

XI Encontro Nacional de Pesquisa em Ciência da Informação - **XI ENANCIB
2010**

GT 2: Organização e Representação do Conhecimento

Metadados textuais e visuais para recuperação de informação em imagens

Comunicação Oral

RESUMO

Esta pesquisa faz uma discussão sobre os conceitos de metadados em diferentes áreas de conhecimento. Em seguida, propõe um sistema de recuperação híbrido que utiliza metadados textuais e visuais para organização e recuperação de informação contida nos desenhos técnicos de projeto de engenharia. Finalmente, apresenta um exemplo de classificação de um documento, utilizando o modelo híbrido proposto. Atualmente, existem duas linhas de pesquisa independentes que estudam recuperação de imagens: uma baseada na descrição textual da imagem e outra baseada no conteúdo da imagem. Atualmente, os sistemas de informação para projetos de engenharia utilizam basicamente metadados textuais que descrevem a imagem. A proposta deste trabalho envolve, além das descrições (metadados de conceito ou de assunto), o uso da própria imagem ou partes dela (metadados de conteúdo) para recuperação. O sistema híbrido proposto representa um grande avanço para a recuperação da informação de imagens e desenhos. Quando o sistema não considera os metadados visuais, ou seja, a organização e recuperação da imagem pela imagem, a recuperação do documento pode ficar restrita a quem a executou ou a descreveu. O conteúdo visual da imagem é um metadado importante para a definição do assunto do documento na íntegra, independente de interpretações e descrições passíveis de duplicidades e erros.

Palavras chave: Metadados textuais, metadados visuais, recuperação de informação.

ABSTRACT

This paper presents a discussion on the concept of metadata in different areas of knowledge. Then, it proposes a hybrid retrieval system that uses textual and visual metadata for information organization and retrieval of engineering technical drawings. Finally, it presents an example of the use of the model proposed to classify a document. Currently, there are two lines of research on image retrieval: one is based on the image textual description and the other is based on image visual content. Existing information systems for engineering drawings use only textual metadata which describe the image. The model proposed uses both image description (concept metadata) and the image or part of the image (content metadata) for retrieval. The hybrid system proposed represents a major advance for information retrieval of images and technical drawing. When the system does not consider the visual metadata, i.e., the organization and retrieval of the image by the image, document retrieval may be restricted to those who performed or described it. The visual content of image is an important metadata for defining the subject of the document in full, regardless of interpretations and descriptions subject to duplications and errors.

Keywords: Text metadata, visual metadata, information retrieval.

1 INTRODUÇÃO

Atualmente o desenvolvimento das tecnologias da informação proporciona uma quantidade imensurável de dados e, nesse contexto, os metadados se tornaram essenciais para que a informação possa ser organizada e recuperada.

O objetivo deste trabalho é fazer uma discussão sobre definição de metadados e a transposição dos conceitos de metadados de Sistemas de Recuperação de Informação textuais para visuais. Outro objetivo é propor um sistema que considera os metadados textuais e visuais para a organização e recuperação de informação em base de dados de projetos de engenharia.

Este trabalho fará, primeiramente, uma discussão sobre a definição de metadados considerando os conceitos da ciência da informação e da ciência da computação. Em seguida, elaborará um modelo de representação de imagens que considera os metadados textuais e visuais presentes no documento. Finalmente, apresenta um exemplo de classificação e definição dos metadados presentes em desenhos técnicos que compõem os projetos de engenharia.

O trabalho conclui que é necessário avançar as pesquisas e definições de conceitos referentes aos documentos de diferentes formas, como, por exemplo, os desenhos técnicos de engenharia. Esses documentos possuem informações que podem ser tratadas automaticamente, considerando metadados textuais e visuais em um sistema híbrido.

2 REVISÃO DA LITERATURA

Metadado é definido como sendo dado sobre o dado. Metadados descrevem os atributos e o conteúdo de um documento original, e se usados de forma efetiva, viabilizam o acesso à informação precisa (Milstead & Feldman,1999).

Segundo Alvarenga (2005), metadado pode ser considerado um ponto de acesso, termo da área da catalogação, e parece ter sido cunhado em contextos externos à ciência da

informação. Os denominados padrões de metadados estão atualmente, presentes no instrumental de software para criação de objetos digitais.

Na ciência da informação considera-se metadado o atributo utilizado para caracterizar uma entidade. Por exemplo, a entidade livro possui atributos tais como autor, título, número de páginas. Os atributos são inerentes à entidade enquanto os metadados são escolhidos entre os atributos de acordo com a necessidade do usuário de um sistema de recuperação da informação. Metadados têm naturezas diversas e podem ser classificados em categorias tais como: técnicos, estruturais, administrativos, bibliográficos.

De acordo com Bax (2001), metadados são considerados linguagens de marca que identificam, de forma descritiva, cada entidade informacional digna de significado presente nos documentos, como, por exemplo, parágrafos, títulos, tabelas, gráficos. A partir dessas descrições os programas de computador tratam a informação contida em documentos eletrônicos. Linguagens de marca foram desenvolvidas por comunidades responsáveis pela criação ou provimento de acesso a fontes eletrônicas interligadas, utilizando-se de métodos e padrões diversos para a organização de recursos digitais presentes na web.

As organizações envolvidas na geração, organização, recuperação e uso de informação eletrônica têm procurado padrões que possam ser adotados para a catalogação. Elementos como o autor, o título, o assunto são exemplos de metadados que podem ser usados para descrever tanto um livro em uma biblioteca manual ou online, quanto para descrever uma *home page*, uma base de dados ou outro recurso eletrônico qualquer para o ambiente Web (Souza et al., 2005).

Para Marchiori (1998) a representação da informação por metadados é uma necessidade e visa criar uma estrutura de descrição padronizada de documentos eletrônicos, para tornar mais efetiva a recuperação de informação em redes eletrônicas. O grande problema para o uso atual de metadados é a insuficiência de padrões.

A informação em mídia eletrônica necessita de métodos apropriados de descrição, pois possui elementos e especificidades que não são contemplados pelos métodos tradicionais de tratamento e descrição, segundo Hakala et al. (1998). É possível adotar-se metadados de formato simples, de fácil entendimento pelos autores, catalogadores, publicadores e demais

pessoas envolvidas com a publicação de documentos eletrônicos como é o caso do padrão Dublin Core.

A Dublin Core Metadata Initiative (DCMI) é uma organização com a função de estabelecer políticas e padrões de armazenamento de dados, bem como desenvolver uma linguagem padrão que facilite o acesso e torne os sistemas de busca de informações mais inteligentes. O padrão DC surgiu no encontro realizado em 1995 e foi concebido para descrever recursos da internet e afirmar a necessidade de uma nomenclatura comum dos metadados. Foi estabelecido um conjunto de elementos para descrição dos dados: título, criador, assunto, descrição, data, tipo, formato, identificador, origem, idioma e relação.

Quando se trata da imagem, o conteúdo visual pode ser classificado, segundo Barreto (2007) em conteúdo primitivo de imagens que refere aos elementos básicos que compõem a imagem e conteúdo complexo de imagens que refere aos padrões de uma imagem que são percebidos por seres humanos como fontes de significados. O conteúdo primitivo são características visuais que podem ser reconhecidas e extraídas automaticamente pelo computador.

Resgata-se nesse sentido o pensamento de Alvarenga, quando afirma que:

[...] a representação do conhecimento no novo contexto digital não compreende somente a substituição do documento primário por uma informação catalográfica. Considerando-se que o documento não se acha fisicamente em espaço diferente dos dados catalográficos, mas no meio comum que a ambos proporciona materialidade, novas formas de se criar índices de recuperação tornaram-se possíveis (2003. p.).

Segundo a autora acima citada, no novo contexto de produção, organização e recuperação de objetos digitais, o trabalho dos responsáveis por bases de dados documentais não se restringe à criação de representações de objetos físicos constantes de um acervo, mas compreendem, muitas vezes o estabelecimento de novos tipos de metadados, muitos dos quais podem ser indicados diretamente nos próprios itens, constituindo-se esses em chaves de acesso a documentos e seus conteúdos, tal como ocorre nas linguagens de marcação. Este seria o caso dos metadados icônicos, tal como proposto neste trabalho, criados pela identificação no próprio item.

Os metadados sejam extraídos automaticamente pelo computador ou descritos pelo homem pode ser classificado de acordo com a relação que tem com a imagem. Segundo

Barreto (2007) os metadados podem ser definidos como: metadados independentes do conteúdo - dados que não concernem diretamente ao conteúdo da imagem ou vídeo, mas estão relacionados com este, como o formato da imagem, autoria, data, local, condições de iluminação, etc.; metadados dependentes do conteúdo - dados que se referem a características consideradas de nível baixo e médio, como cor, textura, forma, relações espaciais e combinações destes; metadados descritivos do conteúdo - dados que se referem ao conteúdo semântico e que concernem às relações das entidades da imagem com entidades do mundo real ou emoções e significados associados a sinais visuais e cenas. O conteúdo primitivo pode ser extraído automaticamente e, atualmente, é uma área estudada por cientistas da computação.

Para Marques et al. (2009), existem atualmente duas abordagens para o problema da indexação de imagens: a primeira é conhecida como indexação com base no conteúdo e a segunda é conhecida como indexação com base em conceitos. Na abordagem que se baseia no conteúdo, a indexação é realizada de modo automático, por algoritmos computacionais, através da extração de características de cor, textura e arranjo de figuras geométricas presentes na imagem. Na indexação com base em conceitos, as imagens são representadas por uma lista de palavras-chave que se referem às informações presentes na imagem. A indexação por conteúdo opera no nível sintático, enquanto a indexação por conceitos opera no nível semântico.

Resumidamente, as técnicas de recuperação da informação, baseadas no conteúdo visual da imagem, utilizam algoritmos que testam a imagem inteira ou parte dela, para identificar imagens semelhantes. Normalmente, utilizam-se atributos referentes ao conteúdo visual da imagem e não se consideram atributos textuais referentes à descrição ou à interpretação da imagem. Um sistema Content-Based Image Retrieval - CBIR de recuperação da informação, baseado no conteúdo visual da imagem, extrai características da imagem, indexa e processa consultas efetuadas pelo usuário. O princípio básico do CBIR requer que as propriedades visuais da imagem sejam usadas no lugar da descrição textual para recuperar dado icônico, com o objetivo da busca em uma base de dados (Baracho, 2007).

Os sistemas de recuperação de imagens por conteúdo são denominados CBIR (Content Based Image Retrieval) ou CBVIR e podem ser construídos para encontrar imagens de duas formas: a busca por exemplo, em que se utilizam como chave de busca as características visuais de uma imagem ou esboço de referência; e a busca textual, realizada a partir da

transcrição de significados ou conceitos contidos na imagem que foram previamente relacionados a características visuais específicas. Na pesquisa em texto, as palavras serão procuradas na base criada a partir da análise de significados implícitos no conteúdo visual, processo denominado recuperação semântica, pois fundamentalmente diz respeito à relação entre um signo e aquilo a que ele se refere. Os sistemas mais reportados para extração de metadados dependentes de conteúdo são os que reconhecem semelhanças entre características visuais, onde usualmente temos como base de pesquisa a similaridade de cores, de formas ou de texturas, ou uma combinação destes parâmetros.

Atualmente, a organização dos projetos de engenharia, com a utilização de sistemas de gerenciamento de documentos, armazena o desenho técnico com informações textuais. Utiliza um sistema de gerenciamento de banco de dados baseado em atributos textuais para indexar e recuperar o desenho técnico do projeto.

3 PROPOSTA DE UM MODELO DE METADADOS PARA DESCRIVER DESENHOS TÉCNICOS

A partir dos conceitos acima foi proposto um modelo para se determinar, em projetos de engenharia, os metadados necessários para sua classificação e caracterização.

O modelo proposto considera inicialmente a classificação do projeto, baseado na análise semântica do desenho e na definição das categorias a que ele pertence. Em seguida, baseado na análise sintática e a extração dos metadados textuais e visuais.

Metadados de classificação do projeto – definição das categorias

Cada projeto de engenharia pode gerar milhares de desenhos técnicos referentes a uma obra específica. Cada ramo da engenharia produz um conjunto de projetos necessários para a construção de um objeto. Especificamente, a engenharia civil engloba o projeto arquitetônico, estrutural, hidráulico, elétrico, de combate e prevenção a incêndio, de ar condicionado, entre outros. Cada etapa de desenvolvimento do projeto de engenharia/arquitetura é composta de um conjunto de desenhos técnicos, que representam, através de várias vistas, o objeto a ser construído. Os desenhos técnicos possuem nomenclaturas e normas próprias de desenvolvimento. Dentre os desenhos técnicos de engenharia/arquitetura, tem-se planta, corte

ou seção, vistas, elevações, fachadas, vista frontal, vista lateral direita, vista lateral esquerda, vista posterior e perspectivas. A planta da edificação é um desenho, uma imagem, que representa uma edificação cortada no plano horizontal e contém a representação dos elementos construtivos. Possui normas técnicas de desenho que direcionam e facilitam a interpretação humana através de recursos, por exemplo, de espessura de linha mais grossa para a representação dos objetos cortados que representam as paredes; ícones de peças sanitárias, representando as instalações sanitárias; ícones de carros representando o espaço destinado à garagem; ícones de portas representando os acessos e ligações entre os cômodos; linhas de cota que contêm as dimensões; linhas tracejadas que representam projeções; retícula quadriculada que representa área molhada, entre outros (FIG. 1).

Esses desenhos podem ser classificados de acordo com três categorias. Estas categorias são descritas à partir da observação do desenho e referem ao conteúdo semântico da imagem. Quando um engenheiro ou pessoa, que tem noção de projeto, observa um determinado desenho técnico, consegue inferir e identificar quais são os metadados referentes ao tipo, processo e forma. É o primeiro tipo de metadados para caracterização do desenho e são determinantes dos demais metadados.

A primeira categoria chamada “Tipo” define o tipo de projeto que está representado no desenho, dentre as várias disciplinas da engenharia. O filtro Tipo define se o projeto é arquitetônico, estrutural, elétrico ou hidráulico.

A segunda categoria chamada “Processo” define o estágio do desenvolvimento do projeto. Define-se o processo através da interpretação do nível de detalhamento do projeto, se ele está na fase de projeto preliminar, projeto executivo, projeto de detalhamento ou projeto de apresentação.

A terceira categoria chamada “Forma” define a representação gráfica do desenho. A representação gráfica do projeto se subdivide em vista, planta, corte, perspectiva e outros.

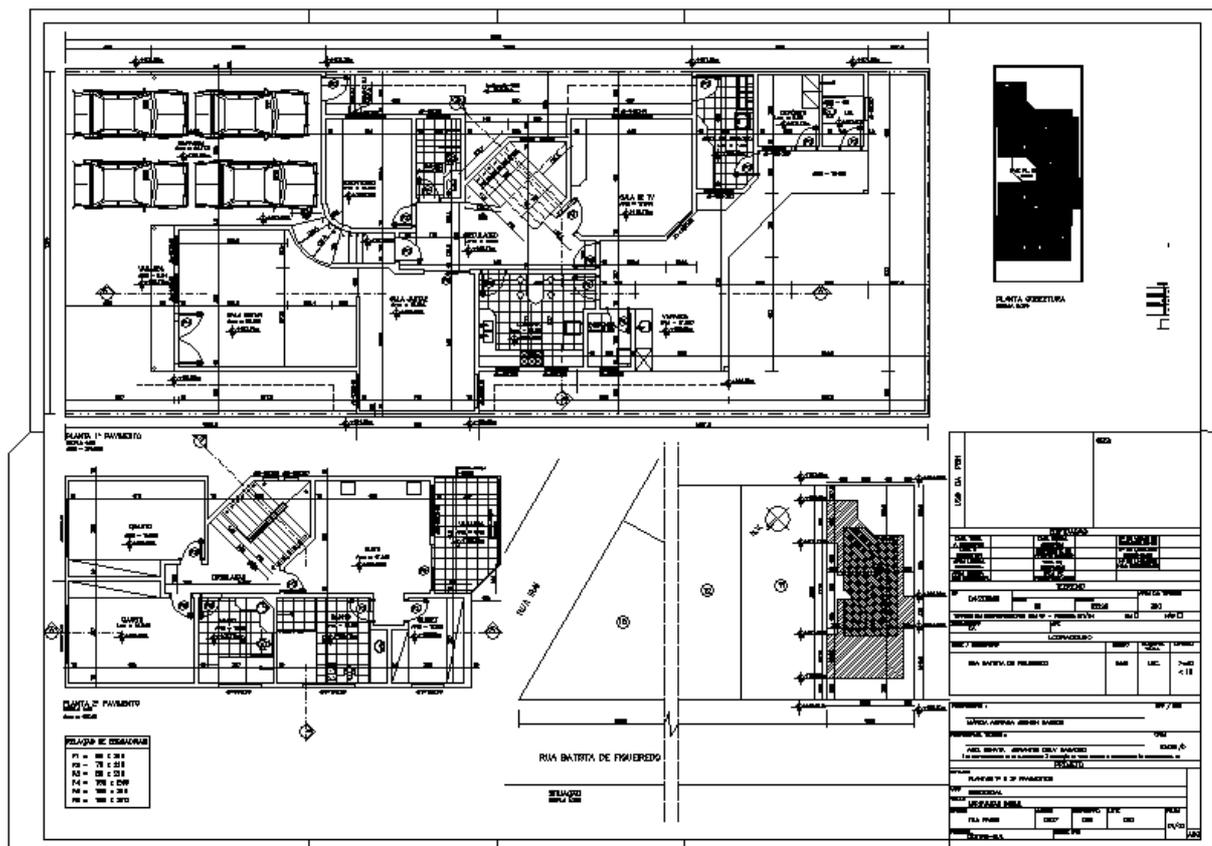


FIGURA 1 – Desenho técnico de um projeto arquitetônico - Planta.

Metadados textuais descritivos: bibliográficos

A partir da análise semântica e da definição das categorias define-se os metadados, baseados na análise sintática da imagem e seus metadados. Os metadados podem ser divididos em bibliográficos, técnicos e visuais. Os metadados visuais são definidos pela classificação do projeto.

Os metadados bibliográficos são utilizados para controle administrativo da imagem e são independentes do conteúdo da imagem. Metadados administrativos, como por exemplo, o código de classificação do desenho; nome do arquivo, data, entre outros. Metadados bibliográficos, como por exemplo, edição, autoria, editora, outros atributos de documentos conforme definido na AACR2, outros atributos de descrição física tais como tamanho da folha.

Não necessariamente estão contidos no documento, mas o descrevem, como por exemplo: extensão, data, tamanho do arquivo digital, nome do software utilizado para o desenvolvimento do projeto e sua versão. A versão do software pode impactar acesso futuros ao documento.

Metadados textuais técnicos

Os metadados técnicos fazem referências às características técnicas do projeto, como por exemplo: área, escala, local. São determinados pela classificação do desenho, podendo variar. São metadados textuais e estão contidos no desenho, geralmente no carimbo ou legenda. Normalmente, carimbo ou legenda do desenho técnico está localizado no canto inferior esquerdo, conforme FIG. 1.

Metadados visuais: assunto

Os metadados visuais são os ícones presentes na imagem. Os metadados visuais permitem a recuperação por assunto. Recupera-se a imagem pela imagem ou parte da imagem.

Uma das categorias de metadados são os metadados de assunto. O assunto de um documento pode ser expresso através de um vocabulário controlado usado para descrever sobre o que o documento versa sem utilizar o conteúdo do documento e pode também ser expresso diretamente através do conteúdo do documento. No caso de documentos que contêm texto, pode-se extrair do texto algumas palavras escolhidas para representar o assunto do documento ou pode-se usar, por exemplo, o vocabulário presente num thesaurus ou num sistema de códigos de classificação. No caso de um documento que contêm outros símbolos, como por exemplo, um desenho técnico de engenharia, pode-se considerar uma representação visual de uma porta ou de um extintor de incêndio como sendo um dos assuntos contidos no desenho. Essa representação visual de conteúdo pode ser usada para representar o assunto de um documento, não precisa ser igual à representação de porta contida no documento (ela pode ter, por exemplo, uma escala ou direção diferente da porta contida no desenho). Nesse caso, da mesma forma que um termo de um thesaurus, ela não é igual, mas representa todas as portas do documento. É, portanto, um elemento intermediário entre o dado presente no documento e o conceito que ele representa e pode ser considerado também um elemento de vocabulário controlado (FIG. 2).

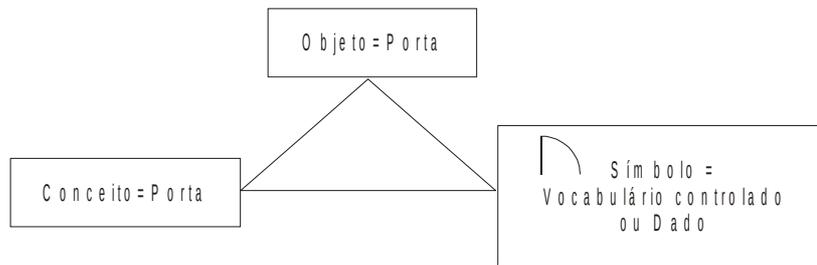
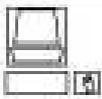


FIGURA 2 - Assunto – vocabulário controlado – dado

A tabela de atributos icônicos a ser utilizada na recuperação do desenho técnico, varia de acordo com a definição das três categorias que definem o tipo, o processo e a forma. A FIG. 3 mostra dois exemplos de metadados visuais para o desenho técnico de acordo a seguinte classificação: Primeira linha - Tipo: Projeto Arquitetônico, Processo: Executivo e Forma: Planta; Segundo linha - Tipo: Projeto Elétrico, Processo: Executivo e Forma: Planta.

Tabela de atributos visuais

Tipo: Projeto Arquitetônico Processo: Executivo Forma: Planta

 computador	 vão	 pia	 porta
---	--	--	---

Tipo: Projeto Elétrico Processo: Executivo Forma: Planta

 lâmpada	 luminária	 tomada 130	 tomada telefone 030	 quadro distribuição	 quadro medidor
--	--	---	--	--	---

FIGURA 3 - Exemplo de duas tabelas de ícones

Utilizando os recursos computacionais de recuperação automática de metadados visuais, os desenhos técnicos de engenharia possuem ícones que podem ser indexados e recuperados através da forma. Forma é um critério que permite identificar na projeção bidimensional parte da estrutura física dos objetos.

Um sistema de recuperação baseado na imagem envolve a extração de características da imagem e, sumariamente, segue os seguintes passos: relaciona características visuais da imagem baseadas em cor, textura e forma; faz a classificação das características visuais da imagem, utilizando-as para alimentar o banco de dados; e, posteriormente, a recuperação da imagem procurada. Esta última é alcançada através de comparação por similaridade. O

sistema de recuperação envolve percepção, representação, codificação e recuperação da imagem em base de dados.

A indexação está ligada à consulta. Para recuperar imagens usando-se metadados de conteúdo (cor, textura e forma), as consultas são expressas através de exemplos visuais extraídos das imagens. O sistema verifica a similaridade entre a imagem-chave usada na consulta e as imagens na base de dados.

4 EXEMPLO DE APLICAÇÃO DO SISTEMA PROPOSTO PARA DESENHOS TÉCNICOS

Através do modelo desenvolvido temos a classificação de um desenho técnico de projeto de engenharia considerando um sistema híbrido de recuperação de informação baseado nos metadados textuais e visuais.

A seguir, apresenta-se um exemplo completo do uso dos elementos do modelo em um desenho técnico. Tem-se a classificação e definição dos metadados do desenho técnico de uma planta de projeto arquitetônico utilizada como exemplo para a definição das categorias, dos atributos administrativos, técnicos e visuais. Esse desenho técnico (FIG. 2) faz parte do projeto arquitetônico de uma residência unifamiliar, aprovado na Prefeitura Municipal de Belo Horizonte em 2002. O projeto completo é composto por quatro formatos de desenhos técnicos, sendo o primeiro composto pelas plantas do primeiro e do segundo pavimento. Todo desenho técnico contém um carimbo, com os atributos técnicos, que compreendem dados relativos à edificação, terreno, logradouro, projeto, proprietário e responsável técnico. Alguns campos estão preenchidos com dados reais e outros estão em branco, propositalmente, para preservar dados particulares e por não interferirem nos testes realizados.

A FIG. 4 apresenta a classificação do projeto com a definição das categorias e metadados presentes no desenho técnico.

Metadados

Classificação do Projeto

Tipo: *Projeto Arquitetônico*
Processo: *Executivo*
Forma: *Planta*

Metadados Bibliográficos

Projeto	Arquivo
Número: <i>00051</i>	Nome: <i>prancha1</i>
Nome: <i>Residência Vila Paris</i>	Tamanho: <i>1.386 Kb</i>
Empresa responsável	Software: <i>AutoCAD</i>
Nome: <i>Targa Engenharia Tecnologia</i>	Extensão: <i>.DWG</i>
Endereço: <i>Rua Bernardo Guimarães</i>	
Data: <i>10/06/2002</i>	

Metadados Técnicos

Escala: <i>1:50</i>	Logradouro:
Edificação:	Endereço: <i>Rua Batista de Figueiredo</i>
Área total a construir: <i>328.10 m²</i>	Código: <i>8421</i>
Área a descontar:	Classificação viária: <i>loc.</i>
Área líquida residencial:	Proprietário:
Área líquida não residencial:	Nome: <i>Márcia</i>
Coefficiente de aproveitamento:	CPF/CNPJ:
Taxa de ocupação:	Responsável técnico:
Taxa de permeabilização:	Nome: <i>Renata</i>
Terreno:	CREA:
CP:	Projeto:
Setor:	Detalhe: <i>Planta 1º e 2º Pavimentos</i>
Quadra:	Uso: <i>Residencial</i>
Área:	Título: <i>Aprovação inicial</i>
Zoneamento:	Bairro: <i>Vila Paris</i>
ADE:	Código: <i>0327</i>
	Quarteirão: <i>008</i>
	Lote: <i>010</i>
	Regional: <i>Centro-Sul</i>
	Folha: <i>01/02</i>

Metadados Visuais

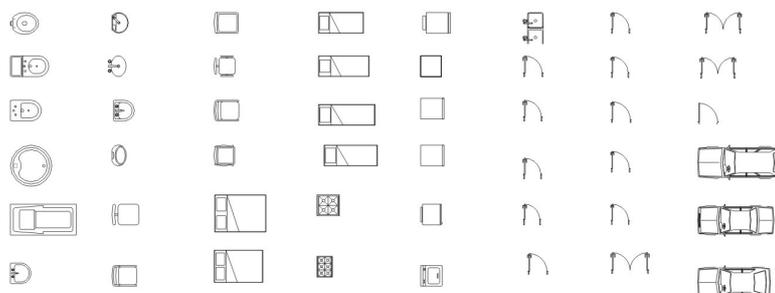


FIGURA 4 – Metadados extraídos de um desenho técnico de projeto de engenharia.

Atualmente, desenhos técnicos de projetos de engenharia, são representados com metadados textuais. É necessário um método híbrido que combina metadados textuais e visuais para melhor e mais completa representação e recuperação.

O método proposto contribui para melhorias no processo de recuperação de informação contido nos projetos de engenharia. Esse tipo de solução interfere diretamente na agilidade de consulta a informação e, auxilia processos de tomada de decisão na engenharia.

Pode ser implementado utilizando os conceitos de recuperação da informação baseada no conteúdo visual da imagem e processada automaticamente.

Uma contribuição desse sistema é a recuperação da informação em diversos contextos de projetos de engenharia. Tal sistema pode responder perguntas tais como:

- Quantos cômodos existem em uma determinada planta? Neste caso, a imagem-chave utilizada é o desenho de uma porta que representa a “query” para a recuperação da informação. A resposta imediata é o número de portas, mas a partir de inferências sobre a resposta obtida pode-se concluir quantos cômodos existem na planta. Essa pergunta pode referir-se a uma planta ou ser submetida a um conjunto de plantas, localizando e apresentando os documentos que contém um determinado ícone, em um diretório que pode conter milhares de documentos, e permitindo, portanto a gestão dos documentos;
- Quantos computadores existem em um projeto? A imagem-chave é o desenho do computador. A partir do resultado, pode-se inferir sobre os laboratórios e infra-estruturas existentes, definir número de computadores existentes em uma edificação e a necessidade de alocação de pessoal necessário para a manutenção, entre outras aplicações.

Através do estudo, pode-se concluir que é importante uma reflexão maior sobre os tipos de metadados com o objetivo de conseguir um consenso e uma padronização dos metadados de desenhos técnicos de projetos de engenharia, principalmente, considerando os metadados visuais.

A tecnologia apresenta inovações para a organização e recuperação da informação de desenhos técnicos de engenharia. As soluções precisam ser discutidas e analisadas e apresentam grandes desafios.

Cada área do conhecimento engloba seu próprio padrão de regras. A junção da Ciência da Computação e Sistemas de informação têm o grande desafio de criar ferramentas de gestão para os dados e soluções inovadoras.

REFERÊNCIAS BIBLIOGRÁFICAS

ALVARENGA, LÍDIA; **A Teoria do Conceito Revisitada em Conexão com Ontologias e Metadados no Contexto das Bibliotecas Tradicionais e Digitais**, 2005. Disponível em: http://www.dgz.org.br/dez01/Art_05.htm

ALVARENGA, L. **Representação do conhecimento na perspectiva da ciência da informação em tempos e espaços digitais**. Encontros Bibli, Florianópolis. n. 15, 2003.

ARAÚJO, A. de A. **RIBC recuperação de informação com base no conteúdo visual**. Belo Horizonte: Universidade Federal de Minas Gerais, 2003.

BARACHO, R. M. A. **Sistema de recuperação de informação visual em desenhos técnicos de engenharia e arquitetura: modelo conceitual, esquema de classificação e protótipo**. 2007. 273 f. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

BARRETO, Juliano Serra. Desafios e avanços na recuperação automática da informação audiovisual. **Revista Ciência da Informação**, Brasília: IBICT, v. 36, n. 3, p. 17-28, Set/Dez 2007.

BAX, Marcelo Peixoto. **Introdução às linguagens de marcas**. Ciência da Informação, v.30, n.1, p.32-38, jan./abr. 2001.

BENACCHIO, ALCIONE; VAZ, MARIA SALETE MARCON GOMES, **Metapadrão – Descrição e Integração de Padrões de Metadados**. Disponível em: http://www.unieuro.edu.br/downloads_2005/ruti_01_04_Metapadrao.pdf. Acesso em: 20/08/2010.

CRESWELL, J.W. **Research design**: qualitative, quantitative, and mixed methods approaches. 2. ed. Thousand Oaks: Sage Publications, 2003. 245 p.

DUBLIN CORE METADATA INITIATIVE. **Dublin Core element set**, version 1.0: reference description [online]. [S.l.], 1998^a. Disponível em: <http://purl.org/dc/documents/recdces-199809.htm>

DUBLIN CORE METADATA INITIATIVE. **Dublin Core metadata element set, version 1.1**: reference description [online]. [S.l.], 1999. Disponível em: <http://purl.org/dc/documents/rec-dces-19990702.htm>

DUBLIN CORE METADATA INITIATIVE. **A user guide for simple Dublin Core**: draft version 5.1 [online]. [S.l.], 1998b. Disponível em: <http://purl.org/dc/documents/wdguide-current.htm>

FUJITA, M. S. L. **A leitura documentária do indexador**: aspectos cognitivos e lingüísticos influentes na formação do leitor profissional. 2003. 321f. Tese (Livre-Docência em Análise Documentária e Linguagens Documentárias Alfabéticas) Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2003.

GOMES, SIONISE ET AL. **Objetos de Aprendizagem Funcionais e as Limitações dos Metadados Atuais**. Disponível em: <http://br-ie.org/pub/index.php/sbie/article/view/406/0>

HAKALA, J.; HANSEN, P.; HUSBY.O.; KOCH, T.; THORBORG, S. **The Nordic metadata project**: final report [online]. [Helsinki]: Helsinki University Library, 1998. Disponível em: <http://linnea.helsinki.fi/meta/nmfinal.htm>.

HJORLAND, B. The concept of subject in information science. **Journal of Documentation**, [S. l.], 1992.

IKEMATU, RICARDO SHOIT, **Gestão de Metadados: Sua Evolução na Tecnologia da Informação**. Disponível em: <http://www.bax.com.br/teaching/courses/bibliotecasdigitais/textos/Gestao%20de%20Metadados%20-%20Sua%20evolucao%20na%20Tecnologia%20da%20Informacao.pdf>. Acesso em: 20/08/2010.

LANCASTER, F. W.; WARNER, A. J. **Information retrieval today**. Arlington: Information Resources Press, 1993.

MARCHIORI, M. The limits of Web metadata, and beyond. **Computer Networks and ISDN Systems**, v.30, p.1-9, 1998.

MARQUES, M. L.; MANINI, M. P.; MIRANDA, A. S. S. **Ontologias: indexação e recuperação de fotografias baseadas na técnica fotográfica e no conteúdo da imagem**. In: Anais do VIII Encontro Nacional de Pesquisa em Ciência da Informação. Salvador: UFBA/ANCIB, 2007, v. 1.

MILSTEAD, J.; FELDMAN, S. Metadata: cataloging **by any other name**. **Online: the leading magazine for information professionals** [online], v.23, n.1, Jan. 1999. Disponível em: <http://www.onlineinc.com/onlinemag/OL1999/medeiros11.html>

NAVES, M.M.L. Estudo de fatores interferentes no processo de análise de assunto. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 6, n. 2, p. 189-203, jul./dez. 2001.

ROSETTO, MARCIA; NOGUEIRA, ADRIANA HYPÓLITO; **Aplicação de Elementos Metadados Dublin Core para de Descrição de Dados Bibliográficos On-Line Da Biblioteca Digital de Teses da USP**. Disponível em: <http://www.sibi.ufrj.br/snbu/snbu2002/oralpdf/82.a.pdf>. Acesso em: 20/08/2010.

SALTON, Gerard; MCGILL, Michel J. **Introduction to modern information retrieval**. New York: McGraw-Hill, 1983. 448 p.

SMITH, E.E.; Medin, D.L. **Categories and concepts**. Cambridge, Massachusetts: Harvard University Press. 1981.

SOUZA, M. I. F.; ALVES, M. das D. R.; SANTOS, A. D. dos; CINTRA, M. A. M. de U.; OLIVEIRA, M. J. de. **Guia para descrição de recurso eletrônico no Sistema Agência utilizando Dublin Core**. Campinas: Embrapa Informática Agropecuária, 2002. 44 p. (Documentos, 11). Disponível em: <http://www.cnptia.embrapa.br/modules/tinycontent3/content/2002/doc11.pdf>. Acesso em: 15 agosto 2008.