

UNIVERSIDADE FEDERAL DE GOIÁS  
INSTITUTO DE INFORMÁTICA

**Pesquisa no Instituto de Informática  
UFG**

**Pesquisas em andamento**

Goiânia  
2008

---

## Apresentação

---

Na década de 70, com a criação do Departamento de Estatística e Informática (DEI), a Universidade Federal de Goiás (UFG) despontava como instituição de referência em informática no cenário estadual. Em 1984, a primeira turma do Curso de Bacharelado em Ciência da Computação era iniciada. Doze anos depois, em 1996, surgiria do DEI o Instituto de Informática (INF). Nos anos seguintes, o INF se fortaleceu, conduzindo, além do curso de bacharelado, diversos cursos de especialização *lato sensu*. Finalmente, em 2004, outro passo significativo na história da unidade da UFG voltada para a pesquisa, ensino e extensão na área de computação seria dado: o Programa de Mestrado *Stricto Sensu* em Ciência da Computação era criado.

O presente documento contribui com esta história, ao reunir os projetos de mestrado em andamento no INF no ano de 2008. Ele foi desenvolvido pela turma de Seminários I, do segundo semestre de 2008, do curso de Mestrado em Ciência da Computação da UFG.

Estamos certos de que esta será uma ferramenta indispensável para a divulgação dos trabalhos de mestrado do INF, permitindo a troca de experiências entre estudantes e professores da UFG e de outras universidades. Além disso, propiciará modelos para a correta atuação dos mestrandos goianos e oportunizará a discussão de variados assuntos.

Neste ano, tivemos 19 trabalhos divulgados, orientados por 14 professores e divididos em 24 projetos. Estes projetos e trabalhos estão distribuídos em quatro linhas de pesquisa: Algoritmos e Grafos, Otimização, Redes e Sistemas Distribuídos, e Sistemas de Informação.

Os professores vinculados ao programa de mestrado são: Ana Paula Laboissière Ambrósio, Auri Marcelo Rizzo Vincenzi, Cedric Luiz de Carvalho, Cláudio Nogueira de Meneses, Diane Castonguay, Eduardo Simões de Albuquerque, Fábio Moreira Costa, Fábio Nogueira de Lucena, Hugo Alexandre Dantas do Nascimento, Humberto José Longo, João Carlos da Silva, Juliano Lopes de Oliveira, Luis Román Lucambio Pérez, Orizon Pereira Ferreira, Plínio de Sá Leitão Júnior, Rommel Melgaço Barbosa, Thierson Couto Rosa, Vagner José do Sacramento Rodrigues e Wellington Santos Martins.

Portanto, este compêndio soma-se ao esforço de melhorar o Curso *Stricto Sensu* de Mestrado em Ciência da Computação da Universidade Federal de Goiás e preparar os

alunos para os desafios que ainda virão.

Agradecemos aos professores e alunos pelo imprescindível apoio.

---

## Nota do Professor

---

À turma de Seminários I do Programa de Mestrado de Ciências da Computação do segundo semestre de 2008, foi dada a tarefa de fazer um compêndio da pesquisa sendo desenvolvida no Instituto de Informática da Universidade Federal de Goiás (INF/UFG). Esta tarefa, aparentemente fácil, mostrou-se um desafio já que esta informação encontrava-se dispersa e em diferentes formatos. Coube aos alunos reunir este material e torná-lo o mais uniforme possível, gerando um documento coeso. Os alunos responderam ao desafio com seriedade, motivação e criatividade.

Apesar de não ser um documento completo, pela própria natureza da pesquisa que está sempre evoluindo e se desenvolvendo, acredito que este documento representa de forma fiel as linhas de pesquisa e os temas que têm norteado a pesquisa no INF/UFG até o momento. Acredito que ao reunir em um único documento esta informação, fique mais fácil divulgar, para a sociedade e para o meio acadêmico, o trabalho sendo desenvolvido.

Espero que este compêndio seja o primeiro de vários que o seguirão, transformando-se em uma referência para quem quer conhecer a pesquisa do INF/UFG, mantendo um registro histórico do trabalho desenvolvido.

Agradeço aos alunos pelo excelente trabalho e desejo a todos muito sucesso nos seus trabalhos de mestrado e na sua vida profissional.

Goiânia, 19 de dezembro de 2008.

Ana Paula Laboissière Ambrósio  
Professora da Disciplina

**Alunos da Disciplina:** Alexandre Cláudio de Almeida, Bruno Ferreira Machado, Elisângela Silva Dias, Fabiana Freitas Mendes, Glauber Boff, Halley Wesley Gondim, Marcio Pereira de Sá, Patrícia Gomes Fernandes, Ricardo Belloti, Victor Ribeiro Silva, Wilane Carlos da Silva.

---

# Sumário

---

<b>1</b>	<b>Algoritmos e Grafos</b>	<b>6</b>
1.1	Álgebras Triangulares de Dimensão Finita Livremente Conexas	7
1.2	Apresentação Máxima de Álgebra	7
1.3	Some Examples of Weakly Shod Algebras	7
1.4	Produtos de Grafos e Grafos Bem-Cobertos	8
1.5	Algoritmos na Teoria de Representações de Álgebra	8
1.6	Álgebras Cluster sem Coeficientes	8
1.7	Algoritmos e Grafos	9
1.7.1	Conjuntos k-Dominantes Eficientes em Grafos	11
1.8	Conjuntos Independentes de Vértices e Grafos Equi-Emparelháveis	13
	<b>Referências Bibliográficas</b>	<b>14</b>
<b>2</b>	<b>Otimização</b>	<b>16</b>
2.1	Otimização Interativa Aplicada a Problemas de Roteamento de Veículos	16
2.1.1	Otimização da Geração de Rotas em Algoritmos BCP para o VRP.	17
	<b>Referências Bibliográficas</b>	<b>18</b>
<b>3</b>	<b>Redes e Sistemas Distribuídos</b>	<b>19</b>
3.1	Computação Sensível ao Contexto	20
3.1.1	LoCoS (Local Context Service)	20
3.2	Identificação por Rádio-freqüência (RFID)	21
3.2.1	Uso de Hadoop para Distribuição de Armazenamento e Processamento em Sistemas de Dados Massivos de Controle de Redes de Sensores Diversos	21
3.2.2	Gerenciador de Dados para uma rede RFID no escopo do projeto SINIAV	22
3.3	Serviços Baseados em Localização (LBS)	24
3.3.1	LBS Shopping	24
3.4	InteGrade3: Gerenciamento de Recursos, Tolerância a Falhas e Escalabilidade em Grades Computacionais Autônomas	25
3.4.1	BIOFOCO III: Software para análise genômica em ambiente cooperativo e distribuído na Região CO	25
3.5	Processamento Paralelo Aplicado à Bioinformática	26
3.6	Paracoccidioides brasiliensis: Abordagens enfocando perfis de expressão gênica e moléculas recombinantes no estudo de interações patógeno-hospedeiro: Aplicações biotecnológicas	26
	<b>Referências Bibliográficas</b>	<b>27</b>

4	Sistemas de Informação	28
4.1	Ensino a Distância	28
4.1.1	Uma Ferramenta de Hipertexto para Apoio à Aprendizagem Colaborativa Apoiada por Computador	29
4.2	Metodologias de Banco de Dados Aplicadas à Recuperação de Informações na WEB	32
4.3	Uso de Tablet PCs e do Método de Aprendizado Baseado em Problemas no Ensino de Algoritmos e Programação	35
4.3.1	Esboço de Fluxogramas no Ensino de Algoritmos	36
4.4	Recuperação de Informação e Extração de Conhecimento em Ambientes Distribuídos e Heterogêneos	38
4.4.1	Extração de Características de Documentos Textuais para Melhoria da Eficácia na Classificação de Textos	38
4.4.2	Grades Semânticas em Comunidades Virtuais	41
4.4.3	Integração de Repositórios Digitais com a Utilização de Especificações Semânticas	43
4.5	Melhoria de Processos de Software em Goiás	44
4.5.1	Uma Abordagem MDA para Geração e Evolução Automática de Esquemas de Bancos de Dados	44
4.5.2	Geração e Manutenção de Interfaces Gráficas com Usuário Baseada em Metamodelos	46
4.5.3	Modelagem e Implementação de Restrições de Integridade Utilizando OCL com Mapeamento para SQL	47
4.5.4	Melhoria de Processos de Governança de TI	49
4.5.5	As Influências e a Gestão da Cultura Organizacional na Melhoria de Processos	51
4.6	Metodologias de Integração de Informações Armazenadas em Bancos de Dados / Web	53
4.6.1	Utilização de Mapas Mentais na Engenharia de Requisitos	53
4.7	Teste de Software	54
4.7.1	Teste Estrutural de Software em Dispositivos Móveis Utilizando J2ME	54
4.8	Identificação de Relacionamento entre Objetos de Bibliotecas Digitais	55
4.9	Resolução de Ambigüidade de Nomes em Bibliotecas Digitais	56
4.10	Propostas de Medidas de Relevância de Conferências	57
	Referências Bibliográficas	57

---

## Algoritmos e Grafos

---

Esta linha de pesquisa compreende o estudo teórico e aplicações de grafos, o desenvolvimento de algoritmos eficientes para a resolução de problemas e a análise teórica de algoritmos.

Os algoritmos e a modelagem de problemas através de grafos são elementos essenciais em qualquer área aplicada da computação.

A análise de algoritmos constitui uma ferramenta fundamental para o profissional de ciência da computação, visto que através dela podemos estudar quando um problema de decisão ou otimização admite algoritmos eficientes (com baixa complexidade), ou quando "possivelmente" não admite (alta complexidade). Os membros desta área desenvolveram e vêm desenvolvendo diversos resultados sobre a complexidade de problemas de decisão e de otimização, seja na tentativa de exibir um algoritmo polinomial ou um certificado para sua dificuldade, que pode ser uma redução de NP-completude, para um problema de decisão. Quando nos deparamos com um problema combinatório difícil, tentamos também projetar e exibir algoritmos polinomiais de aproximação.

Este capítulo apresenta a pesquisa relacionada à linha de pesquisa de Algoritmos e Grafos. Assim, oito projetos foram descritos. Na Seção 1.1 é descrito o projeto de *Álgebras Triangulares de Dimensão Finita Livrementemente Conexas*. Na Seção 1.2 é descrito o projeto denominado *Apresentação Máxima de Álgebra*. Já na Seção 1.3 o projeto *Some Examples of Weakly Shod Algebras* é descrito. Na Seção 1.4, por sua vez, é feita uma descrição do projeto *Produtos de Grafos e Grafos Bem-Cobertos*. A Seção 1.5 apresenta a descrição do projeto *Algoritmos na Teoria de Representações de Álgebra*. A Seção 1.6 traz a descrição do projeto *Álgebras Cluster sem Coeficientes*. Já na Seção 1.7 o projeto *Algoritmos e Grafos* é apresentado. Por fim, na Seção 1.8 o projeto *Conjuntos independentes de vértices e Grafos Equi-emparelháveis* é descrito.

## 1.1 Álgebras Triangulares de Dimensão Finita Livrementemente Conexas

**Coordenadora:** Dra Diane Castonguay - [diane@inf.ufg.br](mailto:diane@inf.ufg.br)

**Colaboradores:**

E. N. Marcos, Universidade de São Paulo, Brasil

C. Novoa, Universidade Católica de Goiás, Brasil

S. Trepode, Universidad Nacional de Mar del Plata, Argentina

As álgebras livremente conexas são uma generalização das álgebras simplesmente conexas. As álgebras de tipo de representação finito (nas quais o número de classes de isomorfismos de módulos indecomponíveis é finito) são álgebras livremente conexas. Este projeto visa desenvolver algumas classes de álgebras livremente conexas.

## 1.2 Apresentação Máxima de Álgebra

**Coordenadora:** Dra Diane Castonguay - [diane@inf.ufg.br](mailto:diane@inf.ufg.br)

**Colaboradores:**

I. Assem, Université de Sherbrooke, Canada

E. N. Marcos, Universidade de São Paulo, Brasil

C. Novoa, Universidade Católica de Goiás, Brasil

S. Trepode, Universidad Nacional de Mar del Plata, Argentina

Seja  $A$  uma álgebra triangular. Dizemos que uma apresentação  $(Q, I)$  de  $A$  é máxima se para qualquer outra apresentação  $(Q, J)$  de  $A$  existe um morfismo sobrejetor do grupo fundamental de  $(Q, I)$  para o grupo fundamental de  $(Q, J)$ . Por conseguinte, se  $A$  possui uma apresentação máxima, temos que  $A$  é simplesmente conexa se e somente se o grupo fundamental de uma apresentação máxima é trivial. Se  $A$  é triangular e sem *doble bypass*, então existe uma apresentação máxima de  $A$  [10].

Os objetivos de nossos trabalhos são de ampliar a classe das álgebras que possuem uma apresentação máxima e, quando for possível, exibir um algoritmo permitindo calcular tal apresentação e seu grupo fundamental. Pretendemos estender este estudo às álgebras que possuem um conjunto finito completo de apresentação maximais.

## 1.3 Some Examples of Weakly Shod Algebras

**Coordenadora:** Dra Diane Castonguay - [diane@inf.ufg.br](mailto:diane@inf.ufg.br)



**Colaboradores:**

F. Huard, Bishop's University, Canada

M. Lanzilotta, Universidad de la República Oriental del Uruguay

As álgebras fracamente shod [5] formam uma classe ampla de álgebra generalizando as álgebras hereditárias, inclinadas, quase-inclinadas e as shod. Neste projeto, queremos caracterizar quais das álgebras, cujo quiver ordinário é um pião, é fracamente shod. As álgebras canônicas são alguns dos exemplos desta classe de álgebras.

## 1.4 Produtos de Grafos e Grafos Bem-Cobertos

**Coordenadora:** Dra Diane Castonguay - diane@inf.ufg.br

Objetivos:

1. Conseguir algumas propriedades para problemas de dominação, independência e coloração envolvendo alguns produtos de grafos, principalmente o cartesiano e o lexicográfico.

2. Estudar a decomposição do produto cartesiano de grafos bem-cobertos. Neste processo, usaremos algoritmos conhecidos de decomposição do produto cartesiano de grafos.

## 1.5 Algoritmos na Teoria de Representações de Álgebra

**Coordenadora:** Dra Diane Castonguay - diane@inf.ufg.br

O principal objetivo deste projeto é estudar os algoritmos existentes proveniente da Teoria de Representações de Álgebra e providenciar versões mais eficientes ou com melhor implementação quando for possível. A maioria destes algoritmos serão proveniente do CREP (Combinatorial REPresentation theory), encontrados em [11]. A partir destes conhecimentos, poderemos desenvolver algoritmos para problemas ainda não abordados deste ponto de vista. Dentro deste estudo, queremos caracterizar alguns problemas da Teoria de Representações de Álgebra que estejam na classe dos problemas NP (ou co-NP) e analisar quais deles são NP-completos (co-NP-completos, respectivamente).

## 1.6 Álgebras Cluster sem Coeficientes

**Coordenadora:** Dra Diane Castonguay - diane@inf.ufg.br

As álgebras cluster, introduzidas por S. Fomin e A. Zelevinsky no ano 2000, formam uma classe de álgebras baseadas no estudo de bases canônicas duales e a positividade de grupos semi-simples. Estas álgebras aparecem como objeto de estudo em várias áreas como geometria, combinatória, física e matemática entre outras. Essencialmente, uma álgebra cluster pode ser descrita da seguinte forma: Seja um inteiro  $n$ , então uma álgebra cluster de dimensão  $n$  é um anel comutativo (unitário) sem divisores de zero, gerado no centro de um corpo prefixado  $F$  por um conjunto, eventualmente infinito, de variáveis cluster. Estas variáveis não são fixadas arbitrariamente. O conjunto das variáveis cluster é a união (não disjunta) de subconjuntos de  $n$ -elementos chamados de cluster, que são relacionados pela relação de mudança.

Para qualquer cluster  $X$  e qualquer variável cluster  $x$  em  $X$ , existe outro cluster obtido de  $X$  substituindo a variável  $x$  por outra  $x'$  relacionada por uma relação binomial, dita de mudança, da seguinte forma:  $x x' = M_1 M_2$ , onde  $M_1$  e  $M_2$  satisfazem condições bem específicas. Todas as variáveis cluster são obtidas recursivamente desta forma a partir de uma semente inicial (dada por um cluster e uma regra de mudança). Este processo é chamado de mutação da semente.

Estudaremos as álgebras cluster, sem coeficientes, obtidas a partir de semente cuja regra de mudança seja dada via um quiver, sem laços, nem ciclos, de comprimento dois (equivalente a uma matriz anti-simétrica).

## 1.7 Algoritmos e Grafos

**Coordenador:** Dr Rommel Melgaço Barbosa - rommel@inf.ufg.br

No ano de 2002, assistiu-se a prova da conjectura forte dos grafos perfeitos por [Chudnowsky et al]. Isto tem uma série de implicações no direcionamento da pesquisa em grafos perfeitos. Nosso interesse é pelo problema de coloração.

Este problema, apesar de difícil, em geral, admite soluções eficazes para certas classes de grafos perfeitos, através do algoritmo dos elipsóides e, portanto, não combinatório. Trabalhamos considerando suas fortes propriedades combinatoriais. Foram resolvidas para classes especiais de grafos (grafos planares, grafos sem garras e grafos sem touros) através de trabalhos conjuntos envolvendo vários pesquisadores, pelos membros deste projeto como colaboração internacional.

Quando uma família considerada é de conjuntos arbitrários, a classe de grafos obtida como grafos de interseção desta família é, simplesmente, a classe de todos os grafos.

Pretendemos atuar em 5 linhas gerais de pesquisa: Hipergrafos e Propriedade de Helly, Grafos de Interseção, Grafos Perfeitos e Partições em Grafos, Algoritmos e

Complexidade, e Conjuntos Independentes. Descrevemos a seguir o referencial teórico associado de cada uma delas.

1 - *Hipergrafos e Propriedade de Helly*: Um hipergrafo é um par  $H=(V,E)$  onde  $V$  é um conjunto de vértices e  $E$  é um conjunto de hiperarestas (subconjuntos não vazios de  $E$ ). Em outras palavras, um hipergrafo é uma subfamília do conjunto das partes de  $V$ . Pretendemos estudar diversas classes de hipergrafos, entre elas as relacionadas a propriedade de Helly. Nesses estudos, faremos uma abordagem do ponto de vista algorítmico e de complexidade.

2 - *Grafos de Interseção*: O problema de caracterizar os grafos de interseção de famílias tendo alguma topologia específica é interessante do ponto de vista teórico e tem aplicações no mundo real. Várias classes de grafos de interseção são definidas considerando-se famílias de subárvores de uma árvore em especial, de caminhos em uma árvore. Consideramos interseções tanto em arestas quanto em vértices. Assim, surgem as classes de grafos Cordais, UV, DV, RDV, UE, DE e RDE. Outras classes de grafos de interseção que temos interesse atualmente são os grafos clique, linha, loop e disco unitário.

O objetivo principal é identificar classes de grafos de interseção, estabelecendo caracterizações, principalmente por subgrafos proibidos e estudar seu reconhecimento. Consideramos também os problemas de otimização restritos a estas classes, em especial, aos problemas de coloração de vértices, cobertura por cliques, conjunto independente, entre outros.

3 - *Grafos Perfeitos e Partições em Grafos*: Apesar de o ano de 2002, ter assistido a prova da Conjectura Forte dos Grafos Perfeitos, muitos problemas permaneceram em aberto, sendo a questão do reconhecimento um deles. Vamos considerar diversas estruturas especiais, tais como conjunto homogêneo e pares homogêneos, importantes quando se considera a questão de decomposição e composição de grafos perfeitos. Ainda na linha de estruturas especiais de grafos, em geral, e em particular na classe dos grafos cordais, vamos estudar partições de grafos em partes que podem ser cliques ou conjuntos independentes.

4 - *Algoritmos e Complexidade*: Observamos que praticamente todas as subáreas de pesquisa em Ciência da Computação possuem uma relação direta ou indireta com a área de Algoritmos e Complexidade.

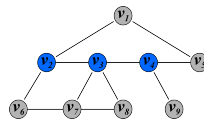
5 - *Conjuntos Independentes em Grafos*: Dentre os problemas relacionados, destacamos o problema da caracterização de grafos bem-cobertos para algumas classes de grafos. Um grafo é bem-coberto quando todos seus conjuntos independentes maximais tiverem a mesma cardinalidade. O problema de reconhecimento de grafos bem-cobertos é Co-NP-completo.

### 1.7.1 Conjuntos k-Dominantes Eficientes em Grafos

**Discente:** Rommel Teodoro de Oliveira - oliveirart@gmail.com

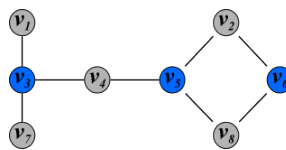
**Nível:** Mestrado

Um conjunto  $D \subseteq V$  é denominado *conjunto dominante* de  $G$  se para todo vértice  $v \in V$  a condição  $|N[v] \cap D| \geq 1$  for satisfeita. Se um vértice  $v$  pertence ao conjunto dominante  $D$ , então  $v$  é chamado de *vértice dominante*. Por outro lado, um vértice estará *dominado* quando possuir um vizinho dominante. No grafo ilustrado na Figura 1.1, o conjunto  $D = \{v_2, v_3, v_4\}$  é um conjunto dominante e seus vértices estão coloridos com a cor azul.



**Figura 1.1:** Grafo que possui conjuntos dominantes.

Diz-se que  $D$  é um *conjunto dominante minimal* se não existir um subconjunto próprio  $D' \subset D$  tal que  $D'$  também seja um conjunto dominante. O tamanho do menor conjunto dominante minimal de  $G$  é denominado *número de dominação* de  $G$ , sendo denotado por  $\gamma(G)$ . Similarmente,  $\Gamma(G)$  corresponde ao tamanho do maior conjunto dominante minimal de  $G$ . Para o grafo  $G$  da Figura 1.2, pode-se citar como exemplo de conjuntos dominantes minimais:  $\{v_3, v_5, v_6\}$ ,  $\{v_1, v_5, v_7, v_8\}$  e  $\{v_1, v_2, v_4, v_7, v_8\}$ . Existem outros conjuntos dominantes minimais neste grafo, entretanto, observe que é impossível obter um conjunto dominante minimal com cardinalidade menor que 3 ou maior que 5. Logo,  $\gamma(G) = 3$  e  $\Gamma(G) = 5$ .



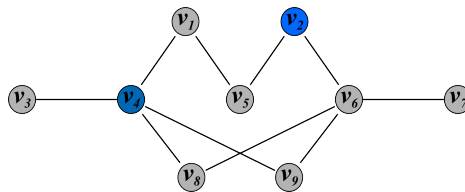
**Figura 1.2:** Grafo que possui diversos conjuntos dominantes minimais, sendo que  $\gamma(G) = 3$  e  $\Gamma(G) = 5$ .

Um problema de decisão envolvendo tais conceitos consiste em decidir se um grafo  $G$  possui um conjunto dominante de cardinalidade menor ou igual a um inteiro  $k$ . [6] demonstraram que este problema é  $\mathcal{NP}$ -Completo para grafos gerais através de uma redução do Problema de Cobertura de Vértices. Por outro lado, [4] apresentaram um algoritmo linear que determina o número de dominação  $\gamma(A)$  para uma árvore  $A$  utilizando um método de rotulação de vértices.

Problemas de conjuntos dominantes em grafos surgem em aplicações de diversas áreas de pesquisas, tais como na computação paralela [9] e redes de computadores

[2]. Em redes de computadores, por exemplo, tal estudo está relacionado ao processo de construção da infra-estrutura uma rede ad hoc sem fio, dentre outras situações. Já em computação paralela, a existência de conjuntos dominantes é útil na construção de algoritmos eficientes para distribuição de recursos entre os processadores que compõem este tipo de arquitetura.

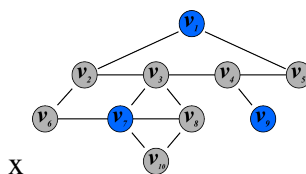
Devido aos diferentes requisitos exigidos por diversas aplicações práticas, muitas variações do problema de dominação têm sido apresentadas na literatura. Por exemplo, um conjunto dominante  $D \subseteq V$  é um *conjunto dominante eficiente* de  $G$  se para todo vértice  $v \in V$  a condição  $|N[v] \cap D| = 1$  for satisfeita. Muitos grafos não possuem conjuntos dominantes eficientes, como o ciclo  $C_5$ . Nestes casos, busca-se determinar a maior quantidade de vértices que podem ser dominados exatamente uma vez, sendo este valor denominado *número de dominação eficiente* e denotado por  $F(G)$  para um grafo  $G$  qualquer. Se  $F(G) = n$ , então  $G$  é eficientemente dominado. Para o grafo  $G$  de ordem 9 da Figura 1.3, tem-se  $F(G) = 8$ .



**Figura 1.3:** Grafo em que  $F(G) = 8$ .

Dado um grafo arbitrário  $G$ , determinar se  $G$  possui um conjunto dominante eficiente é um problema  $\mathcal{NP}$ -Completo [1]. Se  $G$  for uma árvore  $A$ , então existe um algoritmo linear que computa  $F(A)$ , conforme demonstrado por [1].

Neste trabalho, é considerado o conceito da *k-dominância eficiente*: na situação em que para todo  $v \in V$  a condição  $|N[v] \cap D| = k$  for satisfeita, tem-se  $D$  como um conjunto *k-dominante eficiente*. O tamanho do menor conjunto *k-dominante eficiente* é denotado por  $\gamma(G)_{e \times k}$ . No grafo da Figura 1.1, por exemplo, o conjunto  $D = \{v_1, v_7, v_9\}$  é 1-dominante eficiente, ou simplesmente dominante eficiente. Para este grafo, é impossível obter outro conjunto dominante eficiente. Por consequência,  $\gamma(G)_{e \times 1} = 3$ .



**Figura 1.4:** Grafo que possui um conjunto 1-dominante eficiente.

Os objetivos deste trabalho são: i) Demonstrar a complexidade do problema; ii) Caracterizar classes de grafos para os quais é possível determinar  $\gamma(G)_{e \times k}$  através de um algoritmo eficiente; iii) Investigar os valores limites para  $\gamma(G)_{e \times k}$ .

Torna-se necessário investigar a complexidade do problema para classes específicas de grafos.

Neste trabalho, o enfoque da pesquisa está relacionado aos conjuntos  $k$ -dominantes eficientes, para os quais muito pouco se conhece na literatura. Já foram obtidos alguns resultados até o momento. Como se trata de um trabalho em andamento, outras conjecturas propostas estão sendo respondidas para que se estabeleça uma valiosa contribuição em relação ao assunto.

## 1.8 Conjuntos Independentes de Vértices e Grafos Equi-Emparelháveis

**Coordenador:** Dr Rommel Melgaço Barbosa - rommel@inf.ufg.br

Um clássico problema da Teoria dos Grafos com várias aplicações em diversas áreas é a determinação do número de independência de um grafo. Este problema é NP-Completo para grafos em geral, mas é polinomial para grafos bem-cobertos, pois neste caso, basta aplicarmos um algoritmo guloso que encontra um conjunto independente maximal qualquer, visto que todos têm a mesma cardinalidade nestes grafos.

O reconhecimento de grafos bem-cobertos é Co-NP-Completo para grafos em geral e polinomial para grafos livres de  $K_{1,3}$ .

O problema de grafos equi-emparelháveis foi proposto em 1974 por B. Grünbaum [7]. Estudos foram feitos desde então e em 1984 Lesk, Plummer e Pulleyblank [8] apresentaram um algoritmo polinomial que reconhece se um grafo pertence à classe dos grafos equi-emparelháveis.

Embora o reconhecimento de grafos bem-cobertos livres de  $K_{1,3}$  e de grafos equi-emparelháveis seja polinomial, não se conhece uma caracterização destes grafos. Conseguir uma possível caracterização destes grafos ou famílias de grafos nesta classe poderá contribuir para a resolução de problemas existentes bem como futuros problemas em Teoria dos Grafos.

Um conjunto independente de vértices em um grafo  $G = (V, E)$  é um conjunto de vértices  $S \subseteq V(G)$  tal que  $\forall u, v \in S$ , a aresta  $(u, v) \notin E(G)$ . Um conjunto independente é maximal se não é subconjunto próprio de outro conjunto independente e é máximo se tem cardinalidade máxima. A cardinalidade de um conjunto independente máximo em um grafo  $G$  é o número de independência de  $G$  e é denotado por  $\alpha(G)$ . Um grafo  $G$  é bem-coberto, conceito introduzido por Plummer em 1970, se todo conjunto independente maximal de vértices em  $G$  tem a mesma cardinalidade. Um grafo  $G$  é  $Z_m$ -bem-coberto se  $[I] \equiv [J] \pmod{m}$ ,  $\forall I, J$  conjuntos independentes maximais de vértices em  $G$ . Um

grafo  $G$  pertence a classe  $M(t)$  se  $G$  tiver exatamente  $t$  tamanhos diferentes de conjuntos independentes maximais, e a  $I(t)$  se pertencer a  $M(t)$  e estes tamanhos forem números consecutivos.

Um emparelhamento em um grafo conexo  $G = (V, E)$  é um conjunto de arestas  $M \subseteq E(G)$  tal que quaisquer duas arestas não compartilham um vértice. Um emparelhamento é *maximal* se não está contido propriamente em outro emparelhamento e é *máximo* se tem cardinalidade máxima. Um grafo  $G$  é *equi-emparelhável* se todo emparelhamento maximal de arestas em  $G$  é também máximo.

Uma *orientação transitiva* de um grafo  $G$  é uma orientação  $D$  tal que quando  $xy$  e  $yz$  são arestas em  $D$ , também há uma aresta  $xz$  em  $G$  que é orientada de  $x$  para  $z$  em  $D$ . Um grafo simples  $G$  é um *grafo de comparabilidade* se ele tem uma orientação transitiva.

O problema de determinação do número de independência de um grafo é um problema NP-Completo para grafos em geral. Para grafos bem-cobertos este problema torna-se mais simples pois é necessário apenas encontrar qualquer conjunto independente maximal, visto que todos têm a mesma cardinalidade. Mas, verificar se um grafo é bem-coberto é um problema Co-NP-Completo para grafos em geral, como provado independentemente por Chvátal e Slater e em Sankaranarayana e Stewart. Caro, Sebo e Tarsi em [3] provaram que mesmo quando a entrada está restrita a grafos que não contêm um subgrafo induzido isomorfo a  $K_{1,4}$ , o problema permanece Co-NP-Completo.

Embora o problema de reconhecimento de grafos equi-emparelháveis seja polinomial, enquanto o problema de reconhecimento de grafos bem-cobertos é um problema Co-NP-Completo, poucas caracterizações de famílias de grafos equi-emparelháveis são conhecidas. Dentre elas temos os grafos cúbicos equi-emparelháveis, os grafos 3-conexos planares equi-emparelháveis, os grafos bipartidos equi-emparelháveis [8] e os grafos fatores-crítico equi-emparelháveis [8].

A pesquisa tem como principais objetivos:

- Encontrar famílias de grafos bem-cobertos livres de  $K_{1,3}$ ;
- Encontrar famílias de grafos equi-emparelháveis;
- Encontrar propriedades para grafos de comparabilidade bem-cobertos;
- Conseguir propriedades para grafos bipartidos em  $M(t)$  e  $Z_m$ -bem-cobertos.

## Referências Bibliográficas

- [1] BANGE, D. W; BARKAUSKAS, A. E; SLATER, P. J. **Efficient dominating sets in graphs**. Applications of Discrete Math., 1:189–199, 1988.

- [2] BLUM, J; DING, M; THAELER, A; CHENG, X. **Handbook of Combinatorial Optimization**, chapter Connected Dominating Set in Sensor Networks and MANETs, p. 329–369. Kluwer Academic Publishers, 2004.
- [3] CARO, Y; TARSI, M. **Recognizing greedy structures**. *Journal of Algorithms*, 20:137–156, 1996.
- [4] COCKAYNE, E. J; GOODMAN, S; HEDETNIEMI, S. T. **A linear algorithm for the domination number of a tree**. *Information Processing Letters*, 4:41–44, 1975.
- [5] COELHO, F. U; LANZILOTTA, M. A. **Weakly shod algebras**. *J. Algebra*, 1(265):379–403, 2003.
- [6] GAREY, M. R; JOHNSON, D. S. **Computers and Intractability: A Guide to the Theory of NP-Completeness**. Freeman, New York, 1979.
- [7] GRUNBAUM, B. **Matchings in polytopal graphs**. *Networks*, 04:175–190, 1974.
- [8] LESK, M; PLUMMER, M; PULLEYBLANK, W. **Equimatchable graphs**. *Graphs Theory and combinatorics*, New York, 1984.
- [9] LIVINGSTON, M; STOUT, Q. F. **Perfect dominating sets**. *Congressus Numerantium*, 79:187–203, 1990.
- [10] MEUR, P. L. **The fundamental group of a triangular algebra without double bypasses**. preprint 2005, New York, 1984.
- [11] OF BIELEFELD, U. **Combinatorial Representation Theory**. <http://www.mathematik.uni-bielefeld.de/~sek/crep.html>, ?ltimo acesso em Agosto de 2007, 2007.



## Otimização

---

Esta linha visa investigar problemas teóricos e computacionais nas áreas de otimização combinatória e contínua, e o desenvolvimento de métodos para solução destes problemas.

Este capítulo apresenta a pesquisa relacionada à linha de Otimização. O único projeto associado, denominado **Otimização Interativa Aplicada a Problemas de Roteamento de Veículos**, é apresentado na Seção [2.1](#).

### 2.1 Otimização Interativa Aplicada a Problemas de Roteamento de Veículos

**Coordenador:** Dr Humberto José Longo - longo@inf.ufg.br

A expressão Roteamento de Veículos está relacionada a um grande conjunto de problemas de fundamental importância para a área de logística de transportes, em especial no que diz respeito ao uso racional de frotas de veículos, seja para o transporte de passageiros ou de bens e mercadorias.

Um dos principais objetivos de tais problemas é a determinação de rotas para os veículos de uma frota. Este objetivo, em geral, é complexo e difícil de ser alcançado, devido às restrições operacionais que dependem de fatores como, por exemplo, a natureza e características dos itens a serem transportados ou da qualidade com que o serviço deve ser executado. Além disso, os custos envolvidos costumam ser elevados e sensíveis às diferentes variáveis associadas ao problema.

Em geral os problemas de roteamento de veículos referem-se à determinação de rotas para os veículos de uma frota, visando o atendimento de um conjunto de clientes com demandas (passageiros ou mercadorias) a serem transportadas. Considera-se que os veículos operam a partir de um ou mais depósitos base e têm que iniciar e encerrar suas rotas (seqüências de visitas aos clientes) no depósito que é a sua base. O objetivo principal

é a minimização do custo total do transporte, do tempo total a ser gasto ou da distância total a ser percorrida nas rotas.

O número de variáveis a ser considerado no planejamento de rotas para frotas de veículos pode ser bastante elevado. Dentre essas variáveis destacam-se: o número de depósitos onde estão baseados os veículos, o tamanho da frota de veículos, o tipo da frota disponível (homogênea-heterogênea), a capacidade dos veículos, a natureza da demanda, a localização do cliente, o tipo de malha viária (composta de ruas-estradas e pontos de conexão) associada ao problema, os custos envolvidos, o tipo de operação a ser realizada (coleta-entrega), horário de atendimento de um determinado cliente e requisitos de pessoal (número de motoristas, duração da jornada de trabalho, etc.). Além disso, pode ser necessário considerar restrições com respeito ao tempo (limites superiores na duração de uma rota), à quantidade-peso da demanda ou quanto ao horário que o cliente deve ser atendido. Podem existir também restrições adicionais como, por exemplo, uma que restrinja o atendimento de um cliente em particular por um determinado veículo.

Esses problemas podem, ainda, ser classificados em duas classes, dependendo da localização dos clientes na rede viária associada ao problema. Na primeira classe, que agrega a maioria das aplicações mais tradicionais, o atendimento à demanda de um cliente implica na visita de um local específico da malha viária. Na segunda classe estão os problemas em que o atendimento de um cliente implica em percorrer um trecho completo (entre dois pontos de conexão) da malha viária. Um exemplo clássico de problema da primeira classe é o caso em que o objetivo é a construção de uma rota de percurso mínimo que passe exatamente uma vez por todos os pontos de conexão da rede viária associada ao problema. Este problema, conhecido como problema do caixeiro viajante ou TSP (Traveling Salesman Problem), pode ser visto como um problema de roteamento de veículos em que existe apenas um depósito, um veículo de capacidade ilimitada e demandas unitárias associadas aos pontos de conexão da rede viária.

### **2.1.1 Otimização da Geração de Rotas em Algoritmos BCP para o VRP.**

**Discente:** Diego Galindo Pecin - diegopecin@gmail.com

**Nível:** Mestrado

O projeto visa melhorar as atuais soluções para o seguinte problema de Otimização Combinatória: Dada uma frota de veículos de mesma capacidade, associada a um depósito central, encontrar as melhores rotas, para que os veículos atendam as demandas de um conjunto de clientes.

Este problema, conhecido na literatura como CVRP (Capacitated Vehicle Routing Problem) é uma generalização imediata do conhecido problema do Caixeiro Viajante e foi proposto por Dantzig e Ramser em [1].

O CVRP é um dos principais problemas da classe de problemas de Otimização Combinatória conhecida como Roteamento de Veículos. Esta é a classe de problemas que consistem na distribuição de produtos entre depósitos e clientes. Aplicações típicas de roteamento de veículos incluem, por exemplo, limpeza urbana, roteamento de ônibus escolares, determinação de plano de viagens de vendedores, distribuição de combustível em postos de gasolina, remoção de neve, entre outros.

O CVRP é, portanto, um problema de logística de transporte recorrente que tem atraído a atenção dos pesquisadores mais proeminentes da área desde os anos 60. Porém, atualmente, os melhores códigos disponíveis para a sua resolução somente resolvem de forma consistente instâncias com no máximo 135 vértices-clientes [3], [4], [2].

A proposta do projeto é desenvolver um novo mecanismo de geração de rotas válidas baseado em programação dinâmica para o problema CVRP afim de melhorar as atuais soluções (isto é, obter melhores limites inferiores para o valor ótimo de instâncias ainda em aberto ou então resolver instâncias ainda não resolvidas a otimalidade) para um conjunto considerável de instâncias de teste da literatura. Este novo código para a geração de rotas fará uso de parte do código desenvolvido e apresentado em [2] para que partes importantes implementadas possam ser reutilizadas como, por exemplo, a parte de geração de cortes e regras de branching.

## Referências Bibliográficas

- [1] DANTZIG, G. B; RAMSER, J. H. **The Truck Dispatching Problem**. Management Science, 6(1):80–91, 1959.
- [2] FUKASAWA, R; DE ARAGÃO, M; REIS, M; UCHOA, E. **Robust branch-and-cut-and-price for the capacitated vehicle routing problem**. Technical Report 8, Universidade Federal Fluminense, Engenharia de Producao, Niteroi, Brazil, 2003.
- [3] LYSGAARD, J; LETCHFORD, A; EGGLESE, R. **A new branch-and-cut algorithm for the capacitated vehicle routing problem**. Mathematical Programming, to appear., 2003.
- [4] RALPHS, T; KOPMAN, L; PULLEYBLANK, W; TROTTER, L. **On the capacitated vehicle routing problem**.

---

## Redes e Sistemas Distribuídos

---

Essa linha de pesquisa envolve aspectos dos fundamentos e da construção das redes de computadores e dos sistemas distribuídos. Como se sabe, as Redes de Computadores estão hoje presentes praticamente em todos os aspectos da nossa vida, não se limitando mais apenas à comunicação de dados, mas envolvendo a integração de serviços de voz e vídeo numa infra-estrutura comum, tipicamente baseada no protocolo da Internet, o IP.

Está em curso a assim chamada convergência das redes de computadores com as tradicionais redes de telecomunicações tanto para as redes cabeadas como para as redes sem fio. As redes de computadores formam a principal infra-estrutura de suporte aos sistemas distribuídos. Portanto, de uma forma geral, os sistemas distribuídos podem ser definidos como coleções de processos executando sobre um conjunto de computadores conectados por uma rede de computadores, possivelmente dispersos geograficamente.

Nos dias atuais, principalmente depois do estabelecimento da *World Wide Web*, a sociedade tem-se tornado cada vez mais dependente da utilização destes tipos de sistemas. Esta realidade tem induzido pesquisadores a investigarem técnicas para a construção de sistemas distribuídos confiáveis, que podem fornecer serviços de forma continuada, mesmo na ocorrência de falhas de alguns de seus componentes.

O presente capítulo apresenta a pesquisa relacionada a Redes e Sistemas Distribuídos. Assim, seis projetos foram descritos. Na Seção 3.1 é descrito o projeto de *Computação Sensível ao Contexto*. Na Seção 3.2 é descrito o projeto denominado *Identificação por Rádio-Frequência*. Já na Seção 3.3 o projeto *Serviços Baseados em Localização* é descrito. Na Seção 3.4 é feita uma descrição do projeto *InteGrade3: Gerenciamento de Recursos, Tolerância a Falhas e Escalabilidade em Grades Computacionais Autônomas*. Na Seção 3.5 é feita uma descrição do projeto *Processamento Paralelo Aplicado à Bioinformática*. Finalmente, na Seção 3.6 é descrito o projeto denominado *Paracoccidioides brasiliensis: Abordagens enfocando perfis de expressão gênica e moléculas recombinantes no estudo de interações patógeno-hospedeiro: Aplicações biotecnológicas*.

## 3.1 Computação Sensível ao Contexto

**Coordenador:** Dr Vagner José do Sacramento Rodrigues - vagner@inf.ufg.br

Esse projeto trata de aplicações computacionais que podem se adaptar automaticamente a mudanças no ambiente e às necessidades correntes do usuário sem exigir a sua atenção direta. Essas aplicações podem explorar características do ambiente tais como a posição do usuário, pessoas próximas, hora do dia, níveis de ruído, dentre outras, para fornecer serviços e informações adequadas ao contexto.

### 3.1.1 LoCoS (Local Context Service)

**Discente:** Marcio Pereira de Sá - marcio@inf.ufg.br

**Nível:** Mestrado

A computação móvel se torna uma importante ferramenta para os usuários no mundo atual na medida em que permite a estes mesmos usuários diminuir distâncias, economizar recursos (financeiros, de equipamentos, etc), aproveitar melhor o tempo. Entretanto, os dispositivos móveis, como PDAs, smartphones, dentre outros, são naturalmente dispositivos com uma grande limitação de recursos (CPU, memória, energia, etc.). Além disso, um mesmo aparelho móvel pode ser operado em diferentes ambientes e locais, bem como por diferentes usuários e propósitos.

Devido a estas particularidades, a aplicações móveis podem se beneficiar de informações a respeito do ambiente físico e temporal (localização, horário, data, temperatura, luminosidade, nível de ruído e outros), do ambiente operacional (nível de energia, capacidade de memória disponível, uso da CPU, etc.) além de informações sobre seus próprios usuários para adequarem sua execução às mudanças ocorridas no contexto no qual estão inseridas. Todas essas informações são denominadas informações de contexto.

Como exemplo, considere uma aplicação sensível ao contexto que, de acordo com a localização do dispositivo e o horário, pode alterar as configurações de tipo de campanha e volume do toque de um smartphone. Para esta aplicação, o usuário pode definir em suas preferências como e quando o toque do seu aparelho deve ser modificado. Por exemplo, o usuário poderia configurar que durante as reuniões definidas em sua agenda a aplicação deverá configurar automaticamente o tipo de campanha para silencioso ou alerta vibratório, evitando assim que uma chamada inesperada incomode os participantes da reunião correspondente.

Entretanto, na maioria das vezes, desenvolver componentes ou aplicações completas que realizem este processo de obtenção e disponibilização de informações de contexto é uma tarefa árdua e complexa, dispendendo muito tempo e recursos que poderiam

ser investidos no desenvolvimento da lógica de negócio das aplicações propriamente ditas. Desse modo, o desenvolvimento de camadas intermediárias (middleware) que sejam capazes de coletar e disponibilizar informações de contexto facilita o trabalho de desenvolvedores de aplicações sensíveis ao contexto para dispositivos móveis.

Em função disto, o objetivo deste projeto é desenvolver um serviço de provisão de contexto, denominado LoCoS (Local Context Service ou Serviço de Contexto Local), que fornecerá uma camada de abstração aos desenvolvedores de aplicações móveis no que diz respeito à coleta, processamento, representação e acesso às informações contextuais do dispositivo e do usuário.

Este serviço deverá ser extensível e flexível. Por extensível entende-se a facilidade de se adicionar ao LoCoS novos recursos e componentes, permitindo expandir suas capacidades de coleta e processamento de novos tipos de contexto à medida que se faz necessário. O termo flexível indica que o usuário deste serviço (o programador de aplicações sensíveis ao contexto) poderá escolher quais componentes internos do LoCoS deverão ser usados e quais deverão ser desativados em cada aplicação sensível ao contexto.

## 3.2 Identificação por Rádio-frequência (RFID)

**Coordenadores:** Dr Wagner José do Sacramento Rodrigues - wagner@inf.ufg.br e Dr Fábio Moreira Costa - fmc@inf.ufg.br

RFID é a sigla para *Radio-Frequency IDentification* em inglês, que em português significa Identificação por Rádio-Frequência. Trata-se de um método de identificação automática que utiliza sinais de rádio, recuperando e armazenando dados remotamente através de dispositivos chamados de tags RFID.

Uma tag (ou etiqueta) RFID é um transponder, pequeno objeto que pode ser colocado em uma pessoa, animal, equipamento, embalagem ou produto, dentre outros. Ele contém chips e antenas que permitem as tags responder aos sinais de rádio enviados por uma base transmissora. Além das tags passivas, que respondem ao sinal enviado pela base transmissora, existem ainda as tags ativas, dotadas de bateria, que lhes permite enviar o próprio sinal independente da base transmissora.

### 3.2.1 Uso de Hadoop para Distribuição de Armazenamento e Processamento em Sistemas de Dados Massivos de Controle de Redes de Sensores Diversos

**Discente:** Thiago Borges de Oliveira - thborges@gmail.com

**Nível:** Mestrado

O grande aumento na quantidade de dados e necessidade de processamento, advindos da coleta de dados automática por sistemas de sensores e ainda sistemas de indexação de conteúdo, tem lançado novos desafios aos sistemas computacionais existentes. É comum que super-computadores não sejam mais suficientes para se processar estes dados de maneira escalável e com ótimo desempenho. É necessário que se faça aglomerados desses computadores (ou até mesmo de computadores baratos) para se conseguir poder computacional suficiente para manipular uma grande massa de dados.

Empresas e organizações têm feito pesquisas e prototipado sistemas para lidar com esta necessidade. Uma destes esforços é o HADOOP, que é um sistema desenvolvido na organização Apache, com o propósito de distribuir carga - tanto de armazenamento quanto de processamento - que surgiu baseado na especificação da empresa Google de seu sistema de arquivos distribuído (GFS - Google File System) e de seu framework de Map-Reduce [3].

Este sistema tem o propósito de distribuir a carga de armazenamento e processamento através de tarefas de mapeamento (map) e redução (reduce), distribuídas através de um cluster de computadores baratos [3].

Este sistema foi lançado recentemente e é pouco explorado na resolução de problemas existentes. Acredita-se que sua arquitetura consiga resolver problema em muitas áreas da computação e algumas novidades tem sido mostradas em artigos publicados, como na resolução de algoritmos de máquina de aprendizado [2].

O objetivo deste projeto é explorar as capacidades deste sistema para o armazenamento e processamento de dados oriundos de redes de sensores (RFID, Sensores de Temperatura, Sensores de Localização) a fim de se produzir um framework de persistência e processamento distribuídos, utilizando como base clusters implementados em Hadoop.

Os desafios da proposta é conseguir um ganho pelo menos próximo de linear no tempo de processamento dos algoritmos, mediante a adição de nós no cluster. Serão implementados algoritmos baseados no framework Map-Reduce para o armazenamento de dados e a posterior descoberta de eventos nestes dados.

### **3.2.2 Gerenciador de Dados para uma rede RFID no escopo do projeto SINIAV**

**Discente:** Bruno Ferreira Machado - brunoferreira@inf.ufg.br

**Nível:** Mestrado

A pesquisa é relacionada à área de Identificação por Rádio-Frequência (RFID, na sigla em inglês), que é um método de sensoriamento que utiliza leitores específicos equipados com antenas para se comunicar com etiquetas (tags) também equipadas com uma antena. Os leitores podem tanto ler quanto escrever em tags (mas nem todas as tags têm suporte à escrita).

Existem inúmeros modelos de leitores e tags disponíveis no mercado, e cada um pode ter a sua própria representação de dados (alguns utilizam XML, outros utilizam cadeias de caracteres com *tokens* separando os atributos e assim por diante). Foi criado o padrão EPC [4] para solucionar esse problema (dentre outros), mas ele ainda não foi amplamente implantado pela indústria.

Nesse cenário diversificado, cada aplicação deveria compreender a representação de cada leitor com o qual gostaria de se comunicar, possivelmente inviabilizando o desenvolvimento da mesma ou a restringindo a apenas um pequeno grupo de leitores com uma representação comum.

Outra necessidade inerente ao uso de RFID, segundo [5], é a homogeneização dos dados obtidos dos leitores. Mais especificamente, podem ocorrer leituras duplicadas de tags, sendo que essa redundância pode ser prejudicial a alguma aplicação, além de falsos positivos, que seriam tags identificadas mas que não são do escopo do projeto RFID em questão e que podem interferir na aplicação utilizada.

A pesquisa visa solucionar os problemas da diversidade de representações de dados dos leitores RFID e da homogeneização dos dados lidos. A pesquisa terá como foco o projeto SINIAV [1] do Governo Federal, que tem como objetivo instalar uma etiqueta RFID em cada veículo circulando no território nacional para que possam ser identificados automaticamente.

Considerando a escala do projeto, que englobará milhões de tags sendo identificadas por milhares de leitores em centenas de cidades, faz-se absolutamente necessário um serviço que gerencie esses dados e possibilite que diversas aplicações, tanto governamentais quanto privadas, sejam desenvolvidas sem se preocupar com particularidades de diferentes leitores.

A solução proposta na pesquisa é o desenvolvimento do Gerenciador de Dados, um serviço que tratará da coleta de dados da rede e de envio de dados para serem escritos nas tags. Ele se comunicará com os diferentes modelos de leitores da rede RFID e padronizará a representação de dados de todos eles em um esquema comum, para que qualquer aplicação ou outro serviço possa se comunicar com qualquer leitor da rede independente da forma como ele organiza seus dados.

Esse serviço fará parte de um Middleware RFID, que contará também com outros serviços, como o Gerenciador de Eventos, para lidar com eventos na rede; Serviço de Persistência, para armazenar os dados obtidos; Serviço de Descoberta, para informar



outros serviços sobre a inserção de novos leitores na rede; dentre outros.

### 3.3 Serviços Baseados em Localização (LBS)

**Coordenador:** Dr Vagner José do Sacramento Rodrigues - vagner@inf.ufg.br

Um serviço baseado em localização (LBS da sigla em inglês) é um serviço de informação e entretenimento, acessível com dispositivos móveis utilizando alguma rede sem fio e que faz uso da posição geográfica do dispositivo móvel para oferecer um produto ou serviço ao usuário.

#### 3.3.1 LBS Shopping

**Discente:** Adriano Cunha - adriano.cunha@gmail.com

**Nível:** Mestrado

Clientes e lojistas em um shopping mall do tamanho de uma cidade - em breve - será uma realidade proporcionada pela computação móvel e ubíqua. A onipresença de PDAs, celulares e smart-phones, o aumento da capacidade de processamento dos dispositivos móveis, o crescente uso de serviços de comunicação de dados de terceira geração e a maciça inclusão de GPS nos aparelhos mais modernos dão suporte a criação de aplicações capazes de oferecer serviços do interesse do usuário baseados em sua localização (LBSs).

Tradicionalmente, as empresas bem sucedidas atendiam apenas ao mix de marketing (4 Ps), oferecendo um produto adequado a um bom preço em uma loja bem localizada e bem divulgada, mas a tecnologia está alterando radicalmente o panorama de um mercado cada vez mais competitivo. Ao mesmo tempo, os clientes são sobrecarregados de informações, escolher e encontrar produtos a preços razoáveis.

A tecnologia pode ser utilizada para combinar interesses: casar ofertas de produtos com interesses e desejos dos usuários, aproximando desta forma lojas e clientes. O objetivo deste trabalho é criar um sistema que, baseado na localização do usuário, utilizando um dispositivo móvel, ofereça produtos de seu interesse das lojas co-localizadas. Para tanto, o sistema fará o cruzamento da posição do usuário, dos produtos de seu interesse e produtos oferecidos pelas lojas da região em tempo real, bem como, verificará preço e disponibilidade dos produtos e mostrará a localização das lojas em um mapa.

## **3.4 InteGrade3: Gerenciamento de Recursos, Tolerância a Falhas e Escalabilidade em Grades Computacionais Autônomas**

**Coordenador:** Dr Fábio Moreira Costa - [fmc@inf.ufg.br](mailto:fmc@inf.ufg.br)

O Projeto InteGrade (<http://www.integrade.org.br>) visa desenvolver um middleware inovador que permita a utilização de recursos computacionais ociosos existentes em instituições acadêmicas e empresariais para a execução de aplicações científicas, industriais e de entretenimento que demandem alto poder computacional. O middleware é baseado em tecnologias avançadas de objetos distribuídos e pretende dar suporte à execução de aplicações paralelas onde haja um nível significativo de comunicação entre os nós (ao contrário de aplicações tipo bag of tasks onde não há comunicação entre os nós).

O projeto aqui proposto visa o desenvolvimento de uma versão completa do middleware do InteGrade, que opere de modo autônomo em ambientes de grande escala e forneça suporte a ambientes de computação móvel e computação ubíqua. Apesar dos avanços já obtidos e do nível de maturidade do projeto, ainda existem diversos desafios que precisamos atacar. Em particular iremos concentrar a pesquisa em áreas que permitirão o funcionamento da grade de forma mais autônoma, diminuindo a necessidade atual de intervenção de administradores humanos; estas áreas incluem escalonamento inteligente, tolerância a falhas, mobilidade, fornecimento de garantias de qualidade de serviço e armazenamento distribuído de dados de forma robusta. Iremos também explorar o uso da grade para diferentes classes de aplicações e aprimorar a sua escalabilidade.

### **3.4.1 BIOFOCO III: Software para análise genômica em ambiente cooperativo e distribuído na Região CO**

**Discente:**

**Nível:**

O projeto visa estabelecer infra-estruturas de hardware e software para o processamento de análise genômica, notadamente com base em tecnologias de computação em grade. Inicialmente, a implantação de uma grade computacional será realizada, sobre a qual aplicações de análise genômica serão executadas e avaliadas quanto ao seu desempenho. Em seguida, espera-se que a identificação de novos requisitos forneça subsídios para melhorias no middleware de grade utilizado no projeto (InteGrade).

### **3.5 Processamento Paralelo Aplicado à Bioinformática**

**Coordenador:** Dr Wellington Santos Martins - wellington@inf.ufg.br

O processamento paralelo de grande quantidade de dados vem ganhando uma atenção crescente nos últimos anos, e se firmando como uma alternativa viável, e de baixo custo, para a solução de problemas complexos em diferentes áreas.

Com a popularização das redes de computadores em diferentes níveis (WAN, MAN, LAN, Wi-Fi etc) e o crescente aumento do desempenho destas redes e dos processadores (agora com tecnologia multi-core e manycore), problemas antes exclusivos do mundo da supercomputação, podem ser solucionados com a agregação de máquinas e processadores multicore (e manycore – GPUs) de menor custo.

Uma das áreas que tem mais se utilizado destas soluções é a Biologia Molecular, que tem gerado uma grande quantidade de dados a taxas espantosas.

O objetivo deste projeto é o de investigar aplicações da computação paralela no processamento de dados biomoleculares, em especial no problema de comparação de seqüências visto que este problema é fundamental para a bioinformática.

### **3.6 Paracoccidioides brasiliensis: Abordagens enfocando perfis de expressão gênica e moléculas recombinantes no estudo de interações patógeno-hospedeiro: Aplicações biotecnológicas**

**Coordenador:** Dr Wellington Santos Martins - wellington@inf.ufg.br

O fungo dimórfico *Paracoccidioides brasiliensis*, agente etiológico da paracoccidioidomicose apresenta importância médica crescente.

A doença é endêmica nas Américas Central e do Sul, com a maioria das ocorrências no Brasil, o qual apresenta 80% dos casos descritos. *P. brasiliensis* parece ter múltiplos mecanismos de interação com o seu hospedeiro, o que deve incluir a presença de moléculas de superfície, as quais atuariam como fatores de adesão, enzimas proteolíticas e antígenos.

Com o objetivo de estudar a resposta de *P. brasiliensis* ao contacto com células do hospedeiro, serão analisados os perfis de transcrição do fungo isolado de infecção experimental em camundongos, fungo exposto a sangue humano e micélio em transição para leveduras. As ESTs diferenciais serão obtidas através de subtração dos cDNAs (RDA).

Através da comparação de dados obtidos nas diferentes condições experimentais, pretendemos identificar séries de genes especificamente expressos nos modelos utilizados. Utilizando-se abordagens bioquímicas serão analisados o papel de proteínas recombinantes nos estágios iniciais do processo infectivo, avaliando sua ação como adesinas *in vitro*.

Em síntese, utilizando-se uma combinação de abordagens experimentais e de Bioinformática e análises de ESTs geradas em condições de interação patógeno-hospedeiro, será possível obter-se padrões de expressão gênica do fungo em resposta a modelos que mimetizem a infecção humana. Utilizando-se abordagens bioquímicas iniciaremos a descrição do papel funcional de proteínas recombinantes nos estágios iniciais do processo infectivo. Moléculas recombinantes serão testadas quanto ao seu potencial de utilização na indústria farmacêutica.

## Referências Bibliográficas

- [1] BRAUN, V. **SINIAV - Sistema Nacional de Identificação Automática de Veículos**. <http://www.vonbraunlabs.org/siniav/port/index.html>, último acesso em Novembro de 2008, 2008.
- [2] CHU, C. T; KIM, S. K; LIN, Y. A; YU, Y; BRADSKI, G. R; NG, A. Y; OLUKOTUN, K. **Map-reduce for machine learning on multicore**. In: Schölkopf, B; Platt, J. C; Hoffman, T, editors, NIPS, p. 281–288. MIT Press, 2006.
- [3] CUTTING, D. **Hadoop Wiki**. <http://wiki.apache.org/hadoop/>, último acesso em Novembro de 2008, 2008.
- [4] EPCGLOBAL. **EPC**. <http://www.epcglobalinc.org/home>, último acesso em Novembro de 2008, 2008.
- [5] WANG, F; LIU, P. **Temporal management of rfid data**. In: VLDB '05: PROCEEDINGS OF THE 31ST INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, p. 1128–1139. VLDB Endowment, 2005.

## Sistemas de Informação

---

O presente capítulo apresenta os dez projetos associados à linha de pesquisa de Sistemas de Informação. Na Seção 4.1 o projeto *Ensino a Distância* é descrito. O projeto *Metodologias de Banco de Dados Aplicadas à Recuperação de Informações na WEB* é descrito na Seção 4.2. A Seção 4.3 apresenta o projeto *Uso de Tablet PCs e do Método de Aprendizado Baseado em Problemas no Ensino de Algoritmos e Programação*. Já a Seção 4.4 tem a descrição do projeto *Recuperação de Informação e Extração de Conhecimento em Ambientes Distribuídos e Heterogêneos*. A Seção 4.5, por sua vez, apresenta o projeto *Melhoria de Processos de Software em Goiás*. Na Seção 4.6 apresenta o projeto *Metodologias de Integração de Informações Armazenadas em Bancos de Dados / Web*. Já Seção 4.7 descreve o projeto *Teste de Software*. O projeto *Identificação de Relacionamento entre Objetos de Bibliotecas Digitais* é apresentado na Seção 4.8. A Seção 4.9 descreve o projeto *Resolução de Ambigüidade de Nomes em Bibliotecas Digitais*. Finalmente, o projeto *Propostas de Medidas de Relevância de Conferências* é apresentado na Seção 4.10.

### 4.1 Ensino a Distância

**Coordenadora:** Dra Ana Paula Laboissière Ambrósio - [apaula@inf.ufg.br](mailto:apaula@inf.ufg.br)

O Plano Nacional de Educação (Lei N° 010172, de 09/01/2001), prevê ter pelo menos 30% da população brasileira com idade entre 18 e 24 anos com diploma superior antes do fim da década, sendo que 40% da oferta de vagas deve ser feita pelo sistema público de ensino, estabelecendo uma política de expansão que diminui as diferenças entre as diversas regiões do país. Este é um desafio enorme.

A Educação a Distância tem se mostrado um meio importante de levar educação e treinamento a uma população numerosa e diversa. No Brasil, esta modalidade de ensino tem uma longa e importante história que começou com cursos por correspondência. Eram cursos não-formais, que tinham o intuito de educar mas não ofereciam diploma.

Somente recentemente programas de ensino superior foram autorizados a emitir diplomas reconhecidos pelo MEC.

A Lei de Diretrizes e Bases da Educação (LDB 1996) foi a primeira a definir a educação a distância como uma modalidade reconhecida de educação superior. Estabeleceu também várias regras para a implantação desta modalidade de ensino, incluindo o fato de que 20% dos cursos a distância devem ser presenciais em atividades que incluem avaliação e atividades de laboratório. Apesar de ser legalmente possível ofertar cursos superiores a distância, somente em 2005, quando o Ministério da Educação investiu US\$ 7 milhões para apoiar instituições públicas a ofertar treinamento de professores, é que esta modalidade foi de fato adotada pelas universidades públicas. Este projeto resultou na criação de 18 programas voltados para o treinamento de professores do ensino básico e fundamental, com 17,585 vagas ofertadas. O sucesso deste projeto acelerou o processo de criação da Universidade Aberta do Brasil (UAB), uma joint-venture entre universidades públicas, centros tecnológicos e municípios, que pretende oferecer aproximadamente 61,575 vagas somente em 2007. Isto representa uma expansão de 64% sobre o total de vagas ofertadas pelo sistema público de ensino.

A UFG também está investindo em Educação a Distância. Em 2006 foi implantado o primeiro programa a distância da UFG na área de Administração. Além disto faz parte do consórcio CONSET para o ensino de biologia e tem quatro cursos de graduação na UAB e dois de especialização. Isto representa um total de 1050 vagas. Estes cursos são oferecidos no sistema mixto de ensino, com aproximadamente 30-40% das atividades no modo presencial. Estas atividades são realizadas em polos implantados pelos municípios. Antes da implantação destes cursos, a UFG não tinha experiência nesta modalidade de ensino. Assim para que este tipo de ensino possa ser implantado com sucesso na UFG é necessário realizar estudos e treinamento.

Este projeto tem como objetivo realizar pesquisas sobre os diversos aspectos do ensino a distância, incluindo sistemas de gerenciamento de ensino (Learning Management Systems LMS), objetos de aprendizagem, teorias educacionais, etc.

#### **4.1.1 Uma Ferramenta de Hipertexto para Apoio à Aprendizagem Colaborativa Apoiada por Computador**

**Discente:** Luciana Oliveira - luciana.go@gmail.com

**Nível:** Mestrado

A "era da informação" provocou mudanças na sociedade que agora se refletem no ensino, exigindo que as instituições de ensino não atuem somente como transmissoras de conhecimento, mas que criem ambientes estimulantes aos alunos, que valorizem a

iniciativa e a descoberta, que possibilite ao aluno percorrer o caminho do conhecimento de maneira mais motivada, crítica e criativa. Esse ambiente deve também proporcionar um ambiente de parceria, de troca de experiências no ato de aprender e, sobretudo, possibilitar o desenvolvimento do pensamento crítico.

Nesse cenário, a Aprendizagem Colaborativa é uma metodologia que se destaca. Uma das definições de aprendizagem colaborativa é qualquer situação em que as pessoas se reúnem em grupo e são destacadas e consideradas as habilidades e contribuições individuais dos componentes do grupo. Atividades colaborativas envolvem ações em que as pessoas precisam explicar seu raciocínio aos parceiros. Isso traz resultados positivos tanto para quem recebe a explicação, que tem contato com novos conhecimentos ou a uma nova visão do que imaginava saber, quanto para quem explica, pois a verbalização requer elaboração do raciocínio, de modo a ser compreendido pelos demais.

Como o principal objetivo da Aprendizagem Colaborativa é incentivar os alunos ao hábito da pesquisa e, conseqüentemente, fortalecer a capacidade de auto-aprendizagem, as ferramentas de suporte a essa metodologia devem, preferencialmente, disponibilizar maneiras de apoiar o aluno, oferecendo recursos que o auxiliarão nesse processo.

O uso de material didático (apostilas, artigos científicos, e-books, páginas na internet, entre outros) adequado e de fácil acesso é fundamental para ensinar o diálogo permanente; orientar o estudante; motivar para a aprendizagem e também para a ampliação de seus conhecimentos sobre os temas trabalhados; ensinar a compreensão crítica dos conteúdos; instigar o estudante para a pesquisa e; possibilitar o acompanhamento e avaliação do processo de aprendizagem.

Neste sentido, a Internet está cada dia mais presente no cotidiano dos alunos e alterou até mesmo a forma como os textos são lidos. Em materiais didáticos clássicos, a leitura é seqüencial e guiada pelos sumários. Na Internet, a leitura é guiada pelas páginas compostas de vários hipertextos que possibilitam a navegação por múltiplos textos, cuja ligação (link) é determinada por meio de uma palavra. Um link é, na verdade, uma ponte entre produções textuais distintas. Dessa forma, é preciso considerar outra mudança advinda da popularização da Internet: a leitura descontinuada, hipertextual e tematizada.

Nesse sentido, a ferramenta proposta integra um **ambiente de sala de aula eletrônica** que usa apresentações em forma de slides, com **uso de tinta digital** e com **acesso, através de hipertexto, a uma base de conhecimento de material didático para consulta**, seja esse material disponibilizado em um repositório local ou na Internet. Através desta ferramenta, o aluno tem acesso a material complementar que poderá auxiliá-lo na compreensão e aprendizagem do assunto sendo abordado em sala de aula. Para isto basta marcar uma palavra ou conjunto de palavras no texto da apresentação, utilizando tinta digital ou o mouse, para que a ferramenta recupere material semanticamente semelhante

ou complementar. Visando a recuperação de material relevante para o assunto sendo abordado em sala de aula, a ferramenta analisa o contexto onde a consulta está sendo feita e expande a consulta do aluno com termos relacionados.

A identificação do contexto de um documento é realizada pelo ser humano de forma natural. A simples leitura de um título ou de um resumo, por exemplo, pode elucidar o leitor sobre o conteúdo do material que tem em mãos. Quando tratamos de um sistema informatizado, essa tarefa deixa de ser trivial. Isso porque o computador ainda não tem a mesma capacidade de interpretação humana. Por esse motivo, aplicar uma técnica que identifique o contexto de um documento tem sido um desafio para vários pesquisadores ao longo de muitos anos.

A **expansão da consulta** é uma técnica bastante utilizada na recuperação de informação para melhor contextualizar a consulta original do usuário. Basicamente a técnica consiste na adição de termos correlatos aos termos originalmente incluídos na consulta. Para que a consulta possa ser expandida, é necessária a identificação de termos que sejam relacionados aos termos existentes na consulta. Esses termos podem ser sinônimos, variações do tipo stemming, ou termos que estão próximos aos termos da consulta dentro do texto.

O ambiente de sala de aula eletrônica proposto prevê que as aulas sejam ministradas de forma síncrona, estando o professor e os alunos geograficamente em um mesmo ambiente. O **Classroom Presenter (CP)** é o software que inspirou a proposta da aplicação apresentada neste trabalho. Trata-se de um sistema *open source* para uso em equipamentos Tablet-PC, concebido pela Universidade de Washington, que permite o compartilhamento de tinta digital em slides, entre professores e alunos. Nesse cenário todos os alunos compartilham uma tela onde o professor faz a exposição do assunto. O CP permite ao professor interagir com os alunos de várias maneiras: através da realização de anotações na tela com o uso de tinta digital, esclarecendo dúvidas, propondo exercícios, entre outras possibilidades. Os alunos também podem enviar dúvidas para o professor e receber o feedback, de forma individual e privativa.

Em linhas gerais, o aplicativo proposto prevê a inclusão de melhorias ao CP: a qualquer momento, os alunos podem selecionar qualquer palavra contida na apresentação e ter acesso a um conjunto de documentos relacionado às palavras selecionadas com **conteúdo relevante**, compatível com o **contexto** que está sendo exibido na área de trabalho. O objetivo principal é estimular o aluno à prática da pesquisa auto-didata, com ênfase no trabalho em equipe e na troca de conhecimento.



## 4.2 Metodologias de Banco de Dados Aplicadas à Recuperação de Informações na WEB

**Coordenadora:** Dra Ana Paula Laboissière Ambrósio - apaula@inf.ufg.br

A WEB atualmente é uma fonte inesgotável de informações, porém a recuperação destas informações é ainda primária já que os mecanismos de busca muitas vezes não levam em conta a semântica e o contexto em que estas informações se aplicam. Estudos para tentar minimizar estes problemas vêm sendo desenvolvidos, entre eles aqueles voltados para WEB Semântica.

Já a área de Banco de Dados tem uma história no tratamento da semântica. Além disto, a WEB pode em última análise ser vista como um grande banco de dados com características e exigências específicas. Assim, o objetivo deste projeto é estudar diversas metodologias já utilizadas na área de Banco de Dados e verificar como elas podem ser utilizadas na recuperação de informações na WEB, propondo soluções que venham a facilitar e/ou aprimorar os processos de busca, visando desenvolver um sistema de informações para o Centro de Documentação, Informação e Memória da UFG (CeDIM).

O CeDIM se propõe a criar uma rede de produção, análise e circulação de informações, compreendida no sentido mais amplo do termo, envolvendo diversas unidades da Universidade Federal de Goiás. Cabe enfatizar o importante aspecto de inserção regional que a proposta apresenta. Um número expressivo de pesquisas na UFG tem como eixo privilegiado de análise, a produção de um conhecimento comprometido com a realidade regional. Ao lado desta dimensão estritamente acadêmica a explicação de processos regionais apresenta interesse imediato para a sociedade como um todo e o Estado em particular.

As informações que permitem uma análise adequada da região existem e são abundantes. Ocorre que, de forma geral, estão dispersas e carecem de uma sistematização. Existem importantes bases de dados nas Unidades Acadêmicas, nas bibliotecas universitárias, bem como nos órgãos do Estado. No entanto, as informações produzidas são as mais variadas, envolvendo diversos tipos de documentos, incluindo textos, músicas, vídeos, imagens, etc. Assim, este conjunto de informações pode em última análise ser visto como uma rede, onde os conceitos de Web Semântica podem ser aplicados a fim de permitir a recuperação eficiente do conteúdo.

A Web Semântica é caracterizada por associar às informações disponibilizadas na rede, dados de apoio, que atribuem semântica ao conteúdo e permitem que as informações sejam processadas através de regras de inferências e ontologias. Em outras palavras, sistemas poderão interligar documentos por dedução, com base em análise semântica das informações de apoio. Isto deverá resultar em uma indução automática dos resultados

pretendidos, mesmo que tais resultados não estejam diretamente ou sintaticamente conectados entre si.

A associação destes dados de apoio, que podem ser vistos como metadados dentro dos conceitos de banco de dados, exigem um processo de análise de requisitos, onde os diferentes conjuntos de informações, ligadas semanticamente, devem ser analisados a fim de permitir que sejam definidos os atributos que serão usados para melhor descrever o conteúdo destas informações e seus relacionamentos semânticos.

No contexto do CeDIM, esta análise detalhada dos dados, e o uso que se deseja fazer deles, deve ser feita, a fim de definir o modelo conceitual e lógico que melhor especificam os requisitos das informações específicas de cada unidade e que deverá determinar a implementação dos bancos de dados contendo os metadados. O conjunto destes bancos de dados formam um bando de dados federado que será usado para a recuperação das informações na rede. Após esta fase de estruturação e preenchimento das bases de dados, torna-se importante disponibilizar esta informação. Isto será feito através de um sistema de informações inteligente, capaz de entender e processar consultas feitas por usuários leigos, recuperando informações relevantes nas diferentes bases, de forma eficiente e transparente.

A especificação deste sistema de informação inteligente e flexível implica na solução de diversos problemas, entre eles a especificação de mecanismos de recuperação de informações. De fato, os sistemas de banco de dados surgiram como uma ferramenta para ajudar na organização, armazenamento e recuperação de grandes quantidades de informações. Para determinar a organização destas informações, o conceptor estuda o mundo real e o formaliza de tal forma que esta realidade possa ser usada em sistemas automatizados. Este processo de formalização, que ocorre durante a modelagem conceitual, vai influenciar de maneira importante os outros processos que dele dependem, incluindo a recuperação dos dados armazenados nas bases.

Deve-se ressaltar que o processo de modelagem conceitual é extremamente dependente do conceptor que o efetua, e que o mesmo sistema modelado por conceptores distintos pode ter diferentes características. No entanto, são estas características que vão determinar a forma na qual o usuário deverá efetuar uma consulta ao banco de dados. Para que uma consulta obtenha os resultados esperados, ela deve ser feita seguindo fielmente a estrutura de armazenamento adotada, que depende da formalização feita. Isto exige do usuário um profundo conhecimento da estrutura lógica dos bancos de dados, o que nem sempre corresponde à realidade. Na verdade, a maioria dos usuários não estão, e não deveriam estar, preocupados com a estrutura interna do Banco de Dados. Eles querem poder efetuar suas consultas da forma mais natural possível, e esperam obter as informações que desejam, isto é, informações relevantes. Para isto, deve-se flexibilizar o sistema de recuperação de informações permitindo que as consultas feitas por usuários

leigos, que muitas vezes não se adequam às estruturas e/ou conteúdo das bases de dados, possam retornar resultados semanticamente relevantes (consultas flexíveis).

Neste cenário, onde as pessoas interessadas nas informações contidas nas bases de dados provêm de diferentes backgrounds e possuem diferentes expectativas, a necessidade de oferecer mecanismos de consulta eficientes torna-se imprescindível. Para facilitar a interação dos usuários com o sistema de banco de dados, temos que levar em conta não somente as informações disponíveis, mas também o vocabulário usado por estes diferentes grupos, o contexto no qual os usuários estão inseridos e a percepção que eles têm do assunto. Esta diversidade de culturas e expectativas oferece um campo fértil e prático para a pesquisa na área de representação e compreensão de conceitos semânticos e a aplicação de seus resultados na área de consultas flexíveis.

O objetivo desta pesquisa é o estudo de metodologias de banco de dados que podem ser aplicadas na recuperação de informações na Web Semântica. Dentre estas metodologias podemos destacar os processos de análise de requisitos, modelagem conceitual e lógica, integração de esquemas, banco de dados ativos, federados e distribuídos entre outros, e principalmente os mecanismos e linguagens de consulta a banco de dados.

Especificamente, os mecanismos de consulta podem oferecer métodos de flexibilização que identifiquem e eliminem diferenças que ocorrem quando diferentes pessoas formalizam um conceito do Universo de Discurso, permitindo a recuperação de informações relevantes semanticamente para o usuário. Para isto, deve-se fazer uma análise dos métodos de categorização de conceitos e de definições de distâncias semânticas para aplicá-los à um sistema de consultas flexíveis.

Por sua relevância dentro do contexto da Web Semântica, o estudo das metodologias será iniciado com o estudo das técnicas de flexibilização de consultas em banco de dados e como estas podem ser aplicadas na especificação de um sistema para o CeDIM. A primeira parte consiste em estudar os mecanismos de classificação dos conceitos usados dentro de uma aplicação a fim de organizá-los de forma a permitir a identificação de outros conceitos que estejam dentro de uma "distância semântica" desejada. Para isto estuda-se os conceitos usados no contexto, e a noção de similaridade aceita pelos usuários. Depois que esta classificação está feita, ela será usada pelo sistema de consultas flexíveis para "modificar" a consulta do usuário e tentar ajudá-lo na busca das informações desejadas.

O CeDIM se apresenta dentro deste contexto como uma aplicação extremamente interessante. Primeiramente pelo desafio de tratar informações tão diversificadas em conteúdo e estrutura. Segundo pela quantidade de semântica existente nos diversos campos de pesquisa, que devem enriquecer o trabalho.

## 4.3 Uso de Tablet PCs e do Método de Aprendizado Baseado em Problemas no Ensino de Algoritmos e Programação

**Coordenadores:** Dra Ana Paula Laboissière Ambrósio - [apaula@inf.ufg.br](mailto:apaula@inf.ufg.br) e Dr Fábio Moreira Costa - [fabio@inf.ufg.br](mailto:fabio@inf.ufg.br)

Tradicionalmente, o ensino de Algoritmos e Programação é baseado em uma seqüência de tópicos, partindo das técnicas básicas e procurando abordar aspectos mais avançadas de programação, como fluxo de controle e modularização de programas. Tipicamente, os conceitos são expostos primeiro, seguidos de exemplos de sua aplicação e problemas específicos para os quais os alunos devem reproduzir e estender as soluções aprendidas. Contudo, o cenário típico de programação consiste em primeiro abordar o problema para então elaborar uma solução com base na análise do problema e na experiência prévia. Consideramos que esta é uma abordagem mais natural para o aprendiz. De fato, com a metodologia atual, apenas alguns alunos em cada turma apresentam níveis adequados de sucesso ao final do semestre, enquanto os demais parecem não assimilar suficientemente bem os conceitos chave, seus relacionamentos e aplicações. No Instituto de Informática, tentativas foram realizadas no passado no sentido de melhorar o aproveitamento nesta disciplina, notadamente utilizando elementos do método de Aprendizado Baseado em Problemas (Problem-Based Learning - PBL). Os níveis de aproveitamento dos alunos foram significativamente mais altos em comparação com outras turmas. Contudo, o uso do método PBL mostrou problemas com relação à eficiência do ensino e participatividade dos alunos, o que comprometia o cumprimento do programa da disciplina.

Neste projeto, planejamos retomar o uso do método PBL, agora assistido pelo uso de tecnologias de computação móvel e ubíqua. O projeto objetiva investigar e aplicar essas tecnologias para incrementar o ensino na disciplina introdutória de Algoritmos e Programação. Esta disciplina é ofertada pelo INF para diversos cursos de graduação da UFG, incluindo Ciência da Computação, Engenharia da Computação, Engenharia Elétrica, Engenharia Civil, Matemática, Física, Química, Engenharia de Alimentos e Agronomia. Um grande problema nesta disciplina é conseguir que os estudantes adquiram o domínio dos princípios da construção de algoritmos e programas e sejam capazes de aplicá-los para resolver problemas reais. Como parte do projeto, planejamos fazer um re-design da disciplina utilizando PBL como a metodologia central de ensino. As tecnologias de computação móvel e ubíqua serão um fator chave para facilitar a aplicação desta estratégia. Mais especificamente, a sala de aula será dotada de um conjunto de

computadores do tipo tablet PC, um para cada um ou dois alunos, além de um tablet para o professor, um projetor multimídia e uma rede sem fio conectando estes equipamentos.

Usando os tablets, os alunos serão capazes interagir colaborativamente uns com os outros e com o professor de maneira mais efetiva, intercambiando, avaliando e complementando as soluções por eles produzidas para os problemas apresentados em aula. Com o uso de ferramentas apropriadas, os professores da disciplina estarão melhor instrumentados para acompanhar as atividades realizadas pelos alunos durante as aulas, através da submissão automática de tarefas e da análise semi-automática dos resultados e de dados sobre o desempenho dos alunos. Isto permitirá que o professor possa re-orientar seu ensino durante a aula, conforme a necessidade da turma. Os tablets serão também úteis na condução do próprio processo de aprendizado via PBL, uma vez que os mesmos representam uma ferramenta efetiva para a instrumentação do trabalho em grupo e colaborativo.

Uma parte importante do projeto consistirá no desenvolvimento de ferramentas de software, algumas com base em ferramentas já existentes, para dar suporte ao uso dos tablets como parte da metodologia proposta. Finalmente, pretendemos também desenvolver métodos eficazes de avaliação que tirem proveito dos recursos tecnológicos disponíveis no projeto.

### 4.3.1 Esboço de Fluxogramas no Ensino de Algoritmos

**Discente:** Halley Wesley Alexandre Silva Gondim - halley@inf.ufg.br

**Nível:** Mestrado

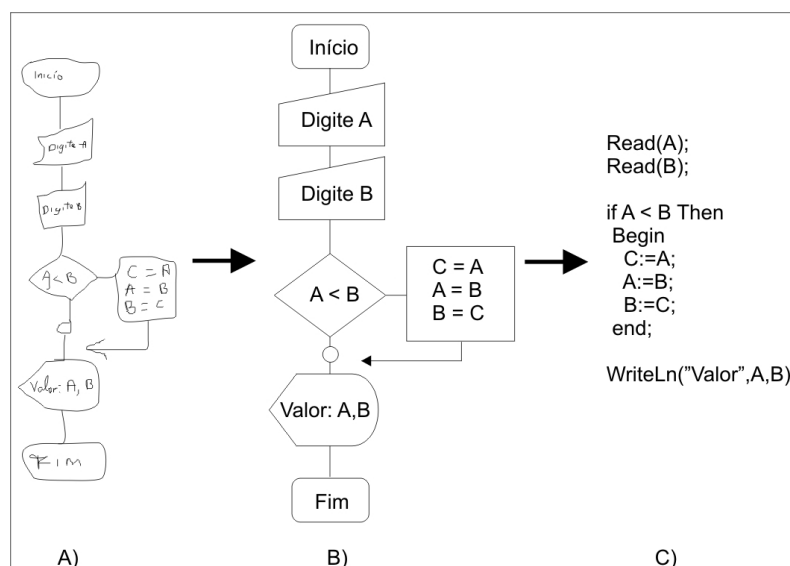
AA disciplina Algoritmos e Programação de Computadores sempre representou um desafio para os cursos de Computação. De fato, este é considerado um dos sete grandes desafios do ensino da computação [33]. Ensinar programação básica no ensino superior tem servido de tema para muitas discussões entre professores da disciplina [3]. Problemas com esta disciplina são apontados como responsáveis pelo grande número de reprovações e desistências em cursos de Computação [30]. Trabalhar esta disciplina torna-se portanto vital para o sucesso dos cursos na área.

Por outro lado, sabe-se que o ser humano tem mais habilidade para interpretar um algoritmo usando formas visuais que em formato de código texto. Isto se deve ao fato que o hemisfério esquerdo do cérebro processa informação oral e lógica, enquanto o hemisfério direito processa informação visual e espacial. Quando o aluno interpreta o código texto, o hemisfério direito não contribui significativamente no processo de aprendizagem. [9]. Assim, a chave para fazer o ensino de algoritmo mais acessível está em utilizar aspectos visuais e textuais, estimulando ambos os hemisférios do cérebro.

O ambiente proposto visa a confecção de esboços de fluxogramas utilizando a tinta digital [42]. Mas, se esboços são tão informais e pessoais, então qual o motivo de se criar uma ferramenta de esboço ao invés de utilizar as atuais ferramentas que criam fluxogramas de modo mais preciso? Pode-se citar dois motivos. O primeiro é que a utilização de uma caneta sempre será uma forma mais natural do que um teclado ou mouse, deixando de forma livre e mais intuitiva as imagens geradas, aliando a facilidade de pensamento à facilidade de esboçar. Segundo e mais importante, os sistemas que possuem as formas de um fluxograma possuem limitações impostas pelos projetistas [45]. Por exemplo, tamanho e ângulos precisos.

O ambiente também possui um reconhecedor das formas básicas do fluxograma e um conversor para uma linguagem. Como cada forma gráfica básica reconhecida do fluxograma possui semântica específica, que se traduz em uma estrutura das linguagens de programação, na execução de um fluxograma, o interpretador analisa o código montado em memória e o executa [50]. Assim, o interpretador nada mais é que a definição de uma linguagem possuindo um conjunto de instruções.

Como o fluxograma possui uma gramática idêntica à da interpretação, é possível realizar a conversão para uma pseudo-linguagem ou para uma linguagem de máquina específica, por exemplo, Pascal [50]. Desta forma o aluno ao esboçar o algoritmo utilizando o auxílio da tinta digital pode, além de ter mais facilidade na anotação, escolher converter o fluxograma para uma linguagem específica, trazendo assim um facilitador para o primeiro contato com a construção de códigos. Um resumo de todo este trabalho se aplica na Figura 4.1.



**Figura 4.1:** Exemplo de esboço de fluxograma (A), reconhecimento do desenho (B) e conversão para código (C)

## 4.4 Recuperação de Informação e Extração de Conhecimento em Ambientes Distribuídos e Heterogêneos

**Coordenador:** Dr Luiz de Carvalho - cedric@inf.ufg.br

Este projeto tem como objetivo a construção de infra-estrutura computacional para apoiar a construção de Comunidades Virtuais de Prática na Web. A infra-estrutura deve ser baseada nos princípios fundamentais da Web Semântica e deve ser estabelecida sobre uma Grade Computacional. Para tanto, deverão ser implementadas ferramentas para a gestão de informações no contexto das comunidades que permitam a construção de Sistemas de Apoio à Decisão.

Existem três subprojetos relacionados a este projeto de pesquisa: “Extração de características de documentos textuais para melhoria da eficácia na classificação de textos”, “Grades Semânticas em Comunidades Virtuais” e “Integração de Repositórios Digitais com a Utilização de Especificações Semânticas”. O resumo de cada um dos trabalhos é apresentado nas subseções posteriores.

### 4.4.1 Extração de Características de Documentos Textuais para Melhoria da Eficácia na Classificação de Textos

**Discente:** Fernando Chagas Santos - fernandosam@gmail.com

**Nível:** Mestrado

Sistemas de Recuperação de Informação (SRI) lidam com o problema de representar, armazenar, organizar e acessar documentos [5]. Uma das principais tarefas de um sistema de recuperação de informações é a recuperação de documentos através de consultas formuladas por um usuário de acordo com a sua necessidade de informação. A parte do sistema responsável por esta tarefa é chamada de máquina de busca.

Antes da década de 90, os SRI eram desenvolvidos principalmente para os usuários pesquisar por informações em coleções específicas, como publicações científicas e bibliotecas digitais. Havia uma padronização na estrutura dos documentos destas coleções, permitindo um certo controle sobre o formato e o conteúdo destes documentos. Além disso, os usuários desses sistemas possuíam treinamentos sobre como formular consultas, permitindo-lhes expressar melhor as suas necessidades de informação [48].

Grande parte dos SRI, desenvolvidos antes da década de 90, forneciam um suporte adequado aos seus usuários e não traziam muitos desafios. No início da década

de 90, o surgimento da *World Wide Web* (WWW)<sup>1</sup>, ou simplesmente *Web*, modificou este cenário e trouxe novos desafios para a área de Recuperação de Informação (RI) e as suas disciplinas relacionadas.

A *Web* é a principal fonte de informação de muitas pessoas e uma das fontes mais utilizadas para a pesquisa de informação [40]. Atualmente existe mais de 570 milhões de endereços (*hosts*) na *Web*<sup>2</sup> e um imenso volume de documentos armazenados com diferentes estruturas, formatos e conteúdos.

Uma estratégia típica para localizar informações na *Web*, chamada busca por padrão, consiste na formulação de consultas em máquinas de busca utilizando palavras-chave. Os problemas desta estratégia são a quantidade e a qualidade dos documentos recuperados em uma consulta. O usuário precisa analisar vários documentos para verificar quais deles são relevantes para satisfazer a sua necessidade de informação.

A *Web Semântica*, uma iniciativa do *World Wide Web Consortium* (W3C), busca melhorar as potencialidades da *Web*, definindo padrões e tecnologias que possibilitem a integração e a combinação de dados na *Web* [59] [6]. Os documentos publicados na *Web* de acordo estes padrões, possibilitarão que as pessoas e os computadores trabalhem de maneira colaborativa [6]. As máquinas de busca por conteúdo é um exemplo de ferramenta neste sentido [11].

As máquinas de busca por conteúdo minimizam os problemas das máquinas de busca por padrão. As palavras-chave formuladas pelo usuário são utilizadas como um guia na obtenção de conteúdos específicos, restringindo o espaço de busca dos documentos recuperados em uma consulta. Desta forma, a quantidade de documentos irrelevantes recuperados em uma consulta e a quantidade de documentos a serem verificados pelo usuário diminuem em comparação com as consultas em máquinas de busca por padrão.

As ferramentas da *Web Semântica*, tais como as máquinas de busca por conteúdo, dependem de documentos semanticamente estruturados. Contudo, cerca de apenas 2 milhões de documentos<sup>3</sup> na *Web* possuem esta estruturação e apesar da maturidade de alguns padrões e tecnologias da *Web Semântica* [59], não é possível controlar o formato e a estrutura na qual os documentos são publicados na *Web*.

Outra estratégia para localizar informações na *Web* consiste no usuário navegar em um diretório *Web* (*Yahoo! Directory*, *Open Directory* e *Google Directory*), onde os documentos da *Web* estão organizados em uma hierarquia de tópicos. Esta estratégia possui vantagens semelhantes às máquinas de busca por conteúdo. Contudo, um dos maiores problemas desta estratégia está na manutenção do diretório *Web*.

---

<sup>1</sup> A WWW é um sistema distribuído de hipertexto que possibilita o intercâmbio eletrônico de documentos hipertexto entre pessoas.

<sup>2</sup> <http://www.isc.org/index.pl?/ops/ds/>

<sup>3</sup> [http://swoogle.umbc.edu/index.php?option=com\\_swoogle\\_stats](http://swoogle.umbc.edu/index.php?option=com_swoogle_stats)



## Problema

A manutenção de um diretório *Web* pode ser manual ou automática. Na manutenção manual, um profissional é responsável pela tarefa de determinar o tópico que um documento pertence. Esta abordagem pode ocasionar os seguintes problemas: documentos podem ser incluídos em tópicos incorretos e tópicos podem não ser atualizados com a criação de novos documentos. Na manutenção automática, a tarefa de determinar o tópico que um documento pertence é feita por um classificador automático. Esta tarefa é chamada de categorização hierárquica de documentos, uma das aplicações da classificação de textos.

A manutenção automática de diretórios *Web* com muitos tópicos e estrutura hierárquica, tais como o *Yahoo! Directory* com mais de 200 mil tópicos, não é eficaz. Uma abordagem híbrida [31], que mistura a manutenção manual e automática de diretórios *Web*, tem sido utilizada para aumentar a eficácia na classificação automática. Nesta abordagem, um classificador automático é executado inicialmente e aqueles documentos que um revisor inclui em uma fila são considerados como documentos de baixa confiança. Em uma próxima iteração, o classificador automático considera estes documentos com baixa confiança para melhorar a sua eficácia.

A classificação de textos é uma disciplina (e uma tarefa) que teve início na década de 60 e, deste então, tem sido utilizada em contextos onde é necessário organizar, selecionar ou adaptar documentos. Até o final da década de 80, a abordagem mais popular para a classificação de texto foi a engenharia de conhecimento [51]. Entretanto, a partir da década de 90 esta abordagem foi perdendo popularidade em favor da abordagem de aprendizagem de máquina [51] e desde então tem sido a abordagem mais utilizada na classificação de textos [54].

Existe basicamente três metodologias de aprendizagem de máquina - supervisionada, não supervisionada e semi-supervisionada - e vários algoritmos para gerar um classificador automático, também chamado de aprendiz. Estudos mostram que as Máquinas de Vetores de Suporte (SVM, do inglês *Support Vector Machines*), são um dos algoritmos mais eficazes para a classificação de textos (e.g., Yang e Liu, 1999; Joachims, 1998; Dumais et al., 1998).

Na metodologia de aprendizagem supervisionada, por exemplo, um processo indutivo constrói um classificador, ou aprendiz, para uma categoria  $c$  a partir da observação das características de um coleção de documentos classificados manualmente, por uma pessoa, sobre a categoria  $c$  [51]. Seja  $D$  um domínio de documentos e  $C$  um conjunto de categorias predefinidas, a classificação de textos é a tarefa de determinar se um documento  $d_j$  pertence a uma categoria  $c_i$  a partir do uso do aprendiz construído na etapa anterior.

Escolher boas características de um coleção de documentos pode melhorar, substancialmente, a eficácia de um classificador [15]. A extração de características busca

encontrar um subconjunto que ao mesmo tempo seja o menor e que otimize um objetivo primário [17]. O objetivo primário da extração de características pode ser: facilitar a visualização e o entendimento dos dados, reduzir o tempo de treinamento e o tempo de teste ou melhorar a eficácia de um classificador [18].

### **Objetivo**

A hipótese deste trabalho é que é possível melhorar a eficácia na classificação de documentos na *Web* a partir da investigação de métodos de extração de características. Desta forma, este trabalho tem como objetivo comparar métodos de extração de características e identificar os mais eficazes na classificação automática de documentos na *Web*. Para alcançar o objetivo do trabalho, os seguintes objetivos específicos precisam ser alcançados:

- Pesquisar métodos de seleção de características.
- Analisar os métodos pesquisados em diferentes coleções de documentos na *Web*.
- Comparar os resultados obtidos na análise dos métodos pesquisados.
- Comparar os resultados obtidos com outros resultados publicados na literatura.

### **4.4.2 Grades Semânticas em Comunidades Virtuais**

**Discente:** Hellen Carmo de Oliveira - hellen.sistemas@gmail.com

**Nível:** Mestrado

O advento da Internet como meio de comunicação ágil, flexível e de baixo custo, e sua adoção em larga escala pelas instituições foram os propulsores das comunidades virtuais. Grupos de pessoas com interesses comuns - em uma instituição ou em várias - se formaram lentamente, se comunicando através de e-mail, chats e websites. Profissionais/estudantes de uma área específica passaram a poder trocar informações relevantes para o seu dia-a-dia, sobre suas “melhores práticas”, a forma como estruturaram seus processos e a compartilhar soluções para os seus problemas mais comuns. Verdadeiras “comunidades” começaram a se formar nas instituições em torno do compartilhamento de suas “práticas”.

Os produtos desse compartilhamento de informações são diversos dados sendo armazenados constantemente sem um controle específico de armazenamento e de uma busca posterior.

As grades computacionais propiciam uma arquitetura interessante que pode ser utilizada no gerenciamento das informações específicas compartilhadas dentro dessas comunidades virtuais.

Para se trabalhar com grades computacionais é preciso ter um ambiente adequado onde haja máquinas conectadas por uma rede para que as tarefas sejam executadas.

Grades computacionais surgiram na década de 90 para tornar possível a execução paralela de aplicações em recursos geograficamente distribuídos. A idéia de grades computacionais é bastante simples: executar aplicações de forma paralela em recursos abundantes sem a necessidade de se investir em um supercomputador, muitas vezes inviável [2].

Um dos mais conceituados *middlewares*<sup>4</sup> de grades existente é o Integrate. O Integrate [16] é um *middleware* para grades oportunistas, isto é, voltado para o compartilhamento de recursos computacionais ociosos. Aplicações, dados, estruturas de rede, *middleware* e equipamentos compõem o universo dos recursos disponibilizados no escopo das grades. A regência organizada desse conjunto promove o uso do Integrate como um meio efetivo sobre o qual aplicações complexas e com grande demanda de recursos podem ser eficientemente executadas.

A infra-estrutura da grade computacional é organizada principalmente para atender requisitos gerados pela demanda de execução de aplicações e do armazenamento e recuperação dos dados produzidos e consumidos por essas mesmas aplicações [58]. Assim, identifica-se a importância da abordagem semântica, como uma alternativa para lidar com a heterogeneidade dos recursos computacionais das grades, das aplicações que nelas são executadas e dos diversos dados que nelas serão distribuídos. Assim, obtém-se uma arquitetura de grade com recursos bem definidos: a Grade Semântica.

As Grades Semânticas são uma extensão da Grade computacional, com informações e serviços bem definidos e significados explicitamente representados, permitindo melhor trabalho em equipe para pessoas e computadores [49]. O uso da palavra “semântica”, nesse contexto, não implica, evidentemente, na captura do significado das informações por parte das máquinas, mas que conteúdo com significado lógico pode ser mecanicamente manipulado pelos computadores com finalidades úteis para os seres humanos[58].

As Grades Semânticas propiciam uma infra-estrutura tecnológica, em que *middlewares* permitem, semanticamente, colecionar recursos de computação, armazenagem, conjuntos de dados, bibliotecas digitais, instrumentos científicos, instituições e pessoas. Assim, as instituições e pessoas rapidamente se unem para formar organizações virtuais para solucionar um problema específico e, assim facilmente ser encontrada uma solução.

A grade semântica procura incorporar a abordagem da Web semântica à grade computacional, desempenhando o papel de uma conexão entre os domínios de grades computacionais e de Web semântica [49].

O “conhecimento para a grade” possibilita ao *middleware* da grade processar

---

<sup>4</sup> *Middleware* é um software que reside entre o sistema operacional e a aplicação a fim de facilitar o desenvolvimento de aplicações, escondendo do programador diferenças entre plataformas de hardware, sistemas operacionais, bibliotecas de comunicação, protocolos de comunicação, formatação de dados, linguagens de programação e modelos de programação [12].

conhecimento sobre disponibilidade e propósito de serviços e recursos da grade, suas políticas de uso, estados e direitos de acesso, em uma extensa lista de possibilidades. Por exemplo, uma arquitetura de grade baseada em um *middleware* flexível e dinâmico deve possibilitar a descoberta e formação dinâmica de organizações virtuais *ad hoc* de recursos dispersos na grade[58].

O uso da arquitetura de Grade Semântica pode facilitar a implantação de aplicações complexas, como representação e recuperação de conhecimento, nas quais comunidades virtuais estarão envolvidas e seus diversos recursos poderão estar compartilhados.

O objetivo geral deste trabalho é propiciar um gerenciamento semântico das informações dentro de uma comunidade virtual utilizando a arquitetura de grades semânticas. Espera-se que se obtenha o controle do armazenamento dos documentos compartilhados de diferentes extensões (.pdf, .doc, .txt etc) de uma comunidade virtual e a posterior busca do conhecimento contido nesses documentos.

O objetivo específico é propor e implementar uma aplicação para o armazenamento distribuído de dados e outra para a busca semântica desses dados no *middleware* Integrate.

### 4.4.3 Integração de Repositórios Digitais com a Utilização de Especificações Semânticas

**Discente:** Renan Rodrigues de Oliveira - cmp.renan@gmail.com

**Nível:** Mestrado

O Centro de Documentação Informação e Memória (CDIM) é um projeto aprovado pela FINEP, cujo objetivo é formar um acervo de documentos, informações e arquivos que contribuam com a melhoria das condições de pesquisa no Estado de Goiás.

Para alcançar estes objetivos, espera-se a construção de um sistema que funcionará como um ponto de acesso único a diversos repositórios de dados distribuídos, contendo dados nos mais diversos formatos.

Além desse objetivo, também é foco desse projeto a conservação de documentos, sons e vídeos que retratam os valores históricos políticos e culturais do Estado, capazes de fomentar estudos e pesquisas voltadas para formulação e avaliação de políticas públicas para Goiás.

Com relação à implementação do CDIM, espera-se desenvolver um sistema que torne possível a realização de tarefas bastante complexas na busca por informações. Espera-se atingir este objetivo com a utilização de tecnologias da Web Semântica.

Um dos desafios é que este sistema compreenda e vincule dados com semântica, isto é, através de ontologias e regras de inferências, torne possível que se encontrem

respostas mais precisas para as consultas, descartando o que é irrelevante para o usuário.

Um dos passos iniciais deste projeto inclui a implantação do sistema Fedora, uma plataforma de código aberto que surge como uma importante alternativa para a implementação de repositórios digitais.

A arquitetura Fedora tem sido adotada por instituições de diversos seguimentos, abrangendo sua utilização a bibliotecas digitais, repositórios institucionais, tecnologias para educação, publicação de conteúdo, distribuições de informações na Web, entre outros.

## **4.5 Melhoria de Processos de Software em Goiás**

**Coordenador:** Dr Juliano Lopes de Oliveira - juliano@inf.ufg.br

O projeto visa desenvolver processos, métodos e ferramentas de Engenharia de Software que possam ser aplicadas em empresas goianas, principalmente naquelas de pequeno porte.

### **4.5.1 Uma Abordagem MDA para Geração e Evolução Automática de Esquemas de Bancos de Dados**

**Discente:** Alexandre Cláudio de Almeida - alexandreclaudio@gmail.com

**Nível:** Mestrado

Existem inúmeras variações de processos de desenvolvimento de sistemas de software e um problema comum a todas elas é a dificuldade para modelar, documentar, construir e evoluir o esquema de banco de dados que armazena as informações relevantes do software. Em geral, as atividades relacionadas com a definição e a modificação do esquema de banco de dados de um software apresentam características negativas. Uma delas é a baixa produtividade da equipe técnica: a necessidade de mapeamento manual entre esquemas conceituais, lógicos e operacionais representa uma barreira para o desempenho dos profissionais envolvidos com a construção de bancos de dados. Outra dificuldade é garantir portabilidade e interoperabilidade da aplicação, pois se os recursos específicos de um SGBD são utilizados, o desempenho do software é melhor, mas a aplicação perde em portabilidade.

Um problema ainda mais grave é a documentação inconsistente ou imprecisa de esquemas, originada da redundância causada pelos diversos níveis de modelos (conceitual, lógico e operacional, por exemplo) de bancos de dados. Cada mudança precisa ser refletida em diversos modelos, o que aumenta o esforço de modificação. Na prática, as

mudanças feitas em modelos de baixo nível (ou diretamente no código) dificilmente são refletidas nos documentos conceituais, tornando-os inúteis.

A abordagem MDA (*Model Driven Architecture*) [29] é uma proposta para reduzir o impacto desses problemas no contexto de desenvolvimento e manutenção do sistema de software, tendo como base a utilização de modelos abstratos para representar os sistemas. O desenvolvimento de software dirigido por modelos permite especificar conceitos relevantes do domínio da aplicação, incluindo regras do negócio, relacionamentos e semântica, usando linguagens de modelagem de alto nível de abstração.

Com estas linguagens são criados modelos do sistema, independentemente da plataforma que será utilizada para a sua implementação. Modelos em UML, por exemplo, são o padrão de fato utilizado pela indústria de software. Esses modelos de alto nível são chamados de PIM (*Platform Independent Model* - Modelo Independente de Plataforma). A idéia central da MDA é a transformação de modelos através do mapeamento automático de um modelo PIM para um ou mais modelos voltados para uma plataforma específica (PSM - *Platform Specific Model*) e, a partir deste, gerar o código da aplicação nesta plataforma. Embora a abordagem MDA tenha sido proposta para resolver problemas de desenvolvimento e manutenção de software, os seus princípios também podem ser aplicados para especificar bancos de dados, já que estes são um componente importante de qualquer software [13]. Portanto, a proposta deste projeto é aplicar a abordagem MDA para identificar, modelar, criar e evoluir o esquema de banco de dados de um determinado software. Neste contexto, os conceitos relevantes do domínio da aplicação devem ser representados em modelos independentes de plataforma (PIM) e posteriormente mapeados para modelos e código voltados para plataformas específicas (PSM), produzindo esquemas de bancos de dados com as restrições de integridade que garantem a consistência das informações armazenadas.

A ferramenta proposta neste projeto contribui em vários aspectos da concepção de um sistema de bancos de dados. A produtividade da equipe de desenvolvimento é maior, pois o projetista do banco de dados (desenvolvedor do modelo PIM) tem menos trabalho para criar um modelo do sistema, visto que os detalhes referentes à tecnologia de implementação podem ser abstraídos deste modelo.

Esses detalhes de implementação são levados em consideração no momento da definição da transformação do modelo PIM para o PSM. Além disso, as ferramentas de transformação e geração de código permitem a redução da quantidade de código que deve ser escrita manualmente por um programador.

Com relação à portabilidade, a utilização de MDA permite que um mesmo modelo PIM possa ser transformado em modelos PSM voltados para plataformas variadas. Para isso, basta utilizar outra definição de transformação correspondente à linguagem alvo, ou seja, a mudança de um SGBD implica na criação de uma nova definição de trans-

formação do modelo conceitual para o dialeto SQL do SGBD alvo.

Em um processo de desenvolvimento tradicional, tanto a documentação quanto a manutenção são atividades muito dispendiosas e que geralmente causam muitos problemas. A maior parte dos desenvolvedores não cria documentação dos sistemas de forma apropriada, pois acha que essa atividade não traz nenhum benefício para ele. No entanto, isso não é verdade. A documentação é essencial para que futuras manutenções no sistema sejam possíveis de serem realizadas.

A utilização do modelo PIM permite que o desenvolvedor trabalhe em um nível mais alto de abstração. Esse modelo deverá ser transformado em um modelo PSM com a utilização de uma ferramenta de transformação. Dessa forma, sempre que alguma alteração seja necessária no sistema, basta modificar o modelo PIM e executar novamente a transformação para o modelo PSM.

Isso faz com que a documentação do sistema esteja sempre atualizada, pois qualquer modificação no modelo PIM irá refletir no modelo PSM que foi gerado a partir dele e também nos esquemas gerados. Além disso, nota-se que o processo de manutenção é simplificado, visto que se resume em reestruturar o modelo PIM e aplicar a transformação de modelos.

## **4.5.2 Geração e Manutenção de Interfaces Gráficas com Usuário Baseada em Metamodelos**

**Discente:** Wilane Carlos da Silva - wilanec@hotmail.com

**Nível:** Mestrado

Fazer com que o processo de desenvolvimento de sistemas seja mais produtivo e, ao mesmo tempo, com uma maior qualidade é um desafio constante para a engenharia de software. A cada ano esse desafio torna-se mais importante devido ao pouco tempo disponível para o desenvolvimento de aplicações e para sua manutenção.

A interface gráfica com usuário tem um papel de destaque na grande maioria das aplicações, já que ela está em contato direto com o usuário. Para este a interface gráfica é o próprio sistema. Ela sendo de baixa usabilidade, por exemplo, pode colocar em risco todo um sistema bem elaborado. Ou o contrário, um sistema cuja aplicação seja mal feita pode ser salvo por uma interface gráfica agradável ao usuário. As constantes modificações que ocorrem no escopo de desenvolvimento de uma interface gráfica, em seu ciclo de vida são diretamente proporcionais à importância dessa interface. O usuário, cada vez mais exigente, deseja que o sistema tenha, resumidamente, uma ótima usabilidade, confiabilidade e configurabilidade e, geralmente, exige várias mudanças até que o software fique com a aparência e comportamentos ao seu gosto. Já o desenvolvedor

visa a criação de um software com boa manutenibilidade, justamente para implementar as modificações necessárias durante o processo sem muito esforço e sem consumir muito tempo, já que, segundo estudos realizados, a fase de manutenção do software costuma ser mais duradoura que a fase de desenvolvimento [43].

Sendo assim, esse trabalho de mestrado propõe a criação de um framework para a geração automática de interfaces, ou GUIs (Graphical User Interfaces), a partir de metamodelos, mais especificamente o MDA [37], introduzido em 2001 pelo OMG (Object Management Group) [38], fazendo com que elas sejam formadas a partir de regras do sistema. Essa geração faz com que o sistema tenha principalmente uma maior manutenibilidade e configurabilidade.

### **4.5.3 Modelagem e Implementação de Restrições de Integridade Utilizando OCL com Mapeamento para SQL**

**Discente:** Glauber Boff - bglauber@gmail.com

**Nível:** Mestrado

O desenvolvimento de aplicações baseadas em bancos de dados, de acordo com o paradigma de Orientação a Objetos (OO) envolve a criação de modelos descrevendo as entidades de negócio (ou entidades do domínio da aplicação), seus relacionamentos, seu comportamento e as regras que determinam a sua consistência com o mundo real modelado.

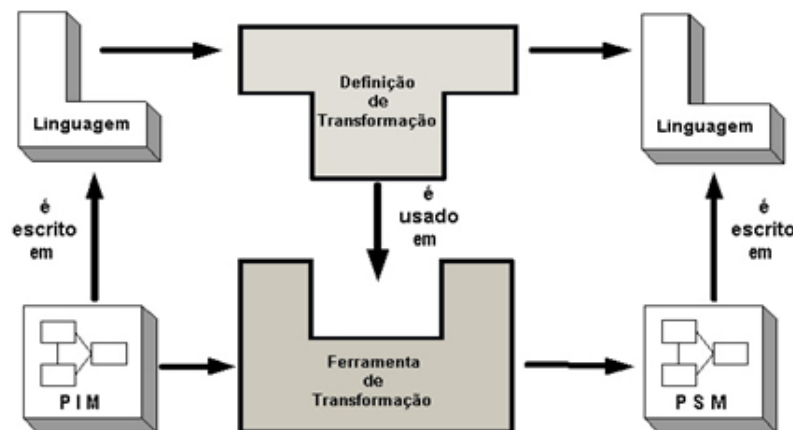
Esses modelos devem ser implementados em uma determinada plataforma computacional, ou seja, em aplicações baseadas em Bancos de Dados (BD) é necessário converter a especificação conceitual de dados, relacionamentos, funções e regras de negócio para a linguagem computacional disponível na plataforma de SGBD, que envolve, tipicamente, a linguagem SQL. Os dados referem-se às propriedades das entidades de negócio e as regras são utilizadas para impor restrições sobre eles [10][55]. No contexto de BD, as regras de negócio são tratadas como Restrições de Integridade (RI) e especificam condições que devem ser atendidas para garantir a consistência entre o esquema de BD e o mundo real representado [52].

Na abordagem tradicional de desenvolvimento de aplicações, os modelos conceituais são mapeados manualmente para as linguagens computacionais específicas, de modo que, caso seja necessário alterar a plataforma de implementação, todo o trabalho manual de mapeamento deve ser refeito. Em contraposição a esta abordagem existe a estratégia baseada em MDA (Model Driven Architecture), que é um framework que utiliza UML e outros padrões propostos pela OMG (Object Management Group) para o desenvolvimento de software baseado em modelos abstratos [29].



Esse desenvolvimento dirigido por modelos permite especificar os conceitos relevantes do domínio da aplicação, as regras de negócio, seus relacionamentos e semântica usando várias linguagens de modelagem conceitual. Como mostra a Figura 1, a estratégia de transformação de modelos facilita o mapeamento automático de modelos independentes de plataforma (PIM - Platform Independent Model) para modelos voltados para plataformas específicas (PSM - Platform Specific Model), podendo ser utilizada para gerar esquemas de banco de dados e restrições de integridade sobre eles. A transformação entre modelos deve ser feita por uma ferramenta de transformação que irá utilizar uma definição de transformação (mapeamento) entre os elementos da linguagem de origem para a linguagem alvo. A definição de transformação representa a vantagem da MDA em relação à abordagem tradicional de desenvolvimento: em caso de haver a necessidade de alterar a plataforma de implementação do modelo (por exemplo, trocar o banco de dados que está sendo utilizado), basta que a definição de transformação seja alterada para incorporar elementos da nova plataforma.

A OCL é uma linguagem criada especificamente para descrever expressões e regras complexas, de forma a complementar modelos especificados em UML [34]. O objetivo da criação da OCL pela OMG foi, além de facilitar o entendimento dos modelos construídos com UML, aumentar o poder de expressão de restrições (ou regras) destes modelos, fato que é reconhecidamente uma limitação dos elementos padrões da UML. A OCL permite especificar diversos tipos de regras, como invariantes, valores iniciais de atributos, e computação de valores derivados.



**Figura 4.2:** Visão Geral da Abordagem MDA [29]

Com a especificação de modelos UML contendo regras de negócio escritas em OCL e a utilização das técnicas de MDA, problemas comuns no desenvolvimento de aplicações para bancos de dados podem ser atenuados. A portabilidade da aplicação será facilitada, pois as técnicas de MDA permitem que um mesmo modelo fonte possa ser transformado em modelos para plataformas variadas, bastando alterar a definição de transformação correspondente à linguagem alvo. Também a produtividade no desenvolvimento

e na manutenção de esquemas de bancos de dados é beneficiada, pois as ferramentas de transformação e geração de código permitem que modificações nas regras de negócio escritas em OCL sejam automaticamente transformadas para o dialeto SQL específico do SGBD adotado [29].

O objetivo deste projeto de mestrado é investigar a utilização de OCL para a modelagem de regras de negócio e a conversão delas para SQL através de uma ferramenta de transformação automática baseada em MDA. Com isso, pretende-se reduzir significativamente o impacto dos problemas citados no desenvolvimento de aplicações em bancos de dados.

#### **4.5.4 Melhoria de Processos de Governança de TI**

**Discente:** Fabiana Freitas Mendes - fabianafreitas@inf.ufg.br

**Nível:** Mestrado

De acordo com Brooks [8], construir software é especialmente difícil por causa das dificuldades inerentes ao software, tais como a complexidade e a invisibilidade, e também das dificuldades presentes na produção, mas que não fazem parte da natureza do software, como, por exemplo, as dificuldades tecnológicas.

Dentre os problemas apresentados por Brooks podemos destacar o fato de o software ser invisível, ou seja, não é espacialmente representável. Assim, são necessários vários diagramas para se conseguir um entendimento parcial do sistema. Além disso, ele sofre muita pressão para ser modificado, mesmo durante seu desenvolvimento e, ao contrário do que se possa imaginar, modificá-lo não é uma tarefa fácil.

Por essas e outras razões, produzir software com qualidade tem sido um desafio para a comunidade de Engenharia de Software. Uma das maneiras de se melhorar a qualidade de software é por meio da melhoria de processos de software.

O uso de processos auxilia o desenvolvimento de software, principalmente quando se está lidando com membros inexperientes. Isso porque a utilização de processos confere disciplina à equipe que desenvolve software. A utilização de processos também permite que as tarefas sejam realizadas sempre de uma mesma maneira. Assim, caso se tenha uma boa maneira de desempenhar uma certa tarefa, o uso do processo garantirá que esta tarefa será sempre realizada desta mesma boa maneira. Por todos estes motivos, o estudo de processos tem sido enfatizado pela Engenharia de Software.

O estudo de processos tornou possível o surgimento de vários modelos de qualidade que descrevem as melhores práticas das equipes que têm desenvolvido software em todo o mundo. Como exemplo de modelos de qualidade pode-se citar o CMMI (Capability Maturity Model Integration) [57] e o MPS.BR (Melhoria de Processos do Software Brasi-

leiro) [41]. Além destes, podemos citar as normas ISO/IEC 12207 [24] e ISO/IEC 15504 [21]. Todos estes documentos refletem a experiência mundial em utilização de processos na construção de software. Estes modelos, entretanto, consideram apenas os processos do nível técnico da organização, ou seja, aqueles que estão diretamente envolvidos com a produção de software.

Existe, entretanto, um outro nível de processos na organização que são importantes para garantir a infra-estrutura necessária para a execução dos processos mais diretamente relacionados à produção de software. Estes processos estão relacionados com a Governança de TI da empresa. Existem diversos modelos de qualidade que tratam estes processos como ITIL [39], Cobit [25], ISO/IEC 20000 [23] e ISO/IEC 17799 [22].

O projeto de pesquisa aqui representado tem por objetivo desenvolver um método de implantação de Melhorias de Processos de Governança de TI em organizações, considerando a existência de outros projetos de melhoria de processo na organização.

O projeto de pesquisa de mestrado de Melhoria de Processos da Governança de TI (MPGTI) está inserido no contexto de Melhoria da TI (MTI). Entende-se que a MTI deve ser desenvolvida considerando quatro aspectos: Melhoria da Governança, Melhoria da Governança de TI, Melhoria Operacional e Integração Entre as Abordagens de Melhoria, conforme pode ser observado na Figura 4.3.



**Figura 4.3:** Melhoria da Tecnologia da Informação

A Governança trata de assuntos como cultura organizacional que é essencial para garantir que as iniciativas de melhoria realizadas na organização sejam, de fato, institucionalizadas.

A Melhoria operacional trata da melhoria dos processos de software da organização e da melhoria dos produtos entregues pela TI.

A Governança de TI (GTI) trata de assuntos como o suporte e a entrega de serviços da TI que atendam às necessidades da organização. Assim, busca-se maximizar o valor agregado da TI a organização por meio, por exemplo, da minimização de riscos.

Para cada um destes tipos de melhoria é possível criar uma abordagem de implantação de melhoria que considere as peculiaridades da área alvo. Estes métodos, entretanto, precisam ser integrados a fim de que se possa trabalhar com uma abordagem única de melhoria na organização.

Além disso, é importante que sejam desenvolvidas ferramentas específicas para apoiar um projeto de MTI em uma organização.

### **4.5.5 As Influências e a Gestão da Cultura Organizacional na Melhoria de Processos**

**Discente:** Patrícia Gomes Fernandes - patriciagomes@inf.ufg.br

**Nível:** Mestrado

Na sociedade atual o uso de software no apoio a diversas atividades tem sido cada vez maior. A gama de atividades apoiadas varia desde atividades mais simples até atividades que envolvem riscos para a vida humana. Diante deste cenário de crescimento de uso e criticidade dos softwares utilizados pela sociedade, as organizações e pessoas responsáveis pelo desenvolvimento e manutenção de tais softwares têm tido cada vez maiores preocupações com a qualidade destes produtos e dos serviços associados [28]. Considerando que a qualidade dos produtos de software está intimamente relacionada com a qualidade dos processos que os geram e mantêm [56], a melhoria de processos de software (MPS) tem sido uma abordagem que está se difundindo entre tais organizações.

Todavia, a implementação de melhorias nos processos das organizações de software não representa uma tarefa fácil. Por se tratar da implementação de mudanças no modo de trabalho das organizações, envolve uma mudança na cultura organizacional. Vários trabalhos na literatura de MPS existente corroboram esta afirmação. Por exemplo, ao analisar orientações dadas no Guia de Grupo de Processo de Engenharia de Software, publicado pelo Software Engineering Institute (SEI) em 1990, um de seus autores adiciona informações que também são importantes mas que não foram considerados. Dentre estas informações há considerações sobre a cultura organizacional. Questões sobre a associação entre a cultura e a resistência à MPS foram colocadas [46].

Em relação à modelos e padrões de qualidade e à certificação, os trabalhos de Ibrahim [19] e de Matsubara [32] enfatizam a necessidade de mudança cultural para que os benefícios decorrentes da MPS sejam realmente desfrutados. Considerando ainda os modelos, o estudo de [35], que avalia um modelo de cultura organizacional contra o *Capability Maturity Model* (CMM), pode sugerir que há suposições culturais neles embutidas que têm sido utilizados como referência para a MPS. Em um estudo de caso sobre a implementação de uma das versões do CMM, o *Capability Maturity Model Integration* (CMMI), Dutton enfatiza que o sucesso da MPS depende fortemente da cultura da organização, assim como da coragem e habilidade dos seus líderes [14]. Neste estudo há uma breve seção sobre os valores da organização em questão, com o argumento que fornecem um forte fundamento para a MPS.

Já em relação às diferentes filosofias das metodologias existentes, Boehm e Turner fazem uma análise comparativa entre métodos ágeis e direcionados por planos, apontando a cultura como um fator que diferencia organizações de modo que um ou outro método seja mais apropriado [7].

Alguns trabalhos colocam a cultura organizacional como fator crítico, como por exemplo em [26], ou como obstáculo, como em [4], cujas conclusões evidenciam que, embora os gerentes pareçam comprometidos com a MPS, muitas vezes a cultura organizacional não a valoriza. No trabalho apresentado em [36], o estabelecimento de mecanismos para fazer a MPS parte da cultura da organização é colocado como um dos fatores críticos de sucesso para a MPS. Além disso, em [19] a cultura é colocada também como um dos fatores que podem levar de projetos de MPS ao insucesso. Em um estudo de caso sobre a implementação do *Capability Maturity Model Integration* (CMMI), um modelo de qualidade, Dutton enfatiza que o sucesso da MPS depende fortemente da cultura da organização, assim como da coragem e habilidade dos seus líderes [14]. Neste estudo há uma breve seção sobre os valores da organização em questão, com o argumento que fornecem um forte fundamento para a MPS.

Em um outro trabalho, Aaen sugere que a MPS envolve a criação de uma cultura, diferindo assim dos demais autores, ao colocar o intento da MPS como sendo a construção de uma cultura que apóie métodos, práticas e procedimentos efetivos, a fim de que se tornem o modo de fazer negócio de uma organização [1].

Assim, muitos trabalhos reconhecem a importância e a influência da cultura organizacional sobre as iniciativas de MPS. Todavia, faltam trabalhos que analisem mais profundamente a natureza das influências da cultura organizacional sobre as iniciativas de MPS, bem como que proponham orientações, seja na forma de diretrizes, estratégias, métodos ou técnicas, que possam orientar tais iniciativas, de modo que a cultura organizacional possa ser utilizada em seu favor.

Neste sentido, o presente trabalho tem por propósito o estudo da literatura existente sobre cultura organizacional, com o intuito de identificar os elementos componentes da cultura das organizações e seus relacionamentos com a MPS, assim como de propor orientações sobre a avaliação e a gerência da cultura, de modo que possa favorecer os efeitos da MPS em organizações que a estiverem adotando.

Adicionalmente, o estudo proposto neste projeto pretende aplicar as orientações propostas em organizações que estão implementando melhorias em seus processos de modo a levantar dados sobre como a cultura pode afetar os efeitos da MPS.

## 4.6 Metodologias de Integração de Informações Armazenadas em Bancos de Dados / Web

**Coordenadores:** Dr João Carlos da Silva - jcs@inf.ufg.br, Ms. Dirson Santos de Campos - dirson@inf.ufg.br, Dr. Fábio Nogueira de Lucena - fabio@inf.ufg.br

Existem muitas fontes de informação no mundo atual. Esta informação é armazenada das mais variadas formas: documentos na web, imagens, sons, diagramas, mapas, tabelas e bancos de dados de variados graus de complexidade. O objetivo principal deste projeto é o estudo de metodologias para integração de dados provenientes de fontes heterogêneas (bancos de dados, web, etc.), de maneira a identificar mecanismos que possam ser aplicados ao processo de identificação e integração de informações semanticamente equivalentes, armazenadas em meios estruturados e/ou semi-estruturados.

### 4.6.1 Utilização de Mapas Mentais na Engenharia de Requisitos

**Discente:** Luciana Nishi

**Nível:** Mestrado

Em “Requirements - Led Project Management: Discovering David’s Slingshot” [47] é ressaltado que quando se compreende a verdadeira natureza do negócio, o melhor a ser feito é adicionar um novo passo na atividade de requisitos: “Inventar algo melhor”.

E a citação do projetista de software Dewys Lasdon reintera isto: “Nosso trabalho é dar ao cliente, no devido prazo e custo, não o que ele precisa, mas o que ele nunca imaginou precisar; e quando ele conseguir, ele reconhecerá como sendo algo que ele sempre precisou ter” (leia mais sobre Lasdon em [44]).

Todo desenvolvedor tem como uma de suas principais metas construir o sistema desejado pelo cliente. Este intuito é valoroso mas difícil de ser realizado, pois compreender a natureza do negócio não é uma tarefa simples como aparenta. Existem várias formas/técnicas para compreender a natureza do negócio (entrevistas, storyboards e diagramas, dentre outras), mas estas formas não garantem que o analista de requisitos compreende o negócio de maneira satisfatória. Ele poderá ter a ilusão de que realmente compreendeu, inventar algo e não ser promissor pois, na verdade não conhecia a "verdadeira" natureza do negócio.

O artigo Requirements Engineering as a driver for innovations [27], apresenta a Engenharia de requisitos como apoiador na transformação de idéias criativas em produto com inovações valiosas, isto é, que a engenharia de requisitos pode ser vista como um

processo criativo onde stakeholders e projetistas trabalham conjuntamente para criar idéias para novos sistemas que são eventualmente denominados requisitos.

A intenção é propor algo para que pessoas envolvidas no processo de desenvolvimento compreendam as necessidades dos clientes e com isso a natureza do sistema. Facilitar a comunicação com os stakeholders, sem que os mesmos tenham que conhecer UML ou modelos matemáticos sem, é claro, perder em precisão.

Inicialmente pesquisas sobre as ferramentas que criam mapas mentais foi realizada, estas ferramentas são utilizadas no estudo para exemplificar as melhorias no uso de mapas na engenharia de requisitos. Exemplificar, como os mapas mentais podem beneficiar os engenheiros de requisitos tendo como referência a IEEE-830[20] e o MPS-BR[53]. Aplicar em um estudo de caso e finalmente relacionar os benefícios e os malefícios do uso de mapas mentais no registro de requisitos.

Investigar como mapas mentais podem ser empregados para a compreensão e o registro de requisitos, isto é, investigar uma forma criativa de "facilitar" a vida tanto de quem escreve quanto de quem lê especificações de requisitos.

## 4.7 Teste de Software

**Coordenador:** Dr Auri Marcelo Rizzo Vincenzi - auri@inf.ufg.br

Este projeto é focado no desenvolvimento de recursos de teste e validação para programas Orientado a Objeto. No contexto deste projeto, atenção especial é dada aos critérios de teste estruturais e baseados em mutação, juntamente com o desenvolvimento de ferramentas de apoio à realização dos testes.

### 4.7.1 Teste Estrutural de Software em Dispositivos Móveis Utilizando J2ME

**Discente:** Gilcimar de Deus

**Nível:** Mestrado

Sob orientação do Prof. Dr. Auri Marcelo Rizzo Vincenzi, o projeto nomeado Teste Estrutural de Software em Dispositivos Móveis Utilizando J2ME, visa mostrar o quanto é importante o teste estrutural para a revelação de erros e na criação de casos de teste.

A pesquisa pretende mostrar que o teste estrutural com o apoio da ferramenta JaBUTi/ME são mais eficientes na revelação de erros que as outras técnicas conhecidas. Uma dessas técnicas é a baseada no conhecimento empírico, também conhecida como 'ad

hoc'. A outra técnica que está sendo comparada é a funcional que avalia o programa como uma caixa preta, sem saber o que se passa em sua estrutura interna.

A disciplina de testes na Engenharia de Software conta com diversas ferramentas que apóiam as mais diferentes técnicas. Como essa disciplina é crítica além de muito importante para o desenvolvimento de um software com qualidade, é necessário medir a eficácia da técnica que está sendo aplicada. Um mundo ainda pouco explorado é o dos aplicativos J2ME, ou seja, dos programas que compoem a maioria dos aparelhos celulares que nos cercam nos dias de hoje. O objetivo da pesquisa é avaliar a ferramenta JaBUTi/ME na revelação de erros para esses tipos de aplicativos.

Através da replicação de pacotes de experimentação é possível coletar informações estatísticas sobre a execução de cada uma das técnicas de teste. Um pacote de experimentação com quatro programas J2ME foi criado e será replicado três vezes, cada replicação com trinta participantes. Curso de teste de software será oferecido para a comunidade e através dos testes feitos no curso, serão coletadas as informações de execução dos mesmos. Todos os programas foram coletados em sítios conhecidos da internet, como sourceforge.net. Um programa será utilizado para treinamento dos participantes nas técnicas, ad hoc, funcional e estrutural. Os outros três programas serão utilizados pelos participantes para aplicarem os conhecimentos adquiridos no curso em cada técnica. Será utilizada a estratégia de seleção fatorial fracional-randomizado para coleta das informações, onde cada indivíduo não aplica todas as técnicas em todos os programas.

Ao final das replicações será feita uma análise dos dados e observado se a ferramenta JaBUTi/ME realmente ajuda na revelação de defeitos nos programas J2ME.

## 4.8 Identificação de Relacionamento entre Objetos de Bibliotecas Digitais

**Coordenadores:** Dr Thierson Couto Rosa - thierson@inf.ufg.br e Dr Cláudio Nogueira de Meneses - claudio@inf.ufg.br

Este projeto visa a obtenção automática de informações sobre relacionamentos entre objetos em uma biblioteca digital de artigos científicos. O objetivo principal é identificar relacionamentos não óbvios entre objetos (autores, artigos, veículos de publicação— anais, revistas indexadas). Por exemplo: a) identificar *comunidades latentes* – grupos de autores que nunca foram co-autores de uma obra, mas que possuem trabalhos fortemente relacionados; b) identificar a evolução de relacionamentos entre subáreas de pesquisa ao longo dos anos e, em consequência disto, descobrir possíveis novos nichos de publicação para autores de uma sub-área. De acordo com Sanderson, existem poucos trabalhos



que utilizam técnicas de mineração de dados para inferir estes tipos de relacionamentos em bibliotecas digitais. Inicialmente este projeto visa inferir estas informações a partir da biblioteca digital da ACM<sup>5</sup>, utilizando algoritmos de mineração de dados e técnicas de otimização. Em uma primeira etapa, algoritmos de mineração serão utilizados para detectar relacionamentos entre os vários objetos. Em uma segunda etapa, técnicas de otimização serão utilizadas para minimizar a chance de que os relacionamentos descobertos sejam falsos.

## 4.9 Resolução de Ambigüidade de Nomes em Bibliotecas Digitais

**Coordenadores:** Dr Thierson Couto Rosa - thierson@inf.ufg.br, Dr Cláudio Nogueira de Meneses - claudio@inf.ufg.br e Eric Patrick Rodrigues de Oliveira (UCG)

É comum o nome de um autor de artigos científicos aparecer escrito de diferentes modos em suas publicações ou em citações feitas a suas obras. Geralmente, as formas distintas de escrita se devem à abreviação ou à omissão de um ou mais componentes do nome. Por exemplo, o nome Thierson Couto Rosa pode aparecer abreviado de diversas formas tais como: T. C. Rosa, T. Couto R., Thierson C. R., Thierson C. Rosa. Este fato dificulta uma série de operações importantes em uma biblioteca digital, como: identificar todas as obras de um mesmo autor, identificar todas as formas distintas utilizadas para fazer referência a um autor, e resolver a qual autor uma citação se refere, em casos em que dois ou mais autores possuem alguma forma idêntica escrita de seus nomes. Problemas semelhantes ocorrem com nomes de conferências e de revistas indexadas que aparecem com partes abreviadas e, às vezes com partes omitidas. As propostas de solução deste problema envolvem a utilização de técnicas de mineração de dados e de teoria de grafos e abordagens semi-automáticas de maneira isolada. Pretendemos combinar estas e outras técnicas para obter soluções melhores do que cada técnica poderia alcançar separadamente. Objetivamos ainda estender os resultados obtidos para outros tipos de ambigüidades, por exemplo, em nomes de conferências e de revistas indexadas.

---

<sup>5</sup>Association for Computing Machinery - [www.acm.org](http://www.acm.org)

## 4.10 Propostas de Medidas de Relevância de Conferências

**Coordenadores:** Dr Thierson Couto Rosa - thierson@inf.ufg.br, Dr Cláudio Nogueira de Meneses - claudio@inf.ufg.br

Anais de conferências são meios importantes de publicação artigos científicos em diversas áreas. Os principais pesquisadores de uma área de pesquisa costumam publicar seus trabalhos em revistas indexadas e/ou anais de conferências. Revistas indexadas são avaliadas pelo *fator de impacto*, que é uma medida bem aceita no meio acadêmico. Contudo, em certas áreas não existem medidas de avaliação de relevância de conferências. Um exemplo disto é a área de ciência da computação. Este trabalho visa propor diversas medidas de relevância de conferências baseadas em distintos critérios: idade da conferência, a importância dos pesquisadores que a organizam, a quantidade e qualidade de outras conferências e revistas indexadas que tenham trabalhos que citam artigos da conferência etc. Nosso trabalho pretende ainda fazer um estudo de correlação e combinação das medidas. O resultado deste trabalho visa servir de subsídio a órgãos governamentais de apoio à pesquisa, como a CAPES, que necessitam medir as relevâncias de conferências.

### Referências Bibliográficas

- [1] AAEN, I. **Challenging Software Process Improvement by Design.** In: PROCEEDINGS OF THE 10TH EUROPEAN CONFERENCE ON INFORMATION SYSTEMS, INFORMATION SYSTEMS AND THE FUTURE OF THE DIGITAL ECONOMY, ECIS 2002, p. 379–390, 2002.
- [2] ALLEMAND, J. N. C. **Serviço baseado em semântica para descoberta de recursos em grade computacional.** Master's thesis, Universidade de Brasília, UNB, 2006.
- [3] ASTRACHAN, O; BRUCE, K; KOFFMAN, E; KÖLLING, M; REGES, S. **Resolved: Objects early has failed.** SIGCSE '05: Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education, St. Louis, Missouri, 2005.
- [4] BADDOO, N; HALL, T; WILSON, D. **Implementing a People Focused SPI Programme.** In: 11TH EUROPEAN SOFTWARE CONTROL AND METRICS CONFERENCE AND THE THIRD SCOPE CONFERENCE ON SOFTWARE PRODUCT QUALITY, 2000.

- [5] BAEZA-YATES, R; RIBEIRO-NETO, B; OTHERS. **Modern information retrieval**. Addison-Wesley Harlow, England, 1999.
- [6] BERNERS-LEE, T; HENDLER, J; LASSILA, O. **The semantic web: Scientific american**. Scientific American, ?(1), May 2001.
- [7] BOEHM, B; TURNER, R. **People Factors in Software Management: Lessons From Comparing Agile and Plan-Driven Methods**. CROSSTALK: The Journal of Defense Software Engineering, 16(12):4–8, 2003.
- [8] BROOKS, F. P. **No silver bullet - essence and accident in software engineering**. In: PROCEEDINGS OF THE IFIP TENTH WORLD COMPUTING CONFERENCE, p. 1069–1076, 1986.
- [9] DA SILVA, M. A. V; ET AL.. **Modi - a proposal of a visual tool to simulate and synthesize software applied to embedded systems**. DEMIC / FEEC / UNICAMP, 2001.
- [10] DATE, C. J. **What, not How - The Business Rules Approach to Application Development**. Addison Wesley, 2001.
- [11] DE LIMA, V. L. S; DAS GRAÇAS VOLPE NUNES, M; VIEIRA, R. **Desafios do processamento de línguas naturais**. SEMISH - Seminário Integrado de Software e Hardware, ?(34):1, 2007.
- [12] DE SOUSA, B. B. **Um serviço de metadados integrado ao middleware de grade mag**. Master's thesis, PPGEE/UFMA, July 2006.
- [13] DUBIELEWICZ, I; HNATKOWSKA, B; HUZAR, Z; TUZINKIEWICZ, L. **Feasibility analysis of mda-based database design**. In: DEPCOS-RELCOMEX '06: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON DEPENDABILITY OF COMPUTER SYSTEMS, p. 19–26, Washington, DC, USA, 2006. IEEE Computer Society.
- [14] DUTTON, J. L. **A CMMI Case Study: Process Engineering vs Culture and Leadership**. CROSSTALK: The Journal of Defense Software Engineering, 15(12):11–14, 2002.
- [15] FORMAN, G. **An extensive empirical study of feature selection metrics for text classification**. The Journal of Machine Learning Research, 3:1289–1305, 2003.
- [16] GOLDCHLEGER, A; KON, F; GOLDMAN, A; FINGER, M; BEZERRA, G. C. **Integrate: Object-oriented grid middleware leveraging idle computing power**

- of desktop machines.** Concurrency and Computation: Practice and Experience, 16(5):449–459, March 2004.
- [17] GUYON, I; CLOPINET, C. **Causal feature selection.** Computational Methods of Feature Selection, 7:63–85, 2007.
- [18] GUYON, I; ELISSEEFF, A. **An introduction to variable and feature selection.** The Journal of Machine Learning Research, 3:1157–1182, 2003.
- [19] IBRAHIM, L. **A Process Improvement Commentary.** CROSSTALK: The Journal of Defense Software Engineering, 21(8):26–29, 2008.
- [20] IEEE-SA STANDARDS BOARD. **IEEE Std 830-1998: IEEE Recommended Practice for Software Requirements Specifications.** Disponível em <<http://ieeexplore.ieee.org/iel4/5841/15571/00720574.pdf?arnumber=720574>>, acessado em janeiro de 2007, 1998.
- [21] ISO/IEC. **Iso/iec tr 15504 software process assessment.** Genebra, 2004.
- [22] ISO/IEC. **Iso 17799 information technology – security techniques – code of practice for information security management,** 2005.
- [23] ISO/IEC. **Iso 20000 information technology – service management,** 2005.
- [24] ISO/IEC. **Iso/iec 12207 information technology – software life cycle processes,** 2005.
- [25] ITGI. **Cobit 4.1 control objectives management guidelines maturity models.** IT Governance Institute, 2007.
- [26] KASSE, T; MCQUAID, P. A. **Factors Affecting Process Improvement Initiatives.** CROSSTALK: The Journal of Defense Software Engineering, 13(8):4–8, 2000.
- [27] KAUPPINEN, M; SAVOLAINEN, J; MÄNNISTÖ, T. **Requirements engineering as a driver for innovations.** re, 0:15–20, 2007.
- [28] KITCHENHAM, B; PFLEEGER, S. L. **Software Quality: The Elusive Target.** IEEE Software, 13:12–21, 1996.
- [29] KLEPPE, A; WARMER, J; BAST, W. **MDA Explained: The Model Driven Architecture: Practice and Promise.** Addison-Wesley, 1st edition, 2003.
- [30] KUMAR, A. N. **The effect of closed labs in computer science i: An assessment.** Journal of Computing in Small Colleges, 2003.

- [31] MANNING, C; RAGHAVAN, P; SCHÜTZE, H. **Introduction to Information Retrieval**. Cambridge University Press, 2008.
- [32] MATSUBARA, T. **Process Certification: A Double-Edged Sword**. IEEE Software, 17(6):104–105, 2000.
- [33] MCGETTRICK, A; BOYLE, R; IBBETT, R; LLOYD, J; LOVEGROVE, G; MANDER, K. **Grand challenges in computing: Education - a summary**. The Computer Journal, 2005.
- [34] MORGAN, T. **Business Rules and Information Systems: Aligning IT with Business Goals**. Addison Wesley, 2001.
- [35] NGWENYAMA, O; NIELSEN, P. A. **Competing Values in Software Process Improvement: An Assumption Analysis of CMM From an Organizational Culture Perspective**. IEEE Transactions on Engineering Management, 50(1):100–112, 2003.
- [36] NIAZI, M; WILSON, D; ZOWGHI, D. **A Maturity Model for the Implementation of Software Process Improvement: an Empirical Study**. The Journal of Systems and Software, 74:155–172, 2005.
- [37] OBJECT MANAGEMENT GROUP. **Mda guide version 1.0.1**. <http://www.omg.org/docs/omg/030601.pdf>, Outubro 2007.
- [38] OBJECT MANAGEMENT GROUP. **The object management group**. <http://www.omg.org>, Janeiro 2007.
- [39] OCG. **Introduction to itil**. TSO, 2005.
- [40] PANSANATO, L. T. E. **Um modelo de navegação exploratória para a infraestrutura da Web Semântica**. Tese de doutorado, USP - São Carlos, Novembro 2007.
- [41] PARA PROMOÇÃO DA EXCELÊNCIA DO SOFTWARE BRASILEIRO SOFTEX, A. **MPS.BR - Guia Geral v 1.2**. [http://www.softex.br/mpsbr/\\_guias/](http://www.softex.br/mpsbr/_guias/). Último acesso em 24 de Outubro de 2007.
- [42] PAULSON, B; HAMMOND, T. **Marqs: Retrieving sketches using domain- and style-independent features learned from a single example using a dual-classifier**. IWUMUI, 2007.
- [43] PETERS, J; PEDRYCZ, W. **Engenharia de Software**. Campus, 2001.

- [44] PETERS, T. **The Circle of Innovation**. Vintage, London, 1999.
- [45] R., D. **Magic paper: Sketch-understanding research**. Massachusetts Institute of Technology., 2007.
- [46] RIFKIN, S. **What I would do differently if I wrote the SEPG Guide today**. Disponível em [www.uces.csulb.edu/spin/media/word/stan\\_rifkin.doc](http://www.uces.csulb.edu/spin/media/word/stan_rifkin.doc). Último acesso em 05/10/2008., 2002.
- [47] ROBERTSON, S; ROBERTSON, J. **Requirements-Led Project Management: Discovering David's Slingshot**. Addison-Wesley Professional, 2004.
- [48] ROSA, T. C. **Uso de Apontadores na Classificação de Documentos em Coleções Digitais**. Tese de doutorado, Universidade Federal de Minas Gerais, Dezembro 2007.
- [49] ROURE, D. D. **Semantic Grid Community Portal**. <http://www.semanticgrid.org/>, último acesso em Novembro de 2007, 2006.
- [50] SANTIAGO, R; S., D. R. L. **Ferramenta de apoio ao ensino de algoritmos**. Seminário de Computação, 2004.
- [51] SEBASTIANI, F. **Machine learning in automated text categorization**. ACM Computing Surveys (CSUR), 34(1):1–47, 2002.
- [52] SILBERSCHATZ, A; KORTH, H. F; SUDARSHAN, S. **Sistema de Banco de Dados**. Editora Campus, 5th edition, 2003.
- [53] SOCIEDADE SOFTEX. **Mps.BR - Melhoria de Processo do Software Brasileiro - Guia Geral (Versão 1.1)**. Disponível em <[http://www.softex.br/mpsbr/\\_guias/MPS.BR\\_Guia\\_Geral\\_V1.1.pdf](http://www.softex.br/mpsbr/_guias/MPS.BR_Guia_Geral_V1.1.pdf)>, acessado em Fevereiro de 2007, 2006.
- [54] STAVRIANOU, A; ANDRITSOS, P; NICOLOYANNIS, N. **Overview and semantic issues of text mining**. SIGMOD Record, 36(3):23–34, 2007.
- [55] THE BUSINESS RULES GROUP. **Defining Business Rules - What Are They Really? rev. 1.3**. <http://www.BusinessRulesGroup.org>, acessado em setembro de 2008, 2008.
- [56] TYRRELL, S. **The Many Dimensions of the Software Process**. Crossroads, 6(4):22–26, 2000.
- [57] UNIVERSITY, S. E. I. C. M. **CMMI for Development v 1.2**. <http://www.sei.cmu.edu/cmmi/>. Último acesso em 29 de Outubro de 2007.

- [58] VIDAL, A. C. T. **Abordagem Semântica Aplicada à Integração e Gerenciamento de Recursos e Aplicações em Grades Computacionais**. PhD thesis, Escola Politécnica da Universidade de São Paulo, 2007.
- [59] W3C. **W3C Semantic Web Activity**. <http://www.w3.org/2001/sw/>, último acesso em Setembro de 2008, 2001.