

Universidade Regional de Blumenau  
Centro de Ciências Exatas e Naturais  
Departamento de Sistemas e Computação



# **Gestão do Conhecimento: Aplicação em Data Mining utilizando a Teoria dos Conjuntos Aproximativos para geração do Capital Intelectual**

**Acadêmico:** Sidnei Schmitt  
**Prof. Orientador:** Dr. Oscar Dalfovo

# Roteiro da apresentação

- ✓ Introdução
  - Objetivos do Trabalho
  
- ✓ Fundamentação teórica
  - Gestão do Conhecimento
  - Data Mining
  - Teoria dos Conjuntos Aproximativos

# Roteiro da apresentação

- ✓ **Desenvolvimento do Trabalho**
  - Especificação
  - Tecnologias e ferramentas utilizadas
  - Operacionalidade
  - Trabalhos Correlatos
  
- ✓ **Considerações Finais**
  - Conclusão
  - Extensões

# Introdução

## ✓ Contextualização

- o conhecimento é a informação apropriada e interpretada pelo ser humano;  
(Dalfovo, 2007)
- o avanço tecnológico facilitou o acúmulo de dados e informações para as organizações;  
(Prass, 2004)
- as técnicas de *Data Mining* buscam extrair as informações do conhecimento em tempo hábil e competitivo.  
(Almeida, 2004)

# Introdução

## ✓ Objetivo geral

estudo e o desenvolvimento de uma aplicação em gestão do conhecimento utilizando *Data Mining* baseado na Teoria dos Conjuntos Aproximativos (TCA), gerando o capital intelectual para as organizações.

# Introdução

- ✓ **Objetivos específicos**
  - demonstrar o potencial do DM para classificação e segmentação de dados baseado na TCA;
  - selecionar os perfis mais adequados dos profissionais cadastrados como capital intelectual;
  - demonstrar graficamente o resultado da mineração dos dados.

# Fundamentação Teórica

- ✓ **Gestão do conhecimento**
  - o mundo atual nos oferece milhares de informações através de diversos meios;  
(Gimenes, 2000)
  - controlar estas informações passou a ser um diferencial para que se possa atingir objetivos;  
(Bernardes, 2001)
  - a gestão do conhecimento é uma metodologia que determina a vantagem competitiva de uma organização;  
(Dalfovo, 2007)

# Fundamentação Teórica

- ✓ **Gestão do conhecimento**
  - a longevidade das organizações está ligada à sua sensibilidade para aprender e se adaptar de forma mais rápida que os concorrentes;  
(Gonçalves e Gonçalves, 2001)
  - para adquirir o conhecimento, as organizações investiram muito em tecnologia e armazenamento de dados.  
(Madeira, 2003)





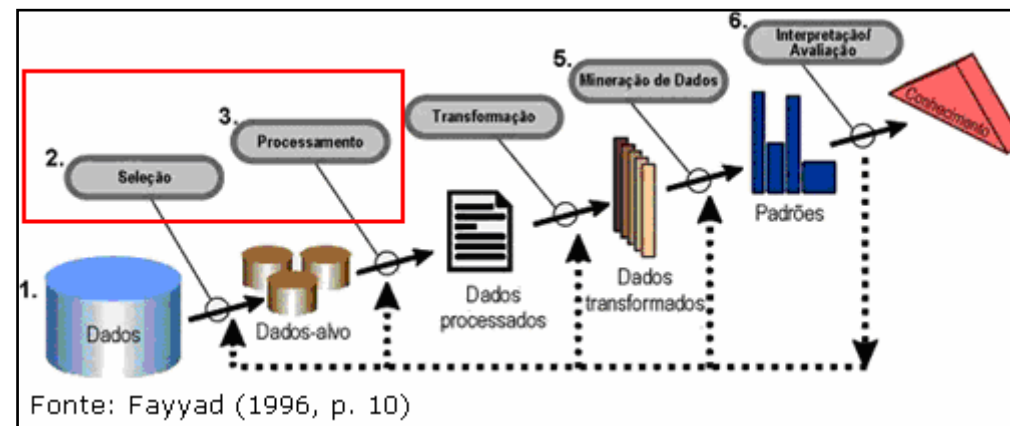
# Fundamentação Teórica

## ✓ Data Mining

- ferramenta capaz de analisar automaticamente bases de dados para obter o conhecimento;
- composto por técnicas e algoritmos especializados que buscam transformar os dados armazenados em informações úteis.
- é uma das etapas do KDD;  
(*Knowledge Discovery in Databases*)

# Fundamentação Teórica

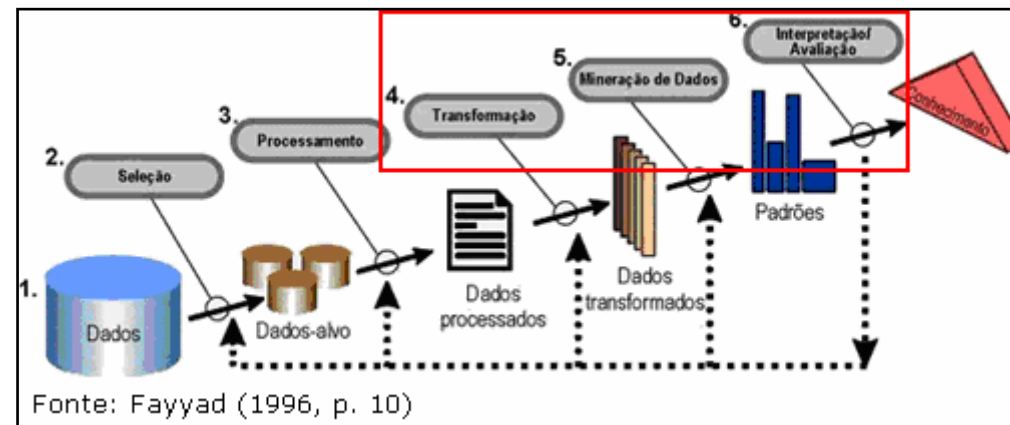
## ✓ Etapas do KDD



- **seleção de dados:**  
identificar e selecionar fontes de dados;
- **pré-processamento e limpeza:**  
tratar inconsistências;

# Fundamentação Teórica

## ✓ Etapas do KDD



- **transformação:**  
conversão do formato dos dados;
- **mineração de dados:**  
procura por padrões existentes baseado em uma técnica;
- **interpretação e avaliação:**  
validação dos resultados pelos analistas.

# Fundamentação Teórica

- ✓ Algumas técnicas de Data Mining
  - raciocínio baseado em casos (MBR):  
procura os vizinhos mais próximos e combinam seus valores;
  - redes neurais artificiais:  
segue a analogia com os neurônios do cérebro humano;
  - árvores de decisão:  
divide os registros montando a árvore de busca;
  - algoritmos genéticos:  
busca encontrar os melhores conjuntos de parâmetros.

# Fundamentação Teórica

- ✓ Teoria dos conjuntos aproximativos
  - desenvolvida por Zdzislaw Pawlak no início da década de 80;
  - ferramenta matemática de redução de dimensionalidade;
  - busca encontrar todos os objetos que produzem um mesmo tipo de informação;
  - mais fácil de ser compreendida e aplicada comparada a outras técnicas.

# Fundamentação Teórica

## ✓ Exemplo TCA

Tabela

- tomando a tabela como exemplo, tem-se as seguintes definições:

$U = \{1,2,3,4,5,6\}$  é o conjunto universo de objetos (pacientes);  
 $C = \{C,M,T\}$  é o conjunto de atributos de condição;  
 $D = \{G\}$  é o conjunto de atributos de decisão.

# Fundamentação Teórica

## ✓ Exemplo TCA

Tabela

- considerando os atributos {M,T}, têm-se os seguintes conjuntos P-elementares:

$Des_p(\{1\}) = \{(M, Sim), (T, Alta)\};$

$Des_p(\{2,5\}) = \{(M, Não), (T, Alta)\};$

$Des_p(\{3,6\}) = \{(M, Sim), (T, Muito Alta)\};$

$Des_p(\{4\}) = \{(M, Sim), (T, Normal)\}.$

# Fundamentação Teórica

## ✓ Exemplo TCA

[Tabela](#)

- sendo  $\gamma_1$  para SIM e  $\gamma_2$  para NÃO temos:

$$P_{\gamma_1} = \{1\} \cup \{3, 6\} = \{1, 3, 6\};$$

$$P^Y_1 = \{1\} \cup \{2, 5\} \cup \{3, 6\} = \{1, 2, 3, 5, 6\};$$

$$P_{\gamma_2} = \{4\};$$

$$P^Y_2 = \{4\} \cup \{2, 5\} = \{2, 4, 5\};$$

$$\gamma P(Y) = \frac{\text{card}(P_{\gamma_1}) + \text{card}(P_{\gamma_2})}{\text{card}(U)} = \frac{3 + 1}{6} = 0,667$$



# Fundamentação Teórica

## ✓ Exemplo TCA

- calculando todas as aproximações temos:

AtributosP	Qualidade Aproximação	Conj. P-elementares
{C,M,T}	0,667	{1},{2,5},{3},{4},{6}
{M,T}	0,667	{1},{2,5},{3,6},{4}
{C,T}	0,667	{1},{2,5},{3},{4},{6}
{C,M}	0,167	{1,4,6},{2,5},{3}
{T}	0,500	{1,2,5},{3,6},{4}
{M}	0,000	{1,3,4,6},{2,5}
{C}	0,000	{1,4,6},{2,3,5}

## Fundamentação Teórica

### ✓ Exemplo TCA

- as  $\gamma$ -reduções de  $P = \{C, M, T\}$  são  $\{M, T\}$  e  $\{C, T\}$ ;
- o  $\gamma$ -núcleo de  $P$  é:  
 $\text{Core}_\gamma(\{C, M, T\}) = \{C, T\} \cap \{M, T\} = \{T\}$ ;
- $T$  é o atributo mais significativo.

# Desenvolvimento do trabalho

- ✓ Especificação
  - levantamento dos requisitos
  - especificação dos diagramas utilizando UML.

# Desenvolvimento do trabalho

## ✓ Requisitos Funcionais

RF01: O sistema deve permitir ao analista visualizar e selecionar o nome das tabelas do banco de dados que devem ser utilizadas para o processo de análise. UC01

RF02: O sistema deve permitir ao analista informar qual o valor de cálculo dos níveis de conhecimento que serão utilizados na seleção de usuários. UC02

RF03: O sistema deve permitir ao analista selecionar quais os atributos devem ser analisados pela TCA. UC03

RF04: O sistema deve permitir ao analista visualizar e selecionar os valores de domínio correspondentes a cada atributo selecionado. UC04

RF05: O sistema deve permitir ao analista adicionar ou remover os atributos durante ou após a entrada de valores de domínio dos atributos. UC05

RF06: O sistema deve permitir ao analista adicionar ou remover as linhas da tabela durante ou após a entrada dos valores de domínio dos atributos. UC06

RF07: O sistema deve verificar ao iniciar, se todas as tabelas de referência estão informadas corretamente. UC07

# Desenvolvimento do trabalho

## ✓ Requisitos Funcionais

RF08: O sistema deve emitir uma mensagem de erro caso haja algum erro nas tabelas de referência informadas. UC08

RF09: O sistema deve verificar se todos valores de domínios dos atributos e se todos os valores de decisão foram informados. UC09

RF10: O sistema deve emitir uma mensagem de erro caso não sejam informados os valores de domínios dos atributos ou valores de decisão. UC10

RF11: O sistema deve emitir uma mensagem de erro sempre que houver algum erro na comunicação com o banco de dados. UC11

RF12: O sistema deve calcular os valores dos níveis de conhecimento dos usuários baseado nos atributos selecionados pelo analista. UC12

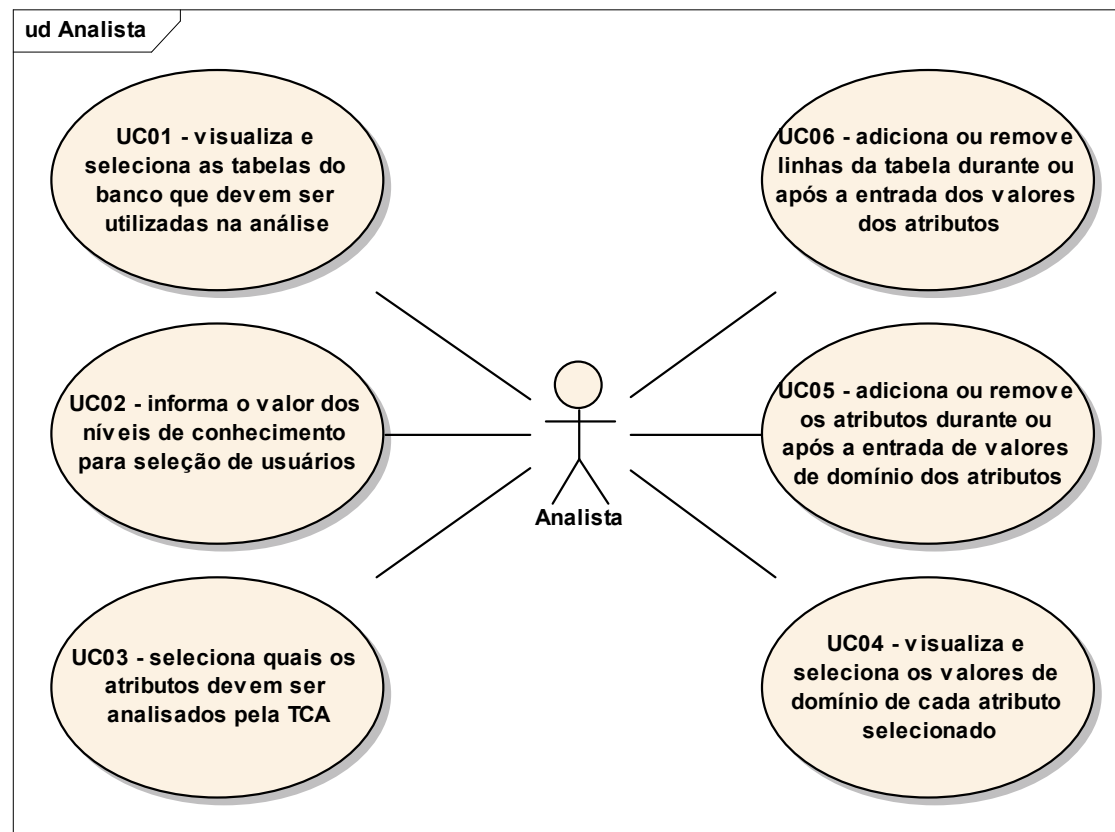
RF13: O sistema deve calcular e informar ao analista qual o atributo mais significativo obtido através da TCA. UC13

RF14: O sistema deve emitir ao analista o resultado da análise da TCA após salvar os valores informados na tabela. UC14

RF15: O sistema deve gerar um arquivo de log com os resultados obtidos no processo da análise da TCA. UC15

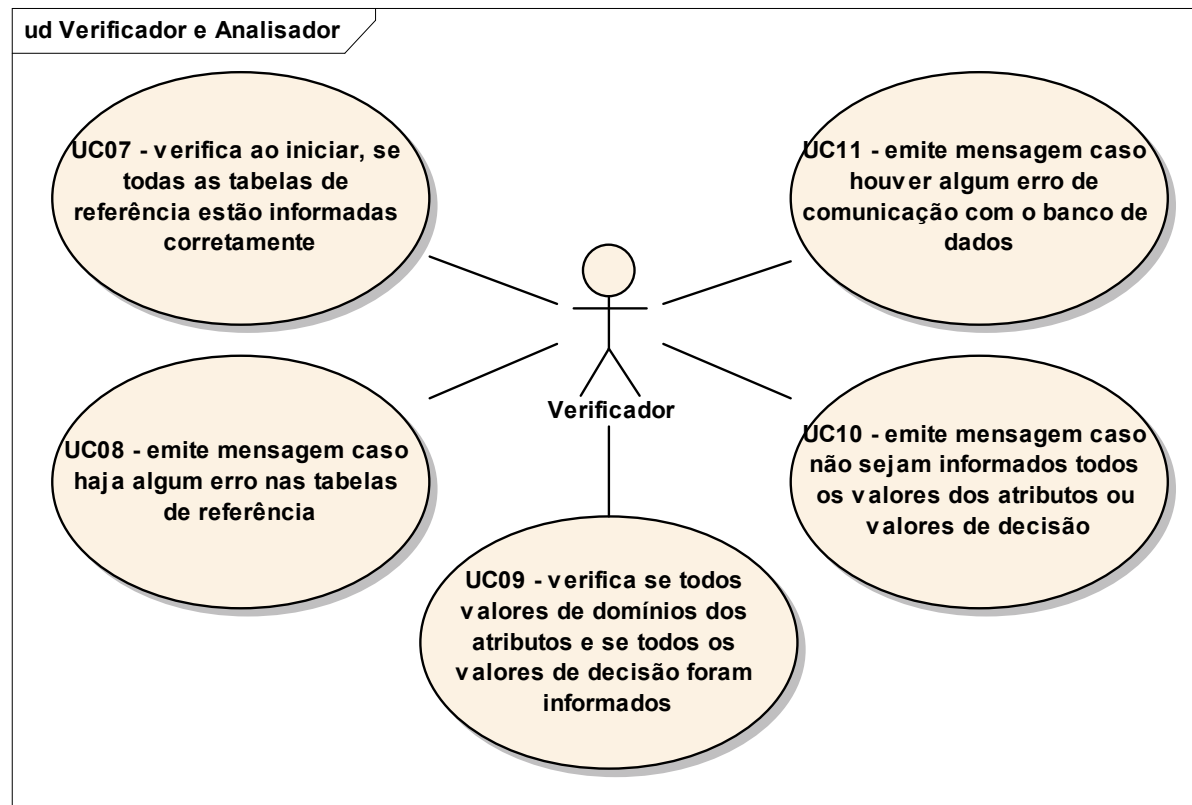
# Desenvolvimento do trabalho

## ✓ Diagrama de casos de uso do analista



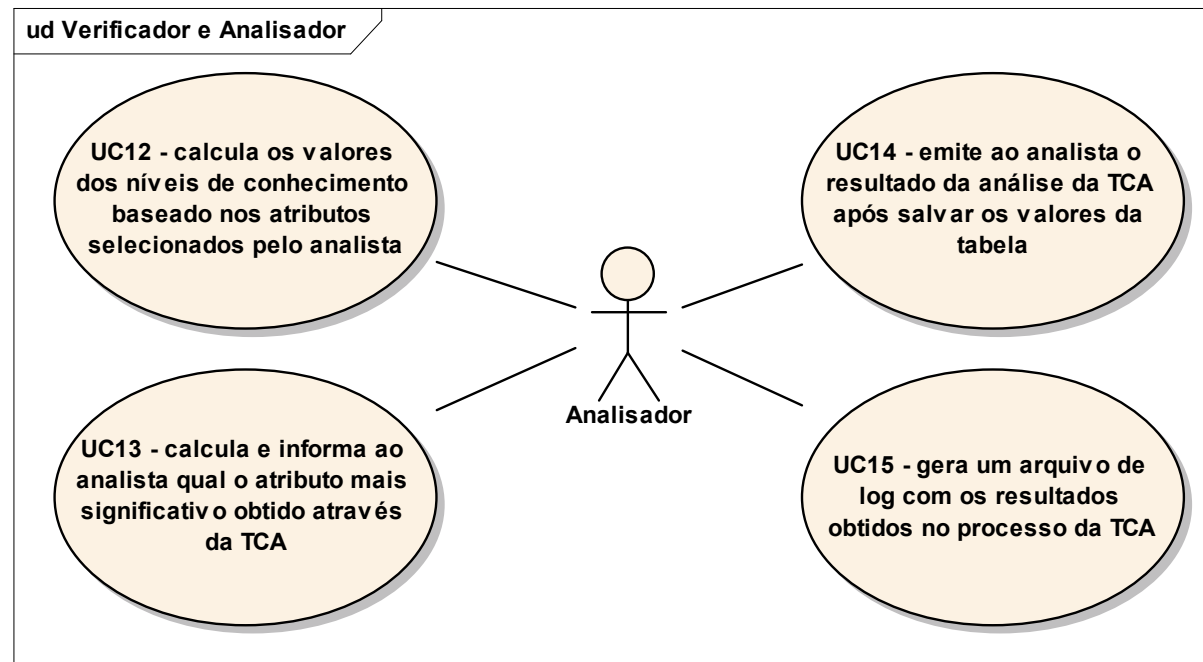
# Desenvolvimento do trabalho

## ✓ Diagrama de casos de uso do verificador



# Desenvolvimento do trabalho

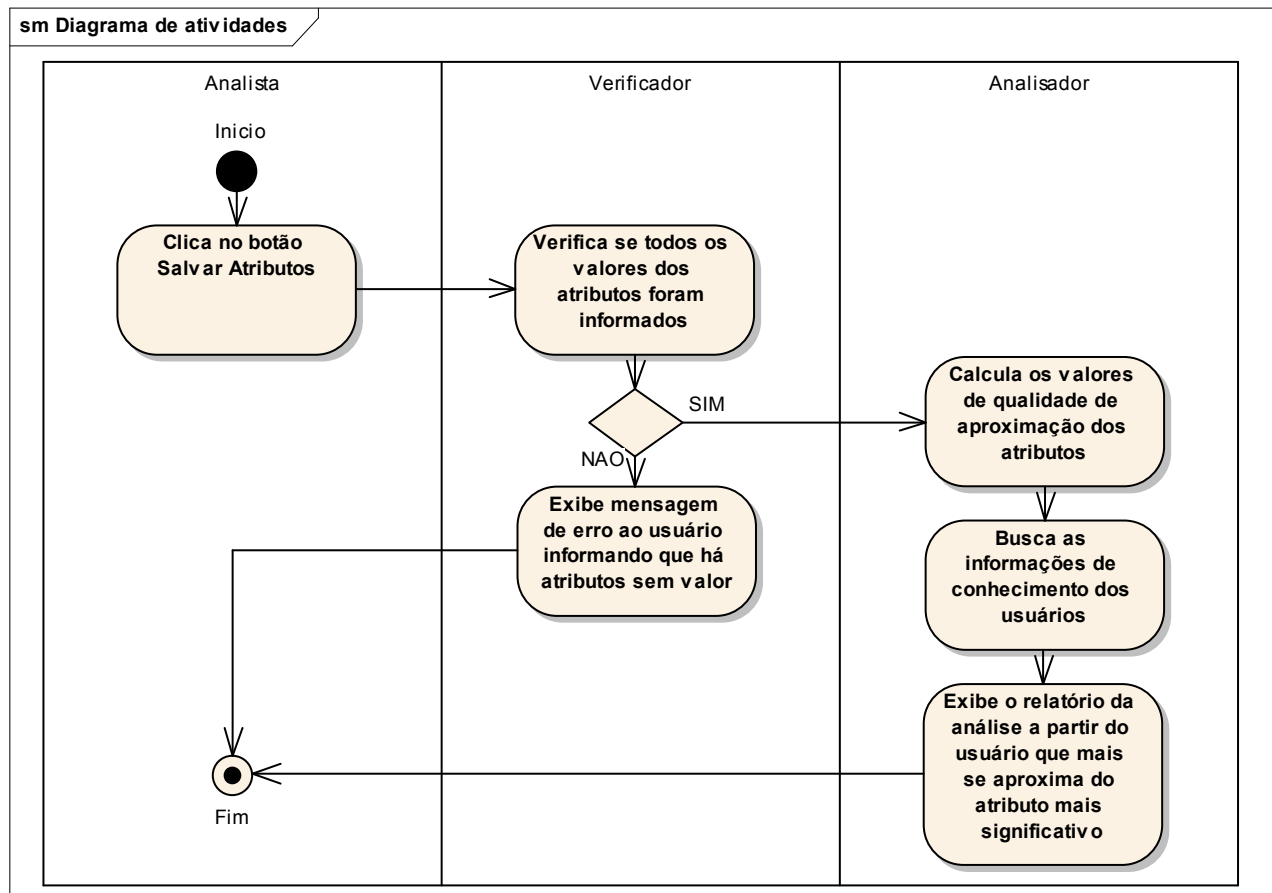
## ✓ Diagrama de casos de uso do analisador





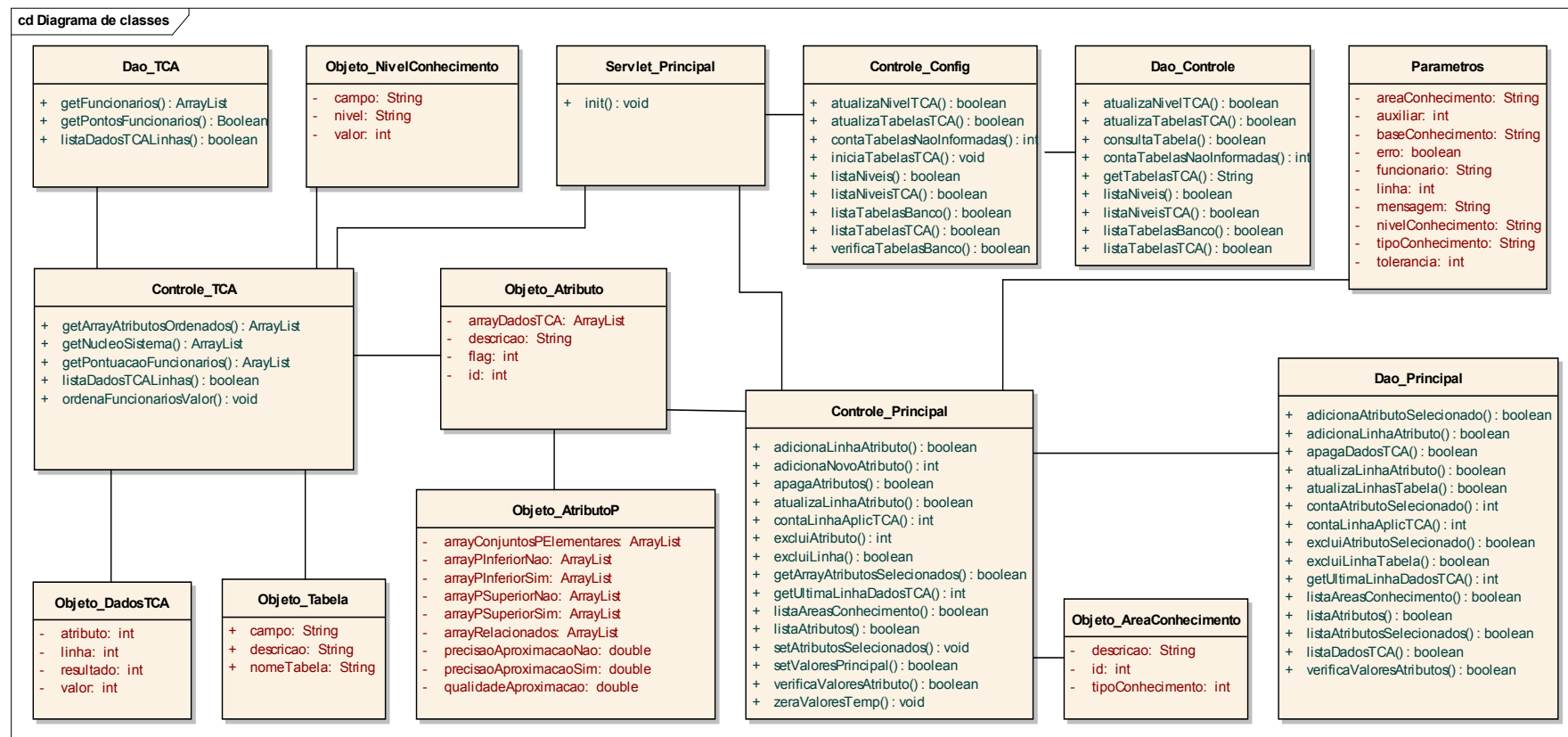
# Desenvolvimento do trabalho

## ✓ Diagrama de atividades do fluxo da análise



# Desenvolvimento do trabalho

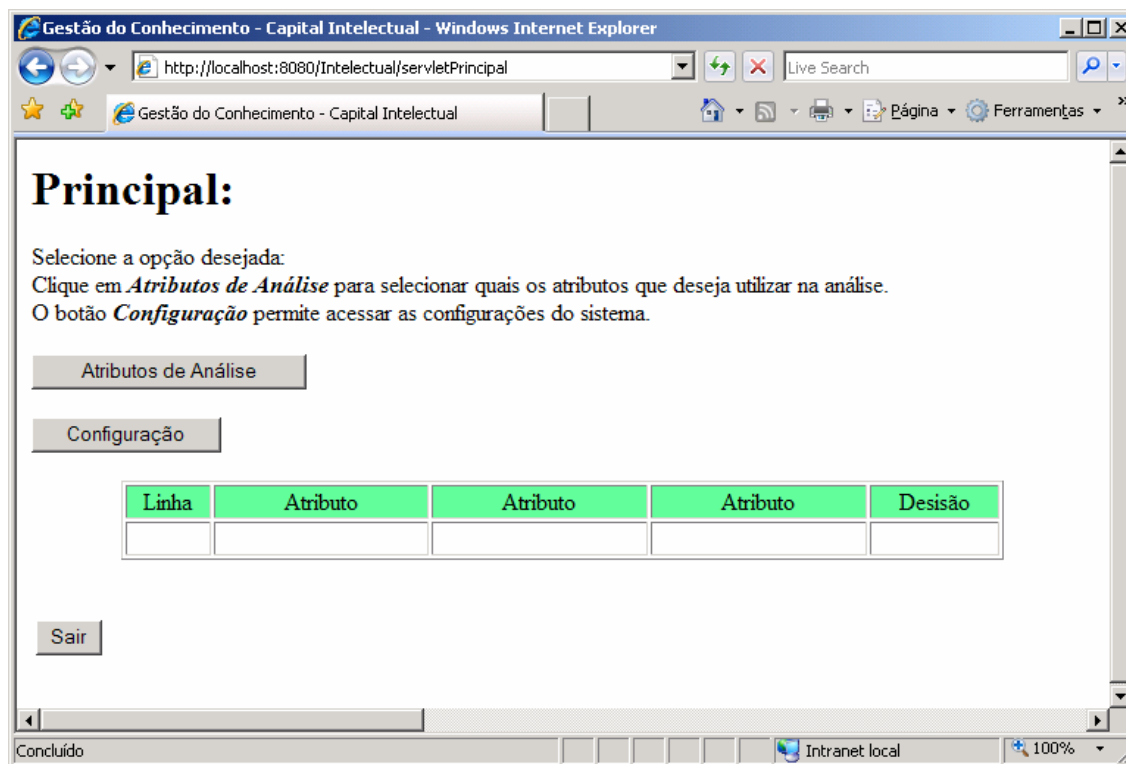
## ✓ Diagrama de classes



# Desenvolvimento do trabalho

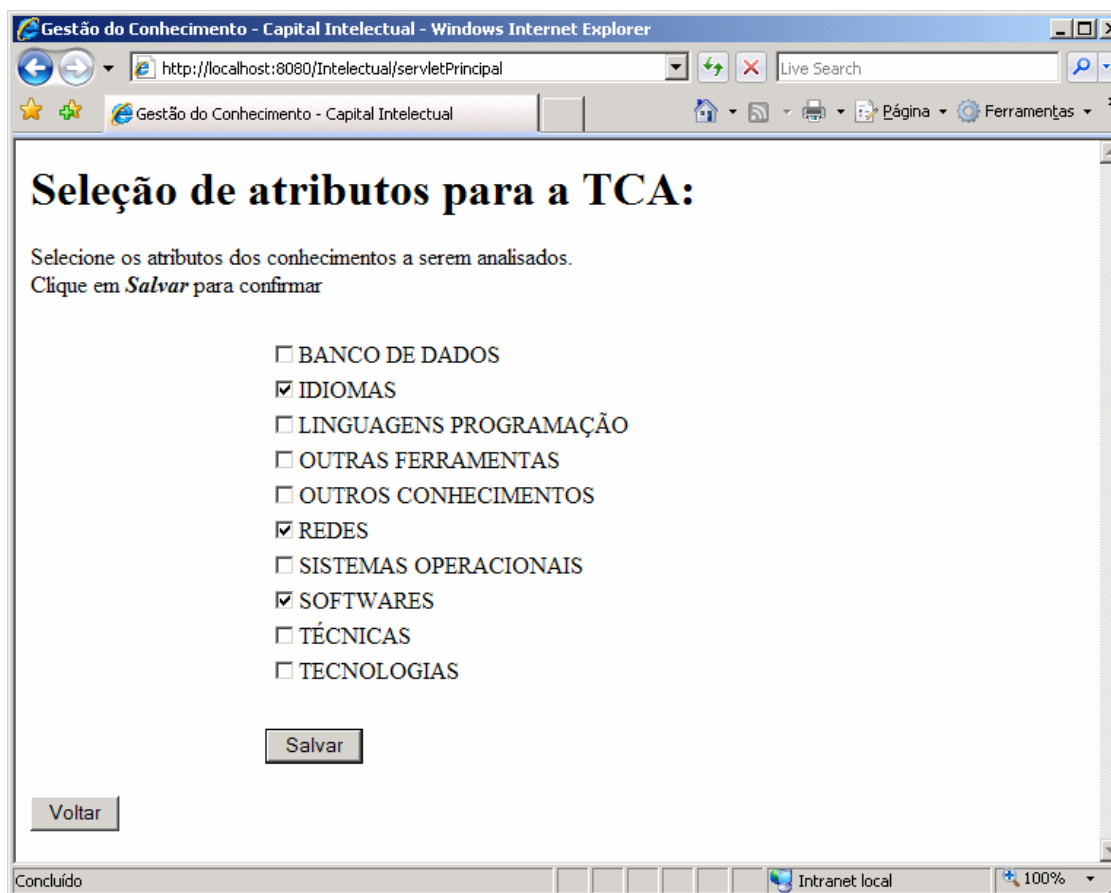
- ✓ Técnicas e ferramentas utilizadas
  - Enterprise Architect - UML:  
criação dos diagramas e casos de uso;
  - JSP e JSTL:  
desenvolvimento das classes e páginas html;
  - MySQL:  
base de dados de informações do sistema;
  - Tomcat:  
servidor de aplicações;
  - Eclipse:  
ambiente de desenvolvimento para implementação do trabalho.

# Operacionalidade



Tela principal do sistema

# Operacionalidade

A screenshot of a web browser window titled "Gestão do Conhecimento - Capital Intelectual - Windows Internet Explorer". The address bar shows "http://localhost:8080/Intelectual/servletPrincipal". The page content includes the heading "Seleção de atributos para a TCA:" followed by instructions: "Selecione os atributos dos conhecimentos a serem analisados. Clique em *Salvar* para confirmar". Below this is a list of attributes with checkboxes: BANCO DE DADOS, IDIOMAS (checked), LINGUAGENS PROGRAMAÇÃO, OUTRAS FERRAMENTAS, OUTROS CONHECIMENTOS, REDES (checked), SISTEMAS OPERACIONAIS, SOFTWARES (checked), TÉCNICAS, and TECNOLOGIAS. At the bottom of the list are two buttons: "Salvar" and "Voltar". The browser's status bar at the bottom shows "Concluído", "Intranet local", and "100%".

**Tela de seleção de atributos para TCA**

# Operacionalidade



**Principal:**

Selecione a opção desejada:  
Clique em *Atributos de Análise* para selecionar quais os atributos que deseja utilizar na análise.  
O botão *Configuração* permite acessar as configurações do sistema.

**3 registro(s) atualizado(s).**

Atributos de Análise

Configuração

Linha	IDIOMAS	REDES	SOFTWARES	Desisão	Excluir
1	Selecione.....	Selecione.....	Selecione.....	Selecione	-><-

Adicionar Linha

Salvar Atributos

Sair

Concluído Intranet local 100%

Tela principal com atributos selecionados

# Operacionalidade

Gestão do Conhecimento - Capital Intelectual - Windows Internet Explorer

http://localhost:8080/Intelectual/servletPrincipal

Principal:

Selecione a opção desejada:  
 Clique em *Atributos de Análise* para selecionar quais os atributos que deseja utilizar na análise.  
 O botão *Configuração* permite acessar as configurações do sistema.

Atributos de Análise

Configuração

Linha	IDIOMAS	REDES	SOFTWARES	Desisão	Excluir
1	Alemão	Linux-Unix	Gráfico-Internet	Sim	-><
2	Inglês	Microsoft-Novel	Gráfico-Internet	Sim	-><
3	Inglês	Linux-Unix	Suíte Office	Sim	-><
4	Alemão	Linux-Unix	Desenvolvimento	Não	-><
5	Inglês	Microsoft-Novel	Gráfico-Internet	Não	-><
6	Alemão	Linux-Unix	Suíte Office	Sim	-><

Adicionar Linha

Salvar Atributos

Sair

Concluído

Intranet local 100%

Tela principal com valores e atributos informados

# Operacionalidade

Gestão do Conhecimento - Capital Intelectual - Windows Internet Explorer

http://localhost:8080/Intelectual/servletPrincipal

Gestão do Conhecimento - Capital Intelectual

Atributos de Análise

Configuração

Linha	IDIOMAS	REDES	SOFTWARES	Decisão	Excluir
1	Inglês	Linux-Unix	Gráfico-Internet	Sim	-><-
2	Alemão	Microsoft-Novel	Gráfico-Internet	Sim	-><-
3	Alemão	Linux-Unix	Suíte Office	Sim	-><-
4	Inglês	Linux-Unix	Desenvolvimento	Não	-><-
5	Alemão	Microsoft-Novel	Gráfico-Internet	Não	-><-
6	Inglês	Linux-Unix	Suíte Office	Sim	-><-

Adicionar Linha

Salvar Atributos

Atributo(s) mais Significativo(s): **SOFTWARES** Índice 50%

COLABORADOR	SOFTWARES	IDIOMAS	REDES
Claudia Geigher	180	96	64
Romulo de Oliveira	168	96	24
Silvana Tomio	168	88	60
Osvaldo Fernandes	160	96	24
Sabrina Ribeiro	156	108	24
Veruska Reis dos Santos	154	72	56

Tela principal com apresentação dos valores



# Obtenção de dados

Qualificação	Proficiência
• Adobe Acrobat	Avançado
• Microsoft Excel	Avançado
• Microsoft Office	Avançado
• Microsoft PowerPoint	Avançado
• Microsoft Word	Avançado
• Microsoft Front Page	Intermediário
• Adobe Photoshop	Intermediário
• Microsoft Access	Intermediário
• Microsoft Visual Studio	Intermediário



Conhecimento - Capital Intelectual - Windows Internet Explorer

http://localhost:8080/Intelectual/ser...

Gestão do Conhecimento - Capita...

valores (pesos) dos *Níveis de Conhecimento* a serem analisados na mineração

Níveis:	Valor:
Nenhum	0
Muito Fraco	0
Fraco / Iniciante	1
Pouco	0
Razoável	0
Satisfatório / Intermediário	2
Bom	0
Muito Bom	0
Excelente / Avançado	4
Certificado	6

Salvar

Concluído

Intranet local 100%



# Desenvolvimento do trabalho

## ✓ Trabalhos correlatos

- gerar um modelo de classificação de dados utilizando entropia e árvores de decisão (Geandro Compolt, 1999);
- aplicar árvores de decisão sobre uma base fornecida pela Central dos Alunos da FURB (Biancca Nardelli, 2000);
- sistema WEB utilizando Data Mining para descobrir informações implícitas, dos perfis dos usuários da Biblioteca Central da FURB (Alberto P. de Jesus, 2004).

# Considerações Finais

## ✓ Conclusões

- o conhecimento é um diferencial competitivo muito importante para as organizações;
- os processos e ferramentas de *Data Mining* são indispensáveis na gestão do conhecimento;
- o objetivo principal de demonstrar a TCA em *Data Mining* para geração do capital intelectual foi alcançado.

## Considerações Finais

### ✓ Extensões

- rotina de análise dos resultados que auxilie o analista a melhorar a busca de dados;
- adicionar outras técnicas de análise para comparação de resultados.

# Obrigado!

**“Quanto maiores somos em humildade,  
tanto mais perto estamos da grandeza”**

**Rabindranath Tagore**

Escritor Indiano (1861 – 1941)

