação;
- Gráfico da precisão para os dez primeiros documentos ordenados dos dois sistemas;
- Gráfico da revocação para os dez primeiros documentos dos dois sistemas;
- Gráfico da curva precisão x revocação em relação a todos os documentos ordenados dos dois sistemas.

Pesquisa 1

Consulta: Tipo de fundação da portaria			
SADP: Portaria			
Número	Relevante	Precisão	Revocação
1	1	100	33
2	1	100	67
3	0	67	67
4	0	50	67
5	0	40	67
6	0	33	67
7	0	29	67
8	0	25	67
9	0	22	67
10	0	20	67
11	0	18	67
12	0	17	67
13	0	15	67
14	0	14	67
15	0	13	67
16	0	13	67
17	0	12	67

Tabela 19 – Ordenação, precisão e revocação da Pesquisa 1 - SADP

Consulta: Tipo de fundação da portaria			
R1Con: Tipo de fundação da portaria			
Número	Relevante	Precisão	Revocação
1	1	100	33
2	1	100	67
3	0	67	67
4	0	50	67
5	0	40	67
6	0	33	67
7	0	29	67
8	0	25	67
9	0	22	67
10	0	20	67
11	0	18	67
12	0	17	67
13	0	15	67
14	0	14	67
15	0	13	67
16	0	13	67
17	0	12	67
18	1	17	100
19	0	16	100
20	0	15	100
21	0	14	100
22	0	14	100
23	0	13	100
24	0	13	100
25	0	12	100
26	0	12	100
27	0	11	100
28	0	11	100
29	0	10	100
30	0	10	100
31	0	10	100
32	0	9	100
33	0	9	100
34	0	9	100
35	0	9	100
36	0	8	100
37	0	8	100
38	0	8	100
39	0	8	100
40	0	8	100
41	0	7	100
42	0	7	100
43	0	7	100
44	0	7	100
45	0	7	100
46	0	7	100
47	0	6	100
48	0	6	100
49	0	6	100
50	0	6	100
51	0	6	100
52	0	6	100
53	0	6	100
54	0	6	100
55	0	5	100
56	0	5	100
57	0	5	100
58	0	5	100
59	0	5	100
60	0	5	100
61	0	5	100
62	0	5	100
63	0	5	100
64	0	5	100
65	0	5	100
66	0	5	100
67	0	4	100
68	0	4	100
69	0	4	100
70	0	4	100
71	0	4	100
72	0	4	100
73	0	4	100
74	0	4	100
75	0	4	100
76	0	4	100
77	0	4	100
78	0	4	100
79	0	4	100
80	0	4	100
81	0	4	100
82	0	4	100
83	0	4	100
84	0	4	100
85	0	4	100
86	0	3	100
87	0	3	100
88	0	3	100
89	0	3	100
90	0	3	100
91	0	3	100
92	0	3	100
93	0	3	100
94	0	3	100
95	0	3	100
96	0	3	100
97	0	3	100
98	0	3	100
99	0	3	100
100	0	3	100
101	0	3	100
102	0	3	100
103	0	3	100
104	0	3	100
105	0	3	100

Tabela 20 – Ordenação, precisão e revocação da Pesquisa 1 - R1Con

Figura 28: Precisão média para os 10 primeiros documentos – Pesquisa 1

Figura 29: Revocação média para os 10 primeiros documentos – Pesquisa 1

Figura 30: Curva Precisão x Revocação para todos os documentos recuperados – Pesquisa 1

Pesquisa 2

Consulta: Tipo de fundação do Ed. Maracaju			
Primeira opção SADP: Fundação			
Número	Relevante	Precisão	Revocação
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0
19	0	0	0
20	0	0	0
21	0	0	0
22	0	0	0
23	0	0	0
24	0	0	0
25	0	0	0
26	0	0	0
27	0	0	0
28	0	0	0
29	0	0	0
30	0	0	0
31	0	0	0
32	1	3	10
33	0	3	10
34	0	3	10
35	0	3	10
36	1	6	20
37	0	5	20
38	0	5	20
39	0	5	20
40	0	5	20
41	0	5	20
42	0	5	20
43	0	5	20
44	1	7	30
45	1	9	40
46	1	11	50
47	0	11	50
48	0	10	50
49	0	10	50
50	0	10	50
51	0	10	50
52	0	10	50
53	0	9	50
54	1	11	60
55	1	13	70
56	1	14	80

Tabela 22 – Ordenação, precisão e revocação da Pesquisa 2 – 1º opção SADP

Consulta: Tipo de fundação do Ed. Maracaju			
Segunda opção SADP: Maracaju			
Número	Relevante	Precisão	Revocação
1	0	0	0
2	1	50	10
3	0	33	10
4	0	25	10
5	1	40	20
6	0	33	20
7	1	43	30
8	1	50	40
9	1	56	50
10	0	50	50
11	0	45	50
12	0	42	50
13	0	38	50
14	0	36	50
15	0	33	50
16	1	38	60
17	1	41	70
18	1	44	80
19	0	42	80
20	0	40	80
21	0	38	80
22	0	36	80
23	0	35	80
24	0	33	80

Tabela 21 – Ordenação, precisão e revocação da Pesquisa 2 – 2º opção SADP

Consulta: Tipo de fundação do Ed. Maracatu			
RICon- Tipo de fundação Maracatu			
Numero	Relevante	Precisão	Revocação
1	1	100	10
2	1	100	20
3	1	100	30
4	1	100	40
5	1	100	50
6	1	100	60
7	1	100	70
8	0	88	70
9	1	89	80
10	1	90	90
11	0	82	90
12	0	75	90
13	0	69	90
14	0	64	90
15	0	60	90
16	0	56	90
17	0	53	90
18	0	50	90
19	0	47	90
20	0	45	90
21	0	43	90
22	1	45	100
23	0	43	100
24	0	42	100
25	0	40	100
26	0	38	100
27	0	37	100
28	0	36	100
29	0	34	100
30	0	33	100
31	0	32	100
32	0	31	100
33	0	30	100
34	0	29	100
35	0	29	100
36	0	28	100
37	0	27	100
38	0	26	100
39	0	26	100
40	0	25	100
41	0	24	100
42	0	24	100
43	0	23	100
44	0	23	100
45	0	22	100
46	0	22	100
47	0	21	100
48	0	21	100
49	0	20	100
50	0	20	100
51	0	20	100
52	0	19	100
53	0	19	100
54	0	19	100
55	0	18	100
56	0	18	100
57	0	18	100
58	0	17	100
59	0	17	100
60	0	17	100
61	0	16	100
62	0	16	100
63	0	16	100
64	0	16	100
65	0	15	100
66	0	15	100
67	0	15	100
68	0	15	100
69	0	14	100
70	0	14	100
71	0	14	100
72	0	14	100
73	0	14	100
74	0	14	100
75	0	13	100
76	0	13	100
77	0	13	100
78	0	13	100
79	0	13	100
80	0	13	100
81	0	12	100
82	0	12	100
83	0	12	100
84	0	12	100
85	0	12	100
86	0	12	100
87	0	11	100
88	0	11	100
89	0	11	100
90	0	11	100
91	0	11	100
92	0	11	100
93	0	11	100
94	0	11	100
95	0	11	100
96	0	10	100
97	0	10	100
98	0	10	100
99	0	10	100
100	0	10	100
101	0	10	100
102	0	10	100
103	0	10	100
104	0	10	100
105	0	10	100

Tabela 23 – Ordenação, precisão e revocação da Pesquisa 2 – RICon

Figura 31: Precisão média para os 10 primeiros documentos – Pesquisa 2

Figura 32: Revocação média para os 10 primeiros documentos – Pesquisa 2

Figura 33: Curva Precisão x Revocação para todos os documentos recuperados – Pesquisa 2

Pesquisa 3

Consulta: Projeto de Instalações elétricas			
SADP:	Elétrica		
Número	Relevante	Precisão	Revocação
1	1	100	1
2	1	100	3
3	1	100	4
4	1	100	5
5	1	100	7
6	1	100	8
7	1	100	10
8	1	100	11
9	1	100	12
10	1	100	14
11	1	100	15
12	1	100	16
13	1	100	18

Tabela 24 – Ordenação, precisão e revocação da Pesquisa 3 – SADP

Consulta: Projeto de Instalações elétricas			
RICon:	Instalações elétricas		
Número	Relevante	Precisão	Revocação
1	1	100	1
2	1	100	3
3	1	100	4
4	1	100	5
5	1	100	7
6	1	100	8
7	1	100	10
8	1	100	11
9	1	100	12
10	1	100	14
11	1	100	15
12	1	100	16
13	1	100	18
14	1	100	19
15	1	100	21
16	1	100	22
17	1	100	23
18	1	100	25
19	1	100	26
20	1	100	27
21	1	100	28
22	1	100	30
23	1	100	32
24	1	100	33
25	1	100	34
26	1	100	36
27	1	100	37
28	1	100	38
29	1	100	40
30	1	100	41
31	1	100	42
32	1	100	44
33	1	100	45
34	1	100	47
35	1	100	48
36	1	100	49
37	1	100	51
38	1	100	52
39	1	100	53
40	1	100	55
41	1	100	56
42	1	100	58
43	1	100	59
44	1	100	60
45	1	100	62
46	1	100	63
47	1	100	64

Tabela 25 – Ordenação, precisão e revocação da Pesquisa 3 – RICon

Figura 34: Precisão média para os 10 primeiros documentos – Pesquisa 3

Figura 35: Revocação média para os 10 primeiros documentos – Pesquisa 3

Figura 36: Curva Precisão x Revocação para todos os documentos recuperados – Pesquisa 3

Pesquisa 4

Consulta: Projeto de Instalação Hidráulica			
SADP: Hidráulica			
Número	Relevante	Precisão	Revocação
1	1	100	3
2	1	100	7
3	1	100	10
4	1	100	14
5	1	100	17
6	1	100	21
7	1	100	24
8	1	100	28
9	1	100	31
10	1	100	34

Tabela 26 – Ordenação, precisão e revocação da Pesquisa 4 – SADP

Consulta: Projeto de Instalação Hidráulica			
RICon: Instalação Hidráulica			
Número	Relevante	Precisão	Revocação
1	1	100	3
2	1	100	7
3	1	100	10
4	1	100	14
5	0	80	14
6	0	67	14
7	1	71	17
8	1	75	21
9	1	78	24
10	1	80	28
11	1	82	31
12	1	83	34
13	1	85	38
14	1	86	41
15	1	87	45
16	1	88	48
17	1	88	52
18	1	89	55
19	1	89	59
20	1	90	62
21	1	90	66
22	1	91	69
23	1	91	72
24	1	92	76
25	0	88	76
26	0	85	76
27	0	81	76
28	0	78	76
29	0	76	76
30	0	73	76
31	0	71	76
32	0	69	76
33	0	67	76
34	0	66	76
35	0	63	76
36	0	61	76
37	0	59	76
38	0	58	76
39	0	56	76
40	0	55	76
41	0	54	76
42	0	52	76
43	0	51	76
44	0	50	76
45	0	49	76
46	0	48	76
47	0	47	76
48	0	46	76
49	0	45	76
50	0	44	76
51	0	43	76
52	0	42	76
53	0	42	76
54	0	41	76
55	0	40	76
56	0	39	76
57	0	38	76
58	0	38	76
59	0	37	76
60	0	37	76
61	0	36	76
62	0	35	76
63	0	35	76
64	0	34	76
65	0	34	76
66	0	33	76
67	0	33	76
68	0	32	76
69	0	32	76
70	0	31	76
71	0	31	76
72	0	31	76
73	0	30	76
74	0	30	76
75	0	29	76
76	0	29	76
77	0	29	76

Tabela 27 – Ordenação, precisão e revocação da Pesquisa 4 – RICon

Figura 37: Precisão média para os 10 primeiros documentos – Pesquisa 4

Figura 38: Revocação média para os 10 primeiros documentos – Pesquisa 4

Figura 39: Curva Precisão x Revocação para todos os documentos recuperados – Pesquisa 4

Pesquisa 5

Consulta: Projeto de estrutura da Portaria				
SADP: Portaria				
Número	Relevante	Precisão	Revocação	
1	1	100	13	
2	1	100	25	
3	0	67	25	
4	1	75	38	
5	1	80	50	
6	1	83	63	
7	1	88	75	
8	0	75	75	
9	0	67	75	
10	0	60	75	
11	0	55	75	
12	0	50	75	
13	0	48	75	
14	0	43	75	
15	0	40	75	
16	0	38	75	
17	0	35	75	

Tabela 28 – Ordenação, precisão e revocação da Pesquisa 5 – SADP

Consulta: Projeto de estrutura da Portaria				
RiCon: Estrutura da Portaria				
Número	Relevante	Precisão	Revocação	
1	1	100	13	
2	1	100	25	
3	1	100	38	
4	1	100	50	
5	1	100	63	
6	1	100	75	
7	1	100	88	
8	1	100	100	
9	0	88	100	
10	0	80	100	
11	0	73	100	
12	0	67	100	
13	0	62	100	
14	0	57	100	
15	0	53	100	
16	0	50	100	
17	0	47	100	
18	0	44	100	
19	0	42	100	
20	0	40	100	
21	0	38	100	
22	0	36	100	
23	0	35	100	
24	0	33	100	
25	0	32	100	
26	0	31	100	
27	0	30	100	
28	0	29	100	
29	0	28	100	
30	0	27	100	
31	0	26	100	
32	0	25	100	
33	0	24	100	
34	0	24	100	
35	0	23	100	
36	0	22	100	
37	0	22	100	
38	0	21	100	
39	0	21	100	
40	0	20	100	
41	0	20	100	
42	0	19	100	
43	0	19	100	
44	0	18	100	
45	0	18	100	
46	0	17	100	
47	0	17	100	
48	0	17	100	
49	0	16	100	
50	0	16	100	
51	0	16	100	
52	0	15	100	
53	0	15	100	
54	0	15	100	
55	0	15	100	
56	0	14	100	
57	0	14	100	
58	0	14	100	
59	0	14	100	
60	0	13	100	
61	0	13	100	
62	0	13	100	
63	0	13	100	
64	0	13	100	
65	0	12	100	
66	0	12	100	
67	0	12	100	
68	0	12	100	
69	0	12	100	
70	0	11	100	
71	0	11	100	
72	0	11	100	
73	0	11	100	
74	0	11	100	
75	0	11	100	
76	0	11	100	
77	0	10	100	
78	0	10	100	
79	0	10	100	
80	0	10	100	
81	0	10	100	
82	0	10	100	
83	0	10	100	
84	0	10	100	
85	0	9	100	
86	0	9	100	
87	0	9	100	
88	0	9	100	
89	0	9	100	
90	0	9	100	
91	0	9	100	
92	0	9	100	
93	0	9	100	
94	0	9	100	
95	0	9	100	
96	0	9	100	
97	0	9	100	
98	0	9	100	
99	0	9	100	
100	0	8	100	
101	0	8	100	

102	0	8	100
103	0	8	100
104	0	8	100
105	0	8	100
106	0	8	100
107	0	7	100
108	0	7	100
109	0	7	100
110	0	7	100
111	0	7	100
112	0	7	100
113	0	7	100
114	0	7	100
115	0	7	100
116	0	7	100
117	0	7	100
118	0	7	100
119	0	7	100
120	0	7	100
121	0	7	100
122	0	7	100
123	0	7	100
124	0	6	100
125	0	6	100
126	0	6	100
127	0	6	100
128	0	6	100
129	0	6	100
130	0	6	100
131	0	6	100
132	0	6	100
133	0	6	100
134	0	6	100
135	0	6	100
136	0	6	100
137	0	6	100
138	0	6	100
139	0	6	100
140	0	6	100
141	0	6	100
142	0	6	100
143	0	6	100
144	0	6	100
145	0	6	100
146	0	5	100
147	0	5	100
148	0	5	100
149	0	5	100
150	0	5	100
151	0	5	100
152	0	5	100
153	0	5	100
154	0	5	100
155	0	5	100
156	0	5	100
157	0	5	100
158	0	5	100
159	0	5	100
160	0	5	100
161	0	5	100
162	0	5	100
163	0	5	100
164	0	5	100
165	0	5	100
166	0	5	100
167	0	5	100
168	0	5	100
169	0	5	100
170	0	5	100
171	0	5	100
172	0	5	100
173	0	5	100
174	0	5	100
175	0	5	100
176	0	5	100
177	0	5	100
178	0	4	100
179	0	4	100
180	0	4	100
181	0	4	100
182	0	4	100
183	0	4	100
184	0	4	100
185	0	4	100
186	0	4	100
187	0	4	100
188	0	4	100
189	0	4	100
190	0	4	100
191	0	4	100
192	0	4	100
193	0	4	100
194	0	4	100
195	0	4	100
196	0	4	100
197	0	4	100
198	0	4	100
199	0	4	100
200	0	4	100

Tabela 29 – Ordenação, precisão e revocação da Pesquisa 5 – RICon

Figura 40: Precisão média para os 10 primeiros documentos – Pesquisa 5

Figura 41: Revocação média para os 10 primeiros documentos – Pesquisa 5

Figura 42: Curva Precisão x Revocação para todos os documentos recuperados – Pesquisa 5

REFERÊNCIAS

AMORIM, S. R. L.; PEIXOTO, L.; NUNES, R.; MADEIRA, L. C. Specification Standards for the New Management Settings of Construction. **Proceedings** IGLC - 2002, Gramado, Brazil.

AMORIM, S. R. L.; PEIXOTO, L. A. Classificação de produtos e serviços com uso de facetas. In: SIMPÓSIO BRASILEIRO DE GESTÃO E ECONOMIA DA CONSTRUÇÃO, 3. **Anais**. São Carlos: SIBRAGEC, 2003. 1 CD-ROM.

ANDERY, P.; CARVALHO JR, A.N.; HELMAN, H Looking for What Could be Wrong: an Approach to Lean Thinking. In: Sixth Annual Conference of the International Group for Lean Construction - IGLC-6. **Proceedings**. Guarujá. 1998.

ANDREASEN, T. "An approach to knowledge-based query evaluation", **Fuzzy Sets and Systems**, v.140, n.1, Elsevier, 2003.

ANDRESEN, J. et al. A Framework for measuring IT innovation benefits. **Electronic Journal of Information Technology in Construction**, v.5, p. 57-72, 2000.

ANDRESEN, J. L.; CHRISTENSEN, K.; HOWARD, R.W. Project Management with a Project Web. **Electronic Journal of Information Technology in Construction**, v.8, p. 29-41, 2003.

ANTAC. **CDCON - Diretrizes Gerais para o Projeto**. 2002.

ANTAC. **CDCON - Terminologia e codificação para construção**. Disponível em: <www.cdcon.ufjf.br/>. Acesso em 18 ago. 2003.

BAEZA, R. Y.; RIBEIRO, B. N. **Modern Information Retrieval**. Addison Wesley, 1999.

BARSOTTI, R. **A informática na biblioteca e na documentação**. São Paulo:Polis/apb, 1990.

CALDAS, C.H.S.; SOIBELMAN, L. Avaliação da Logística de Informação em Processos Inter-Organizacionais na Construção Civil. In: SIMPÓSIO BRASILEIRO DE GESTÃO DA QUALIDADE E

ORGANIZAÇÃO DO TRABALHO NO AMBIENTE CONSTRUÍDO, 2. **Anais**. Fortaleza: ANTAC, 2001. 1 CD-ROM.

CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO; FUNDAÇÃO GETULIO VARGAS. **Macrossetor da Construção Produz Quase 20% do PIB (Release)**. 2002. Disponível em: <<http://www.moradia.org.br/BuscaDados.asp>>. Acesso em: 11 fev. 2003.

CAMBIAGHI, H. et al. **Diretrizes Gerais para Intercambialidade de Projetos em CAD: integração entre projetistas, construtoras e clientes**. São Paulo, Pini, 2002. 199p.

CARDOSO, L. R. A.; ABIKO, A. K.; GONCALVES, O. M.; BARBOSA, A. L. S.F; INOUE, K. P.; HAGA, H. C. R. Proposição de um Modelo para a Cadeia Produtiva da Construção Habitacional Urbana: Desenho e Fluxos. In: XXII Encontro Nacional de Engenharia de Produção, Curitiba. **Anais**. ENEGEP 2002, 2002.

CRUZ, T. **Workflow: a Tecnologia que vai Revolucionar Processos**. 1 ed. São Paulo: Atlas. 1998. 222 p.

CSI. **Masterformat : master list of numbers and titles for the construction industry**. Toronto: The Construction Specifications Institute. 1995. 2v.

DAVENPORT, T.H.; PRUSAK, L. **Conhecimento Empresarial**. São Paulo: Campos, 1999.

DAWOOD, N.; AKINSOLA, A.; HOBBS, B. Development of automated communication of system for managing site information using Internet technology. **Automation in Construction**. v.11, n.5, p. 557-572, 2002

DRI-WEFA; F. W. DODGE, **Global Construction Study 2002**. McGraw Hill Construction, 2002.

DURKIN, J. **Expert Systems: Design and Development**. Prentice-Hall, New Jersey, 1994.

EDMUNDS, A.; MORRIS, A. The problem of information overload in business organizations: a review of the literature. **International**

Journal of Information Management, v. 20, n.1, p.17-28, 2000.

FARHOOMAND, A.; DRURY, D. Managerial Information Overload. **Communications of the ACM**, v. 45, n. 10, p.127-131, 2002.

FUSARO, P.S. **Uma Linguagem de Consulta para um Sistema de Recuperação em Texto Completo**. Dissertação (Dissertação em Ciência da Computação), Belo Horizonte, UFMG, 1988.

HARMAN, D. The TREC Conferences, in K. Spark Jones and P. Willet, editors, **Readings in Information Retrieval**, Morgan Kaufmann, San Francisco, chapter 4, 247-256, 1997.

HARROD, G. aecXML & IFC. **CADinfo**, Digital Business Media. In: <<http://www.cainfo.net/editorial/aecxml.htm>>. Acessado em 18 de agosto de 2002.

HASSELL, S; BERNSTEIN, M.; BOWER, A. **The Role of Information Technology in Housing Design and Construction**. RAND Science and Technology Policy Institute, CF-156-OSTP, 2000. Disponível em: <<http://www.rand.org/publications/CF/CF156/>>. Acesso em 15 set. 2002.

IAI **aecXML**. Disponível em: <<http://www.iai-na.org/aecxml/mission.php>>. Acesso em: 18 ago. 2003.

IAI **IFC2x Technical Guide**. International Alliance for Interoperability, 2000.

ISO **Classification of information in the construction industry**. ISO Technical Report 14177:1994(E). Geneva: International Organisation for Standardisation. 3732-3742, 1994.

KARP, P.D. The Design Space of Frame Knowledge Representation Systems. Artificial Intelligence Center. **SRI AI Center Technical Note #520**. 1993.

KOSOVAC, B.; VANIER, D.J.; FROESE, T.M. Use Of Keyphrase Extraction Software For Creation Of An AEC/FM Thesaurus, **Electronic Journal of Information Technology in Construction**, Vol.5, 2000, p.25, 2000.

LANTELME, E.M.V. **Proposta de um sistema de indicadores de qualidade e produtividade para a construção civil.** .Dissertação (Mestrado em Engenharia Civil), Porto Alegre, RS. UFRGS, 124p, 1994.

LAURINDO, F.J.B. **Um estudo sobre a avaliação da eficácia da tecnologia de informação nas organizações.** Tese (Doutorado em Engenharia). São Paulo, Escola Politécnica da USP, 2000.

LICHTENSTEIN, N.B. **Patologia das construções: procedimento para formulação do diagnóstico de falhas e definição de conduta adequada à recuperação de edificações.** Dissertação (Mestrado). São Paulo, Escola Politécnica da USP, 1985.

LOH, S.; WIVES, L.K.; FRAINER, A.S. Uma abordagem para busca contextual de documentos na Internet. **Revista de Informática Teórica e Aplicada, RITA**, v.4. Porto Alegre, 1997.

MAHER, M.L.; SIMOFF, S.J. Ontology-based multimedia data mining for design information retrieval. In: Computing in Civil Engineering, ASCE. **Proceedings**. pp. 212-223, 1998.

MELHADO, S.B. **Gestão, cooperação e integração para um novo modelo voltado a qualidade do processo de projeto na construção de edifícios.** Tese (Livre-docência). Escola Politécnica da USP, São Paulo, 2001.

MICALI, J.F.M. **Um modelo para integração da indústria da construção civil.** São Paulo, 2000. 151p. Tese (Doutorado) - Escola Politécnica, Universidade de São Paulo.

MIZZARO, S. A. Cognitive Analysis of Information Retrieval. In: Information Science: Integration in Perspective - CoLIS2, 1996, Copenhagen, Denmark. **Proceedings**. The Royal School of Librarianship, 1996. p.233-250.

MULLIER, D.; DIXON, M. Authoring Educational Hypermedia Using a Semantic Network. **World Conference on Educational Multimedia, Hypermedia and Telecommunications 2000(1)**, p.295-300. Available: <<http://dl.aace.org/1259>>.

NASCIMENTO, L.A.; SANTOS, E.T. Estratégia para Aprimoramento da Eficiência e Produtividade na Construção Civil através da Análise dos Processos de um Empreendimento. In. **V Congresso de Engenharia Civil**, 2002, Juiz de Fora - MG, CD-ROM, UFJF, 2002a.

NASCIMENTO, L. A.; SANTOS, E.T. Barreiras para o uso da Tecnologia da Informação na Indústria da Construção Civil. WORKSHOP NACIONAL GESTÃO DO PROCESSO DE PROJETO NA CONSTRUÇÃO DE EDIFÍCIOS, 2., 2002b, Porto Alegre. **Anais**. Porto Alegre: PUC-RS, 2002b. 1 CD-ROM

NASCIMENTO, L. A. ; SANTOS, E. T. Sistema Baseado em Conhecimento para Recuperação de Informação em Repositórios de Documentos de Projetos da Indústria da Construção Civil. III Workshop Gestão do Processo de Projeto na Construção de Edifícios. **Anais**. Belo Horizonte, MG, 2003.

NEWTON, P. Diffusion of IT in the Building and Construction Industry. **Building for Growth Innovation Forum**, Sydney, 1998. Disponível: <<http://www.isr.gov.au/industry/building/invforum/>>. Acesso em: 16 set. 2002.

OCCS Development Committee, **The Overall Construction Classification System - A Strategy for Classifying the Built Environment** - Preliminary draft for review and comment, 2001. Disponível em <<http://www.occsnet.org>>. Acesso em 14 out. 2003.

ONESTAT. **Most people use 2 word phrases in search engines according to OneStat.com**. Amsterdam, SC, 2004. Disponível em: <http://www.onestat.com/html/aboutus_pressbox27.html>. Acesso em: 18 fev. 2004.

PORTER, M.F. An algorithm for suffix stripping, in K. Spark Jones and P. Willet, editors, **Readings in Information Retrieval**, Morgan Kaufmann, San Francisco, 313-316, 1997.

PICORAL, R.B.; SOLANO, R.S. O uso da extranet na coordenação de projetos: aplicação em estudo de caso. In.: Workshop Nacional Gestão do Processo de Projeto na Construção de Edifícios, 1, 2001, São Carlos. **Anais**. São Carlos: UFSCar, 2001. CD-ROM

RIBEIRO, J.M.P. **Algoritmo de Stemming**. Disponível em: <<http://www.fe.up.pt/~jrbeiro/ari/stemming.html>>. Acesso em: 24 nov. 2003.

RIJSBERGEN, C.J. **Information Retrieval**. Butterworths, 1979.

RIVARD, H. A Survey on the Impact of Information Technology on the Canadian Architecture, Engineering and Construction Industry. **Electronic Journal of Information Technology in Construction**, v.5, p. 37-56, 2000. Disponível em: <<http://www.itcon.org/2000/3/>>. Acesso em: 16 set. 2002.

ROE, A.; REINA, P. Learning to Share is Tougher Than Anyone Anticipated - Industry fragmentation impedes automation. **E-construction**. McGraw-Hill. Disponível em: <<http://enr.construction.com/features/technologyconst/archives/010813c.asp>>. Acesso em 16 set. 2002.

SALTON, G; BUCKLEY, C. "Term-Weighting Approaches in Automatic Text Retrieval", **Information Processing and Management**, v. 24, n. 5, 513-523, 1988.

SCHMITT, C. M. Documentação de projetos para edificação: solução para este quebra-cabeças. In: Encontro Nacional de Tecnologia do Ambiente Construído, 5, 1993, São Paulo. **Anais**. São Paulo: ANTAC, 1993. v.2, p.637.

SANTOS, E. T; NASCIMENTO, L. A. Recuperação de Informação em Sistemas de Informações na Construção Civil: o caso das Extranets de Projeto. **Seminário TIC2002 - Tecnologia da Informação e Comunicação na Construção Civil - Oportunidades e Futuro**, Curitiba, PR, 2002.

SANTOS, E. T; NASCIMENTO, L. A. Knowledge-Based Information Retrieval in Project Extranets. **Proceedings: Seventh International Conference on the Application of Artificial Intelligence to Civil and Structural Engineering**. Paper 3, Stirling: Civil-Comp Press, 2003.

SCHWEGLER, B.R. et al. Near- Medium- and Long-Term Benefits of Information Technology in Construction. **CIFE Working Paper #65**, 2001. Disponível em: <<http://www.stanford.edu/group/CIFE/online.publications/WP065.pdf>>. Acesso em: 16 set. 2002.

SILVA, M.F.S; HEINECK, L.F.M. Equipes de projeto de edificações e seu potencial como equipes de gestão de conhecimento: uma reflexão preliminar. In.: Workshop Nacional Gestão do Processo de Projeto na Construção de Edifícios, 2, 2001, São Carlos. **Anais.** São Carlos: UFSCar, 2001. CD-ROM.

SOIBELMAN, L.; CALDAS, C.H.S. O uso de extranets no gerenciamento de projetos: o exemplo norte americano. **Anais do ENTAC 2000.** v.1, p. 588-595, 2000.

SOIBELMAN, L.; CALDAS, C.H.S. Automated classification and integration of text-based information in construction inter-organizational systems. **ASCE-FIAPP Conference at Virginia Tech** - Blacksburg, VA, 2001.

STEWART, R.A.; MOHAMED, S.; DAET, R. Strategic implementation of IT/IS projects in construction: a case study. **Automation in Construction.** v.11, n.6, p. 681-694, 2002.

SULZBACH, S.I. Modelo de representação do conteúdo na base de conhecimento de um sistema tutor inteligente. **Seminário de EAD – Representação do Conhecimento UFRGS.** Porto Alegre, 2001.

TEICHOLZ, P. Productivity Trends in the Construction Industry. PowerPoint presentation. **AISC Annual Convention**, 2000.

TOLEDO, R.; ABREU, ALINE F.; JUNGLES, A.E. A difusão de inovações tecnológicas na indústria da construção civil. In: ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO, 8, 2000, Bahia. **Anais.** Bahia: ANTAC, 2000.

TZORTZOPOULOS, P. **Contribuições para o desenvolvimento de um modelo do processo do projeto de edificações em empresas construtoras incorporadoras de pequeno porte.** Dissertação (Mestrado). Universidade Federal do Rio Grande do Sul, Porto Alegre, 1999.

VICKERY, B.C. **Techniques of Information Retrieval.** Butterworths, London, 1970.

ZAZO, A.F.; FIGUEROLA, C.G.; BERROCAL, J.L.A.; RODRÍGUEZ, E. Term Expansion Using Stemming and Thesauri in Spanish. In: C.

Peters, editor, Results of the CLEF 2002 Cross-Language System Evaluation Campaign. **Working Notes for the CLEF 2002 Workshop**, p. 177-183, Rome, 2002.

ZEGARRA, S.L.V.; FRIGIERI Jr., V.; CARDOSO, F.F. A tecnologia da informação e a indústria da construção de edifícios. In: Simpósio Brasileiro de Gestão da Qualidade e Organização do Trabalho, 1. Recife, 1999. **Anais**. I SIBRAGEQ, Recife : GEQUACIL/UPE, 1999. v. 1, p. 71-80.

WATSON, A.; DAVOODI, M. Transferring project documents and associated metadata between company document management systems and project extranets. In: ESM@ART 2002 CONFERENCE, 2002, Salford. **Anais**. Salford: University of Salford, 2002.