UNIVERSIDADE DO EXTREMO SUL CATARINENSE – UNESC

CURSO DE CIÊNCIA DA COMPUTAÇÃO

PATRICIA JOSÉ PORFIRIO

APLICAÇÃO DE ONTOLOGIAS E DATA MINING PARA DESCOBERTA DE CONHECIMENTO

CRICIÚMA, NOVEMBRO DE 2008

PATRICIA JOSÉ PORFIRIO

APLICAÇÃO DE ONTOLOGIAS E DATA MINING PARA DESCOBERTA DE CONHECIMENTO

Trabalho de Conclusão do Curso apresentado para a obtenção do Grau de Bacharel em Ciência da Computação da Universidade do Extremo Sul Catarinense.

Orientadora: Profa. MSc. Merisandra Côrtes de Mattos
Co-orientador: Prof. Esp. Miguel Moretti

CRICIÚMA, NOVEMBRO DE 2008

PATRICIA JOSE PORFIRIO

Aplicação de Ontologias e *Data Mining* para Descoberta de Conhecimento

Submetido ao corpo docente do Curso de Ciência da Computação da Universidade do Extremo Sul Catarinense como um dos requisitos para obtenção do grau de Bacharel em Ciência da Computação.

Profa. MSc. Ana Claudia Garcia Barbosa
Coordenadora do Curso de Ciência da Computação

Banca Examinadora:

Profa. MSc. Merisandra Côrtes de Mattos (UNESC)
Orientador

Prof. Esp. Miguel Moretti (UNESC)
Co-Orientador

Prof. MSc. Eduardo Menna da Silva (UNESC)

Profa. MSc. Priscyla Waleska Simões (UNESC)

Aos meus pais, José e Helena que sempre me apoiaram e ao meu filho Gabriel Vitor.

AGRADECIMENTOS

Agradeço a Deus que me deu forças para continuar seguindo em busca dos meus sonhos.

Agradeço também:

A minha família que é tudo pra mim me proporcionando, amor, carinho e compreensão nas horas em que mais foi preciso.

A meu filho Gabriel Vitor, que me proporciona momentos inesquecíveis com seu sorriso demonstrando todo seu amor.

A minha orientadora e professora Merisandra, que esteve sempre pronta a me ajudar.

Enfim, a todos os meus amigos de graduação e professores que contribuíram para a conclusão desta pesquisa.

RESUMO

O processo de descoberta do conhecimento em bases de dados, em especial o *data mining,* se faz necessário para identificar as relações existentes entre as informações armazenadas nas organizações. Anterior a etapa de *data mining* nestas bases de dados, deve-se realizar o pré-processamento a fim de organizar, selecionar e preparar as informações para a execução do *data mining*. Apesar disso, pode-se ter conjuntos de dados inconsistentes. De mesma forma, na fase de pós-processamento os conhecimentos podem estar redundantes e não compreensíveis. Mediante isso, pode-se utilizar as ontologias como uma alternativa para padronização das informações. Na realização deste estudo utilizou-a uma base de dados referente ao diagnóstico de doença coronariana, e desenvolveu-se ontologias associadas ao processo de *data mining* em dois momentos distintos. Primeiramente, estruturaram-se os conhecimentos pelas ontologias anteriormente a etapa da *data mining.* No segundo momento realizou-se o *data mining* na base de dados e desenvolveu-se a ontologia dos conhecimentos descobertos. Com isso, objetivou-se a análise dos benefícios obtidos pelo desenvolvimento de ontologia antes e após a etapa de *data mining*. Resultou na confirmação dos conhecimentos na área, demonstrou que a aplicação da ontologia no pré-processamento do processo de DCBD permitiu uma melhor preparação dos dados, além de resultar em uma forma melhor de interpretação do conhecimento na fase de pós-processamento.

Palavras-chave: *Data Mining*, Classificação, Algoritmo ID3, Ontologias, Metodologia Methontology, Doença Coronariana.

ABSTRACT

The process of knowledge discovery in databases, *data mining* in particular, it is necessary to identify the relationship between the information stored within organizations. Prior to the stage of *data mining* in these databases, you must complete the pre-processing in order to organize, select and prepare information for the implementation of *data mining*. Nevertheless, it has been inconsistent data sets. From Likewise, in the post-processing knowledge may be redundant and not understandable. Through this, you can use the ontologies as an alternative to standardization of information. In conducting this study used it to a database concerning the diagnosis of coronary disease, and it was developed ontologies associated with the process of *data mining* in two different times. First, it is structured by the knowledge ontologies earlier stage of the data mining. In the second phase took place in the data mining database and developed to ontology of knowledge discovered. With that aimed to the analysis of the benefits gained through the development of ontology before and after the stage of *data mining*. Resulted in the confirmation of knowledge in the area, showed that the application of ontology in the pre-processing of the DCBD allow better preparation of the data, and result in a better way of interpreting the knowledge in the post-processing.

Keywords: *Data Mining*, Classification, ID3 algorithm, Ontologies, Methodology Methontology, Coronary Heart Disease.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE SIGLAS

CART	C*lassification and Regression Trees*
DCBD	Descoberta de Conhecimento em Bases de Dados
DCC	Diagnóstico de Doença Coronariana
DST	Doenças Sexualmente Transmissíveis
IC	Inteligência Computacional
ID3	*Iterative Dichotomiser3*
KDD	*Knowledge Discovery in Databases*
KIF	Knowledge Interchange Format
KLS	Knowledge Systems Laboratory
OIL	*Ontology Inference Layer*
OKBC	Open Knowledge Based Connectivity
OWL	Ontology Web Language
PROLOG	*Programing Logic*
RDF	*Resource Description Framework*
TOVE	Toronto Virtual Enterprise
UNESC	Universidade do Extremo Sul Catarinense
XML	*EXtensible Markup Language*

SUMÁRIO

1 INTRODUÇÃO

A Descoberta de Conhecimento em Bases de Dados (DCBD) se faz necessária para a identificação das relações e padrões ainda não conhecidos, bem como para confirmar os já existentes. Este processo é composto por várias etapas dividindo-se em pré-processamento, *data mining* e pós-processamento. A etapa de *data mining* é considerada uma das principais, pois é efetivamente responsável pela busca do conhecimento (GOLDSCHMIDT; PASSOS, 2005).

O pré-processamento nas bases de dados deve ser realizado anteriormente a etapa de *data mining*, a fim de organizar, selecionar e preparar as informações. Apesar disso, pode-se ter conjuntos de dados inconsistentes, como por exemplo, atributos com diferentes denominações que possuem o mesmo significado. Isto pode comprometer a aplicação do *data mining*. Assim, estas bases de conhecimentos podem ser representadas por meio de ontologias como uma alternativa para padronização das informações (CÍSARO, NIGRO, XODO, 2008, tradução nossa).

O termo ontologia teve origem nos pensamentos filosóficos de Aristóteles e significava o estudo do ser, referenciava um único princípio aplicando categorias para apenas um domínio (MAEDCHE, 2002, tradução nossa). As ontologias se constituem em uma recente aplicação na área de Inteligência Artificial que utiliza a representação e estruturação de dados, bem como a organização de conhecimentos descobertos (CASTOLDI, 2003).

Dessa forma, o desenvolvimento de ontologias previamente ao *data mining* pode possibilitar uma descoberta de conhecimento mais significativa, já que os dados serão estruturados anteriormente na fase de pré-processamento (CÍSARO, NIGRO, XODO, 2008, tradução nossa).

Do mesmo modo, na etapa de pós-processamento do DCBD, muitas vezes tem-se redundância de conhecimentos, ou eles não estão organizados e estruturados de uma forma que facilite a sua análise e entendimento. Assim, pode-se gerar ontologias posteriormente a etapa de *data mining* para que se estruture e organize os conhecimentos descobertos. De acordo com Câmara et al (2001) as ontologias permitem a identificação de classes específicas de objetos e relacionamentos que existam em um domínio de conhecimento.

Mediante isso, esta pesquisa constituiu na análise do benefício para a descoberta de conhecimento proporcionado pela construção das ontologias previamente e após a etapa de *data mining*. A fim de realizar esta análise foi:

a) gerada a ontologia para estruturação e padronização de conhecimento, referente ao diagnóstico de doenças coronarianas, anteriormente a aplicação de *data mining*, por meio da tarefa de classificação e do algoritmo ID3[1];
b) aplicada a etapa de *data mining* na base de dados e posteriormente construiu-se uma outra ontologia para organização do conhecimento descoberto.

1.1 OBJETIVO GERAL

Analisar a aplicação das ontologias para a estruturação do conhecimento antes e após a etapa de *data mining*.

[1] Algoritmo utilizado na tarefa de classificação para indução de árvores de decisão, trabalha apenas com dados nominais o que o diferencia de outros métodos (KANTARDZIC, 2003, tradução nossa).

1.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos desta pesquisa são:

a) compreender o processo de descoberta de conhecimento e *data mining*;
b) entender a representação e estruturação de conhecimento por meio de ontologias;
c) utilizar a tarefa de classificação pelo algoritmo ID3 na etapa de *data mining*;
d) gerar uma ontologia de aplicação para a base de dados do diagnóstico de doença coronariana;
e) realizar a etapa de *data mining* nos dados organizados pelas ontologias;
f) aplicar a etapa de *data mining* e representar por meio de ontologias o conhecimento descoberto;
g) empregar a Shell Orion Data Mining Engine;
h) utilizar a ferramenta Protégé para a construção de ontologias.

1.3 JUSTIFICATIVA

O aumento na capacidade de armazenamento das informações nos dispositivos tem estimulado a necessidade de transformar grandes quantidades de dados em informações realmente úteis. Tendo-se o *data mining* como uma ferramenta essencial para auxiliar na descoberta de conhecimentos significativos de maneira eficiente.

De acordo com Nigro, Císaro e Xodo (2008, tradução nossa) atualmente um dos problemas mais importantes e desafiadores do *data mining* é a representação da

descoberta do conhecimento. Este processo de busca da informação precisa ter uma fase de seleção dos dados e técnicas necessárias para uma diminuição do número de atributos redundantes, tornando os conhecimentos compreensíveis em todo o processo.

Neste contexto as ontologias estão sendo utilizadas, automatizando o processo de descoberta de conhecimento (NIGRO; CÍSARO; XODO, 2008, tradução nossa). A combinação de *data mining* com ontologias é ainda recente e pouco explorada, porém tem sido empregada como base para a representação do conhecimento adquirido (SILVA, 2004).

A criação de uma ontologia sobre o conhecimento de um domínio, no caso da base de dados referente ao diagnóstico de doença coronariana, conforme Noy e Mcguinness (2001, tradução nossa), contribui para organizar e formalizar conceitos, relações e características comuns do domínio considerado.

Além disso, de acordo com Menzies (1999), a utilização de ontologias proporciona vantagens como: acesso mútuo dos componentes as informações e funcionalidades; pesquisa e navegação; utilização de meta-conhecimento[2] para auxiliar na navegação e ampliar consultas, evitando a reconstrução de componentes já existentes.

A construção da ontologia no pré-processamento e pós-processamento do processo de DCBD para a descoberta de conhecimento foi realizada utilizando-se a ferramenta Protégé, pois é disponibilizada gratuitamente, possuindo uma interface simples e de fácil utilização. Além disso, suporta a metodologia Methontology[3] que foi escolhida para o desenvolvimento da ontologia.

---

[2] Consiste na reutilização de um conhecimento para uma nova definição do mesmo domínio (NIGRO; CÍSARO; XODO, 2008, tradução nossa).

[3] Metodologia utilizada que possibilita por meio de suas fases a construção da ontologia de forma detalhada (LINHALIS, 2007).

A etapa de *data mining* na base de dados foi realizada empregando a Shell Orion Data Mining Engine, pois esta se encontra em desenvolvimento por acadêmicos e pesquisadores do Grupo de Pesquisa em Inteligência Computacional Aplicada da UNESC, sendo disponibilizada gratuitamente.

A tarefa de *data mining* utilizada neste processo foi a classificação por ser entendida como uma das mais importantes e responsável pela busca de uma função que permite associar cada registro corretamente (GOLDSCHMIDT; PASSOS, 2005).

Neste estudo optou-se pelo uso do método de árvores de decisão, pois quando utilizado em conjunto com a tecnologia de indução de regras apresenta os resultados em um formato com priorização. Dessa forma, a regra mais importante consiste no primeiro nó da árvore e as demais são mostradas em nós seguintes, constituindo-se assim em uma das suas principais vantagens, além de serem de fácil entendimento para a maioria das pessoas (SERRA, 2002).

Dentre os algoritmos de árvores de decisão para a tarefa de classificação, escolheu-se para utilização nesta pesquisa o algoritmo ID3 que, conforme Luger (2004) é interessante pela representação que produz do conhecimento aprendido, capacidade de gerenciamento da complexidade e potencial para trabalhar com dados ruidosos.

1.4 ESTRUTURA DO TRABALHO

Este trabalho possui seis capítulos, sendo o primeiro composto pela introdução, objetivos e justificativas.

A Descoberta de Conhecimento em Base de Dados, é comentada no Capítulo 2, apresentando-se definições, contextualização de suas etapas com ênfase ao *data mining*, tarefas e métodos.

No Capítulo 3 é abordado o conceito de representação do conhecimento que é fundamental para o entendimento do restante da pesquisa, com ênfase as ontologias, apresentando-se tipos, componentes, metodologias e ferramenta para a sua construção.

A utilização de ontologias e *data mining* é abordada no Capítulo 4, sendo de suma importância nesta pesquisa. Alguns exemplos da aplicação de ontologias e *data mining* em conjunto são apresentados no Capítulo 5.

No Capítulo 6 o trabalho desenvolvido é apresentado, sendo descrito todo o processo de desenvolvimento e os resultados obtidos. Por fim, tem-se a conclusão e as sugestões de trabalhos futuros.

2 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

Devido à evolução tecnológica o volume de dados armazenados nas organizações vem crescendo consideravelmente aumentando assim a necessidade de ferramentas capazes de extrair conhecimentos, tendências e padrões úteis que facilitem a tomada de decisão (HAN; KAMBER, 2001, tradução nossa). Mediante isso, muitas pesquisas sobre o processo de descoberta de conhecimento estão sendo desenvolvidas para analisar, compreender, solucionar e desenvolver estratégias sobre as informações armazenadas.

A Descoberta de Conhecimento em Bases de Dados (DCBD) ou *Knowledge Discovery in Databases* (KDD) originou-se no campo da Inteligência Computacional (IC) com o objetivo de explorar grandes bases de dados (HAND; MANNILA; SMYTH, 2001, tradução nossa).

DCBD é o processo que analisa e encontra em grandes volumes de dados conhecimentos novos, compreensíveis pelo ser humano, padrões desconhecidos, tendências que sejam úteis no processo de decisão (REZENDE, 2005).

Utiliza-se de técnicas oriundas de várias áreas, dentre elas: estatística[4], inteligência computacional[5], reconhecimentos de padrões[6], aprendizado de máquina[7] e banco de dados[8], sendo, portanto, multidisciplinar (GOLDSCHMIDT; PASSOS, 2005).

---

[4] Ciência que estuda os números é um conjunto de técnicas e métodos que têm por objetivo obter, organizar e analisar dados estatísticos (SERRA, 2002).

[5] Conjunto de técnicas que visa o desenvolvimento de sistemas inteligentes no intuito de imitar o comportamento humano de aprendizado, percepção e raciocínio. (CAVERSAN; ANDRADE, 2005).

[6] Identifica os objetos por meio de suas características atribuindo para eles padrões, número de categorias ou classes, tem por objetivo desenvolver, dispositivos que possam ser implementados por computador com maior precisão (PAL; MITRA, 2004, tradução nossa).

[7] Tem o objetivo de fazer com que os programas aprendam com dados que são estudados, tornando esses programas aptos a tomarem decisões diferentes de acordo com o aprendizado dos dados estudados (SERRA, 2002).

[8] Local de armazenamento de conjuntos de dados dispostos em uma estrutura que permite a manipulação dos mesmos, agrupa registros e pode ser acessado por meio de sistemas gerenciadores de dados (HAN; KAMBER, 2001, tradução nossa).

DCBD é constituído por etapas iterativas e interativas, ou seja, nas descobertas dos conhecimentos adquiridos, ou falta de informações relevantes o usuário poderá decidir pela retomada dos processos ou uma nova seleção de dados (FAYYAD et al, 1996, tradução nossa).

Conforme Goldschmidt e Passos (2005) três etapas operacionais básicas constituem o processo de DCBD: pré-processamento, *data mining* e pós-processsamento. Estas etapas são dividas em sub processos (Figura 1):

a) **pré-processamento**: compreende todo o processo de preparação dos dados para a etapa de *data mining* envolvendo as funções de seleção, limpeza, codificação e enriquecimento de dados;

b) ***data mining***: é a principal etapa do processo de DCBD, onde são definidos os métodos e tarefas que serão utilizados para a busca efetiva dos conhecimentos. Esse processo é essencial e a sua execução envolve a aplicação de algoritmos sobre a base de dados;

c) **pós-processamento**: compreende a fase de análise, visualização e interpretação dos conhecimentos gerados, onde os especialistas em DCBD e no domínio de aplicação avaliam e definem novas alternativas para a investigação dos dados.

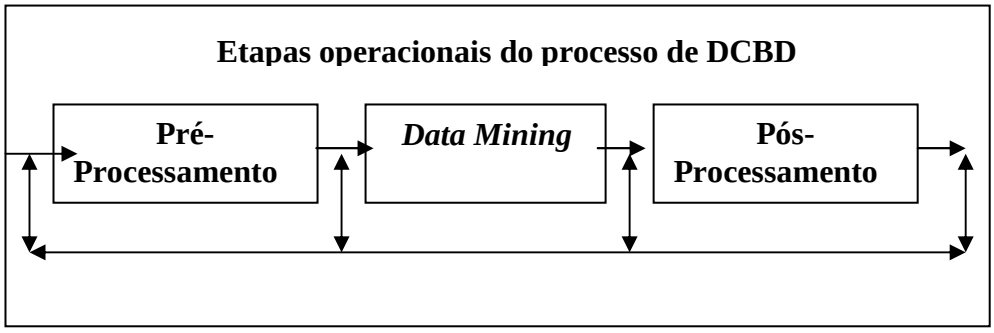


Figura 1. Etapas do Processo de DCBD
Fonte: Adaptado. GOLDSCHMIDT, R.; PASSOS, E. (2005).

Assim, considerando-se os avanços tecnológicos no que se refere ao armazenamento e processamento dos dados, bem como o papel fundamental desempenhado pela informação nas organizações, o processo de DCBD em especial a etapa de *data mining*, constituíram-se em uma das técnicas mais exploradas e utilizadas atualmente (GOLDSCHMIT; PASSOS, 2005).

2.1 DATA MINING

Data mining é um processo de descoberta de padrões significativos em grandes bases de dados por meio de métodos e tarefas. Descobrir relações entre os dados e identificar conhecimentos úteis que possam ajudar na previsão de tendências são seus principais objetivos (WITTEN; FRANK, 2005, tradução nossa).

Conforme Pal e Mitra (2004, tradução nossa) *data mining* consiste na aplicação de algoritmos para a análise de dados de forma a produzir padrões e construir vários modelos de dados.

O processo de *data mining* basicamente consiste em predição e descrição. Predição envolve descobrir valores desconhecidos. Descrição, contudo visa encontrar padrões que descrevem os dados a serem interpretados por especialistas (KANTARDZIC, 2003, tradução nossa). Assim, conforme os objetivos que se tem com a descoberta de conhecimento aplicam-se algumas tarefas de *data mining*, tendo-se as principais citadas pela literatura: associação, clusterização, estimativa, previsão de séries temporais e classificação.

A tarefa de associação consiste na descoberta de relacionamentos significativos entre os dados armazenados e identifica conjuntos de itens que ocorrem simultaneamente em uma base de dados (SERRA, 2002). A associação é muito utilizada

em marketing no processo de descoberta de regras associativas entre os produtos vendidos, para pesquisar e determinar a disposição dos produtos nas prateleiras das lojas, em catálogos, no comércio em geral, de modo que os itens descobertos que possuem relação, estejam sempre próximos (DIAS, M. 2001).

Tarefa de Clusterização também chamada de agrupamento, é utilizada para segmentar as informações contidas na base de dados, agrupando em um mesmo *cluster* os atributos com valores similares, de forma que um conjunto de dados tenha propriedades que os diferencie de outros *clusters* (KAMBER; HAN, 2001, tradução nossa). O resultado da clusterização é uma descrição generalizada de cada conjunto, sendo importante para uma análise das características dos dados na base de conhecimento (KANTARDZIC, 2003, tradução nossa). A clusterização pode ser utilizada, por exemplo, para agrupar clientes por região do país, por comportamento de compra similar, entre outros (SILVA, 2001).

A estimativa é uma tarefa conhecida também por regressão, compreende a busca por funções que atribuam um valor para algum item de dado ainda desconhecido a partir de um atributo real. Alguns exemplos de sua aplicação: estimativa da probabilidade de um paciente sobreviver considerando-se um conjunto de exames médicos; a demanda de consumo de um produto novo (FAYYAD et al, 1996, tradução nossa).

Tarefa de Previsão de séries temporais consiste na descrição de um subconjunto de fenômenos ocorridos e na detecção de padrões em itens de dados ordenados no tempo. Os estudos de séries temporais têm sido utilizados em problemas reais, auxiliando na tomada de decisões. O objetivo maior do uso da análise de séries temporais é a geração de modelos voltados à previsão de valores futuros. Pode-se citar como exemplo: o consumo de energia elétrica mensal de uma casa registrado durante

um ano; as vendas mensais de um produto, entre outros (GOLDSCHMIT; PASSOS 2005).

A tarefa de classificação de acordo com Larose (2005, tradução nossa) consiste em utilizar um grupo de dados pré-definidos desenvolvendo um modelo composto por padrões que distinguem as classes e são aplicados para novas classificações. Esta tarefa como foi utilizada no desenvolvimento da presente pesquisa é abordada com ênfase a seguir.

2.1.1 A Tarefa de Classificação

A Classificação consiste na análise de dados para a definição de padrões (modelos) que diferenciem as classes. Esta tarefa pode ser entendida como a procura por uma função, que permita associar corretamente cada registro X_i de uma base de dados a um único rótulo categórico Y_j, denominado classe (GODSHIMIT; PASSOS, 2005).

Na tarefa de classificação o objetivo é construir um modelo conciso, contudo para isso utiliza-se de uma variável alvo que possa agrupar os dados em hierarquias, tendo o valor correspondente a uma classe. Essa variável é dada para atribuir uma categoria a um item de dado (LAROSE, 2005, tradução nossa).

A variável alvo uma vez identificada pode ser aplicada a novos registros de forma a prever o grupo em que estes se enquadrem, tendo-se a finalidade de usar o modelo resultante para atribuir valores ao item de dado com classe desconhecida (HAN; KAMBER, 2001, tradução nossa).

Esta tarefa pode ser utilizada, por exemplo, para classificar o atributo renda (variável alvo) em elevada, média e baixa. Neste caso, cada registro que contém a variável alvo é analisado bem como as variáveis de entrada. Supondo que se queria

classificar a pessoa não somente pela renda, mas também pela idade, primeiramente serão examinados os registros com a variável idade e posteriormente busca-se a renda definida como variável alvo (SERRA, 2002).

A tarefa de classificação é aplicada por meio de alguns métodos que são compostos por algoritmos. Assim, de acordo com as propriedades, características dos dados e o resultado a ser atingido define-se o método mais adequado. Alguns métodos que podem ser utilizados na tarefa de classificação:

a) **redes neurais artificiais**: são modelos abstratos dos processos do cérebro humano. Uma rede neural é uma estrutura que consiste em um número de nodos conectados (chamados neurônios) com as ligações direcionais. Redes Neurais têm a capacidade de aprendizado baseada na habilidade das mesmas modificarem seus parâmetros de acordo com as iterações realizadas (LAROSE, 2005, tradução nossa);

b) **algoritmos genéticos**: baseados no desenvolvimento natural das espécies são métodos de busca e otimização do processo de evolução, freqüentemente são aplicados para a resolução de problemas difíceis. (KANTARDZIC, 2003, tradução nossa). As gerações são criadas aplicando operadores genéticos tais como: seleção[9], cruzamento[10] e mutação[11], até que todas as regras dentro de uma geração satisfaçam as possibilidades de desenvolvimento (HAN; KAMBER, 2001, tradução nossa);

[9] Operação que determina quais os cromossomos vam formar a próxima geração (KANTARDZIC, 2003, tradução nossa).

[10] Faz a recombinação dos cromossomos enquanto cria duas novas gerações trocando os subseqüentes da direita e esquerda entre os cromossomos escolhidos na seleção (LAROSE, 2005, tradução nossa).

[11] Altera os valores dos genes sorteados aleatoriamente por meio de probabilidades, os cromossomos da nova geração podem ter os genes alterados (LAROSE, 2005, tradução nossa).

c) **raciocínio baseado em casos**: método que utiliza experiências e soluções passadas para resolver um novo problema. Os resultados exatos são dados pela distância dos vizinhos (DIAS, M. 2001). As funções usadas para encontrar os vizinhos mais próximos e combinar os valores para fazer uma previsão, são os principais elementos do método, uma vez encontradas tendem a ficarem estáveis, mesmo que sejam inseridos novos exemplos para outras categorias nos dados já conhecidos (FERNANDES, 2003);

d) árvores de decisão: método que consiste na construção de uma árvore na forma *top-down*[12], classificando os atributos com maior ganho de informação para o atributo de saída selecionado (KANTARDZIC, 2003, tradução nossa). Este método foi empregado no desenvolvimento desta pesquisa, sendo, portanto apresentado a seguir.

2.1.1.1 Árvores de Decisão

As árvores de decisão são meios de representar os resultados obtidos pela etapa de *data mining*, onde os nós demonstram uma decisão sobre um atributo em particular (WITTEN; FRANK, 2005, tradução nossa).

A árvore é gerada de forma recursiva de acordo com os valores dos atributos, os nós são testados sucessivamente, e quando uma folha é alcançada o atributo é classificado de acordo com a classe atribuída à folha. Se o atributo testado em um nó for nominal[13], a quantidade de filhos é geralmente o número de valores possíveis

[12] Consiste em construir árvores de decisão na forma de cima para baixo, pois a classificação parte da raiz da árvore (LUGER, 2004).

[13] Nomes dados a atributos para a sua identificação (GOLDSCHMIDT; PASSOS, 2005)

do atributo. Neste caso onde existe um valor possível para cada um, o mesmo atributo não será mais reexaminado na árvore (LAROSE, 2005, tradução nossa).

Um exemplo de árvore de decisão para a classificação de dois atributos de entrada X e Y é dado na Figura 2. Onde todas as amostras com valores X > 1 e Y = B pertencem à classe 2, enquanto as amostras com valores X < 1 pertencem à classe 1, para qualquer que seja o valor da característica Y (KANTARDZIC, 2003, tradução nossa).

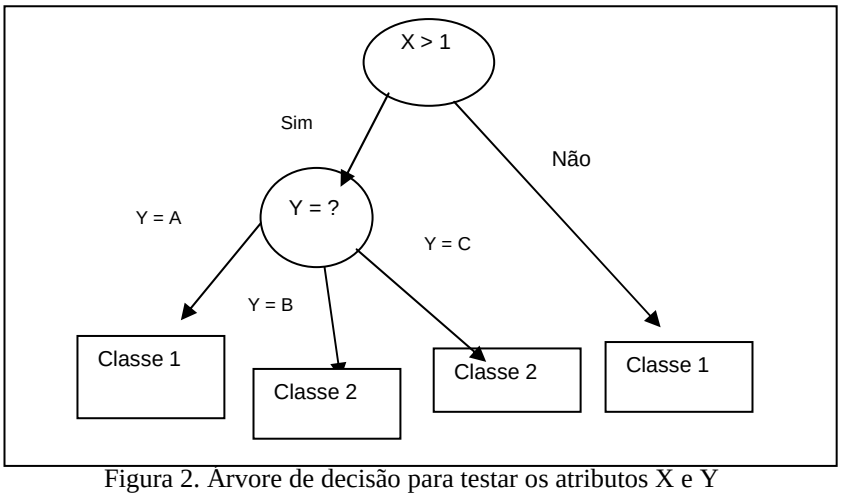


Figura 2. Árvore de decisão para testar os atributos X e Y
Fonte: KANTARDZIC, M. (2003, tradução nossa)

Goldschmidt e Passos (2005) afirmam que uma árvore de decisão é um modelo de conhecimento onde cada nó representa uma decisão acerca de um atributo que determina como os dados estão particionados pelos nós filhos. A raiz da árvore, inicialmente, contém toda a base de dados, tendo-se exemplos de diferentes classes misturados. Mediante a escolha de um ponto de separação[14] divide-se a base de dados em dois ou mais conjuntos associados a um nó filho.

[14] Condição que melhor separa as classes e envolve um dos atributos do problema (GOLDSCHIMIT; PASSOS, 2005).

Dentre os algoritmos utilizados para a indução de árvores de decisão tem-se o ID3 como o primeiro desenvolvido, e sucessivamente C4.5[15] e CART[16] (GOLDSCHIMIT; PASSOS, 2005).

O algoritmo ID3 foi utilizado no desenvolvimento desta pesquisa, pois de acordo com Luger (2004), é o mais utilizado para a indução de árvores de decisão em grandes bases de dados por permitir trabalhar com dados ruidosos.

2.1.1.2 Algoritmo ID3

O Algoritmo *Iterative Dichotomiser 3* (ID3) foi desenvolvido em 1983 por Ross Quinlan na Universidade de Sydney na Austrália (SERRA, 2002).

A forma como o ID3 constrói árvores de decisão é de cima para baixo (*top-down*), pois faz a classificação dos dados de uma forma recursiva, escolhendo um atributo para um nó, partindo da raiz, selecionando uma propriedade para testar os nós descendentes até que certos critérios de parada sejam atingidos ou que todos os atributos sejam usados (LUGER, 2004).

O ID3 opera dados nominais, o que o diferencia dos outros métodos. Esse algoritmo utiliza a lógica e matemática para processar, organizar e simplificar um grande conjunto de dados (KANTARDZIC, 2003, tradução nossa).

A seguir tem-se a forma com que o algoritmo constrói a árvore de decisão sobre um conjunto de dados (KANTARDZIC, 2003, tradução nossa):

a) seleciona um atributo para ser raiz da árvore e cria tantas partições quantos valores tiver esse atributo;

b) utiliza a árvore gerada para classificar o conjunto de treinamento;

---

[15] Algoritmo que trabalha tanto com atributos nominais e numéricos (SERRA, 2002).
[16] Algoritmo que processa apenas valores numéricos (LAROSE, 2005, tradução nossa).

c) se todos os exemplos em uma folha tiverem o mesmo valor para o objeto de saída, retorna ao nó folha este valor;

d) caso contrário, cria um nó com um atributo que ainda não foi utilizado em seus nós ancestrais, e gera todas as partições possíveis para ele, a seguir retorne ao segundo passo.

O algoritmo ID3 faz a classificação do conjunto de amostras de treinamento escolhendo um atributo que possua o maior ganho de informação, aquele que contribua com certa quantidade de informação que melhor o classifique o conjunto de treinamento (LUGER, 2004).

O ganho de informação é um cálculo estatístico utilizado para a escolha do melhor atributo de todo o conjunto que está sendo classificado. Entretanto, para ser compreendido tem-se que conhecer o conceito e entropia. A entropia é a medida que faz o cálculo da homogeneidade do conjunto, sendo aplicada para realizar a avaliação da aleatoriedade da variável a prever na classe (KANTARDZIK, 2003, tradução nossa).

O cálculo do ganho de informação é obtido por (GOLDSCHMIDT; PASSOS, 2005):

a) considerando a partição da base associada ao nó análise:

$$\inf(S) = -\sum_{j=1}^{k} \frac{freq(C_{j,}S)}{|S|} x \log_2 \left(\frac{freq(C_{j,}S)}{|S|} \right) bits$$

Onde:

- S representa a partição de dados;

- *Fre*(C_j , S) representa o número de vezes em que a classe C_j ocorre em S;

- | S | número de casos do conjunto S;

- *k* indica o número de classes distintas.

b) ganho de informação de cada atributo considerando a partição da base associada ao nó análise:

$$\inf_x(T) = \sum_{i=1}^{n} \frac{|T_i|}{|T|} x \inf(T_i)$$

Onde:

- T representa a número de ocorrências na partição em análise;

- T_i representa a quantidade de ocorrências de uma classe contida no T.

c) seleção do atributo com maior ganho de:

$$gain(X) = \inf(T) - \inf_x(T)$$

Um exemplo de árvore de decisão gerada pelo ID3 é mostrado na Figura 3, onde se realizou a classificação do conjunto de dados história de crédito de aplicações de financiamento dado na Tabela 1.

Tabela 1. Árvore de decisão para a avaliação do risco de crédito

Risco	**História de Crédito**	**Dívida**	**Garantia**	**Renda**
ato	ruim	alta	nenhuma	$0 a $15k
ato	desconhecida	alta	nenhuma	$15 a $35k
moderado	desconhecida	baixa	nenhuma	$15 a $35k
alto	desconhecida	baixa	nenhuma	$0 a $15k
baixo	desconhecida	baixa	nenhuma	acima de $35k
baixo	desconhecida	baixa	adequada	acima de $35k
alto	ruim	baixa	nenhuma	$0 a $15k
moderado	ruim	baixa	adequada	acima de $35k
baixo	boa	baixa	nenhuma	acima de $35k
baixo	boa	alta	adequada	acima de $35k
alto	boa	alta	nenhuma	$0 a $15k
moderado	boa	alta	nenhuma	$15 a $35k
baixo	boa	alta	nenhuma	acima de $35k
alto	ruim	alta	nenhuma	$15 a $35k

Fonte: LUGER,G. (2004)

Observa-se na árvore gerada os atributos que continham o maior ganho de informação na classificação da avaliação de risco de crédito.

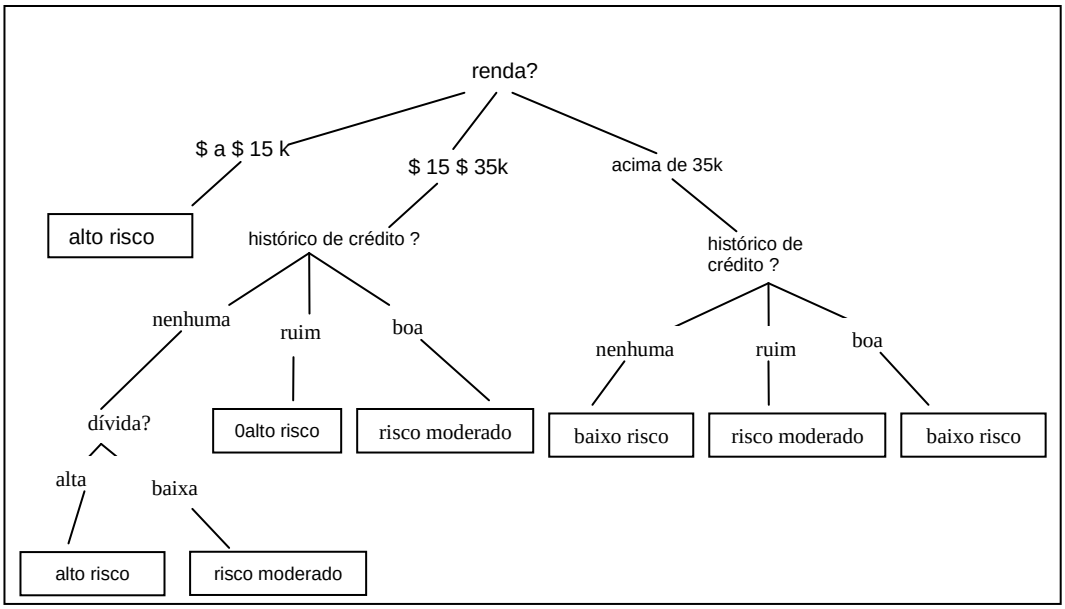


Figura 3. Árvore de decisão para a avaliação de risco de crédito.
Fonte: LUGER,G. (2004)

Neste capítulo foram apresentados os conceitos de *data mining*, classificação, método de árvores de decisão e do algoritmo ID3, fundamentais para entendimento da pesquisa realizada. No entanto, para aplicar a técnica de *data mining* pode-se utilizar ferramentas computacionais que apresentam esses conceitos implementados.

Atualmente, têm-se disponíveis vários softwares para este fim, porém em sua maioria são comerciais. Mediante isso, tem-se em ambientes acadêmicos algumas iniciativas de implementação de ferramentas gratuitas, dentre as quais a *Shell Orion Data Mining Engine*, desenvolvida pelo Grupo de Pesquisa em Inteligência Computacional Aplicada da Universidade do Extremo Sul Catarinense e utilizada neste trabalho.

2.2 SHELL ORION DATA MINING ENGINE

A *Shell* Orion *Data Mining Engine* tem sido desenvolvida com o intuito de disponibilizar diferentes tarefas, métodos e algoritmos. Até o momento encontram-se implementados os seguintes módulos: associação pelo algoritmo Apriori; classificação pelo método de indução de árvores de decisão (ID3 e CART por meio do critério de Gini); clusterização pelos métodos de redes neurais (algoritmo de Kohonen), particionamento (K-means) e lógica *fuzzy* (Gustafson-Kessel).

Esta *shell* busca oferecer uma interface de fácil utilização, tendo-se na Figura 4 a sua interface principal.

A *Shell* Orion está sendo implementada no ambiente de desenvolvimento NetBeans 5.5 que utilizada arquitetura Java, por ser gratuita e multiplataforma disponível para *download* no *site* www.sun.org.

Figura 4. Interface Principal da Shell Orion Data Mining Engine
Fonte: BORTOLOTO, L. (2007)

Esta ferramenta permite conexão com diversos sistemas gerenciadores de banco de dados, sendo necessário apenas o cadastro (Figura 5) do *driver* de conexão (CASSETARI JÚNIOR, 2008).

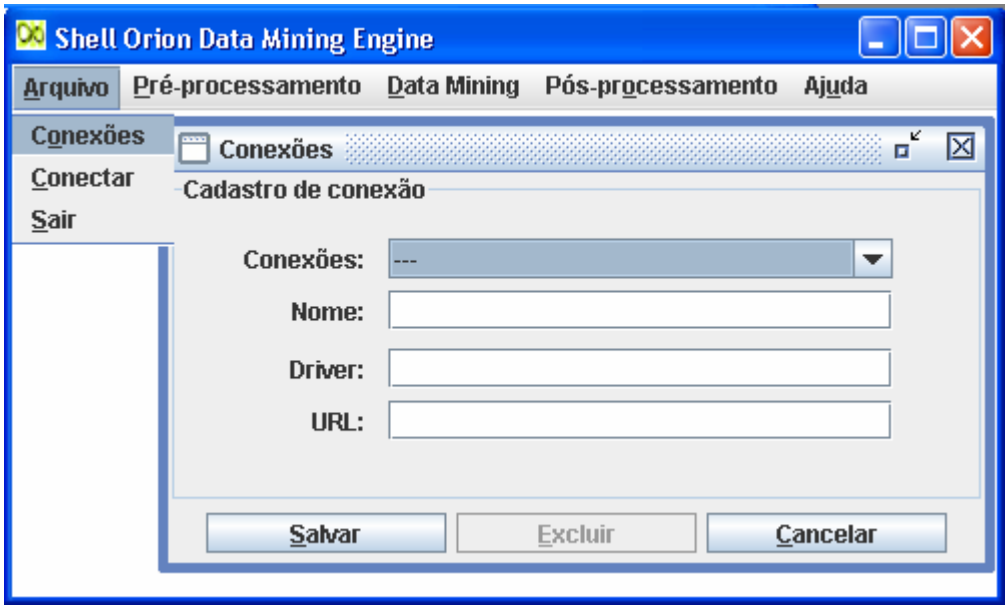


Figura 5. Tela de cadastro de conexão Shell Orion Data Mining Engine
Fonte: BORTOLOTTO, L. (2007).

Após seu cadastro, o usuário pode se conectar com o sistema gerenciador de banco de dados por meio do menu *Arquivo/Conectar* informando apenas o nome da base, seu usuário e senha. Concluída a conexão o próximo passo é a escolha da tabela a qual será realizada as etapas de *data mining* nos módulos disponíveis.

O módulo de classificação da *Shell* Orion tem implementado o método de indução por árvores de decisão pelo algoritmo ID3, que faz a classificação dos dados a partir do atributo de saída selecionado e o nível de profundidade da árvore informada pelo usuário gerando assim as regras de classificação (PELEGRIN, 2005).

Na Figura 6 tem-se um exemplo deste método, onde o objetivo de saída escolhido foi sexo e o nível de profundidade da árvore dois, ou seja, o número de atributos que vai ser testado de acordo com seu ganho de informação. Neste caso

diagnóstico e faixa etária foram os que obtiveram maior relevância na determinação do sexo.

Além de gerar as regras este módulo disponibiliza também a visualização da árvore de classificação construída (Figura 7).

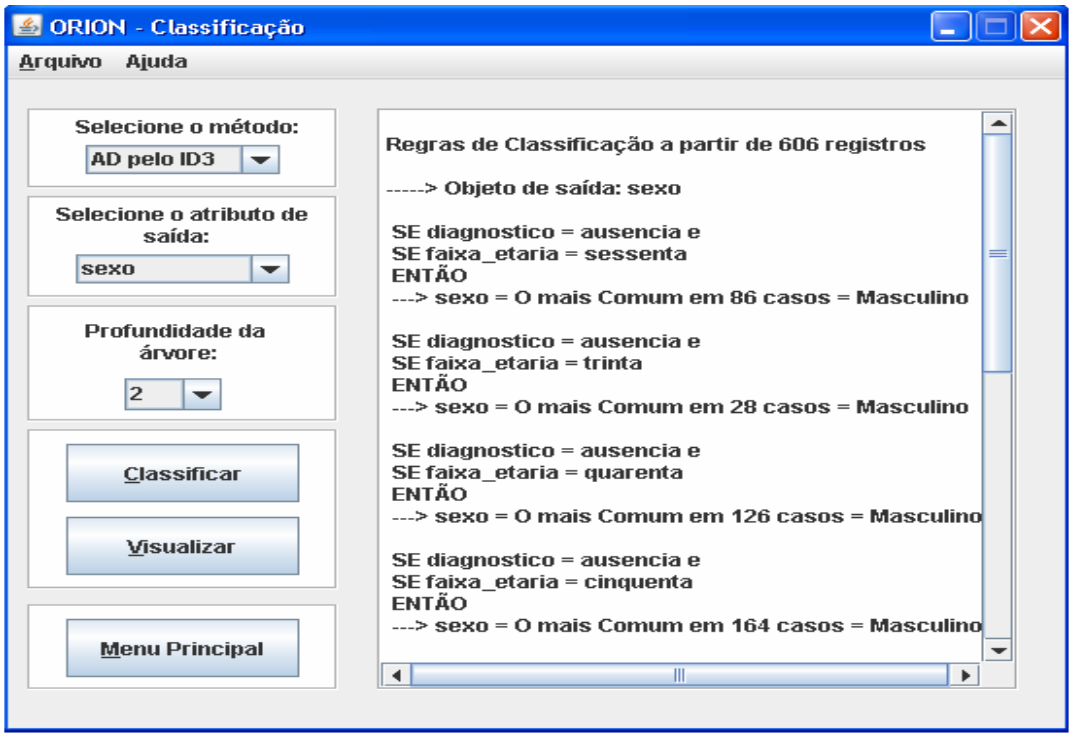


Figura 6. Classificação pelo algoritmo ID3
Fonte: PELEGRIN, D. (2005).

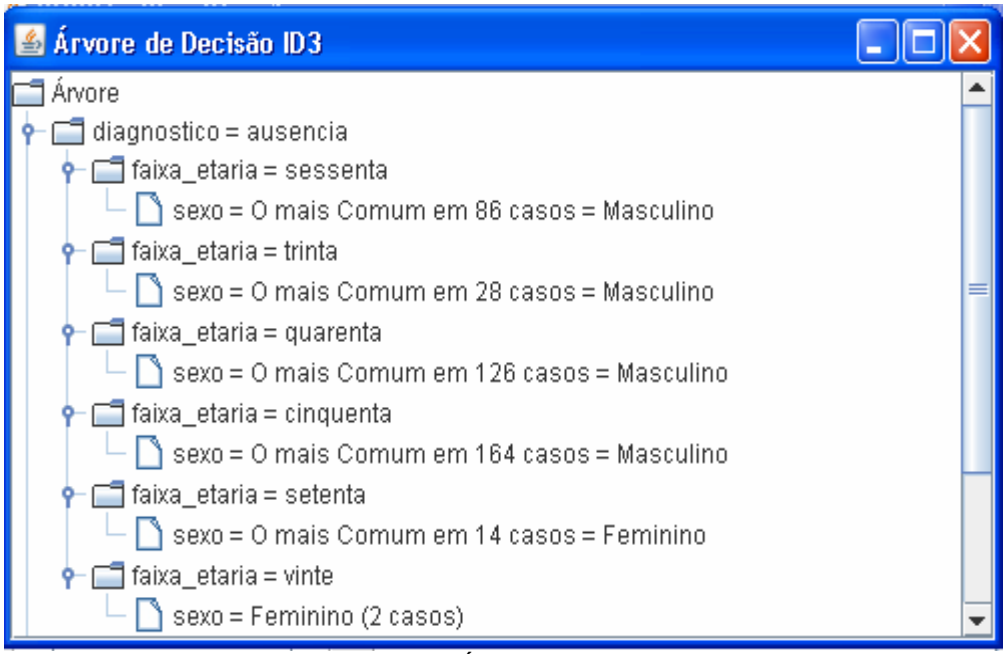


Figura 7. Árvore de decisão
Fonte: PELEGRIN, D. (2005).

Ainda no módulo de classificação tem-se desenvolvido o algoritmo CART pelo critério de Gini implementado (Figura 8) (RAIMUNDO, 2007). Este método faz a divisão das informações em uma árvore de dados de forma recursiva de acordo com critérios pré-estabelecidos (SERRA, 2002).

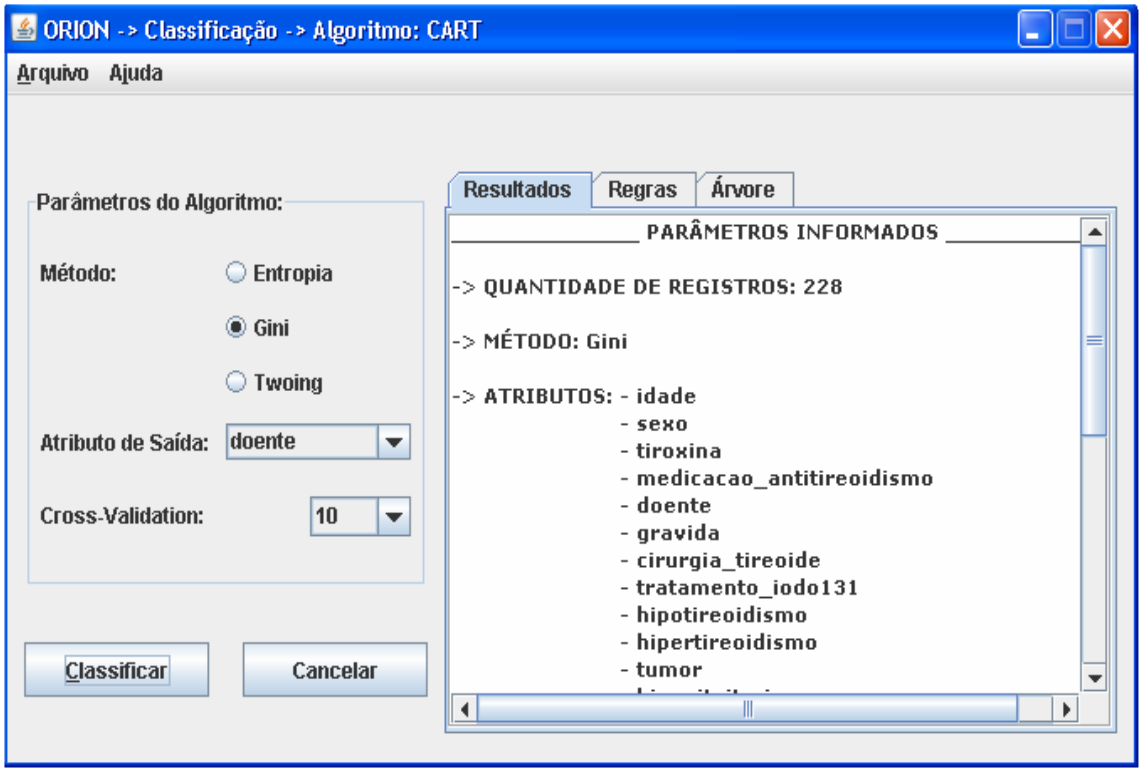


Figura 8. Classificação pelo algoritmo CART (pelo critério de Gini)
Fonte: RAIMUNDO, L. (2007)

O módulo de associação da Shell (Figura 9) tem implementado o algoritmo Apriori (CASAGRANDE, 2005).

De acordo com Hand, Mannila e Smyth (2001, tradução nossa) o algoritmo Apriori tem a finalidade de agrupar os elementos mais freqüentes em uma base de dados de forma recursiva, gerando as regras de associação.

Na *Shell Orion Data Mining Engine* tem-se que definir os parâmetros de suporte e confiança, que se fazem necessários para a confiabilidade das regras associativas geradas.

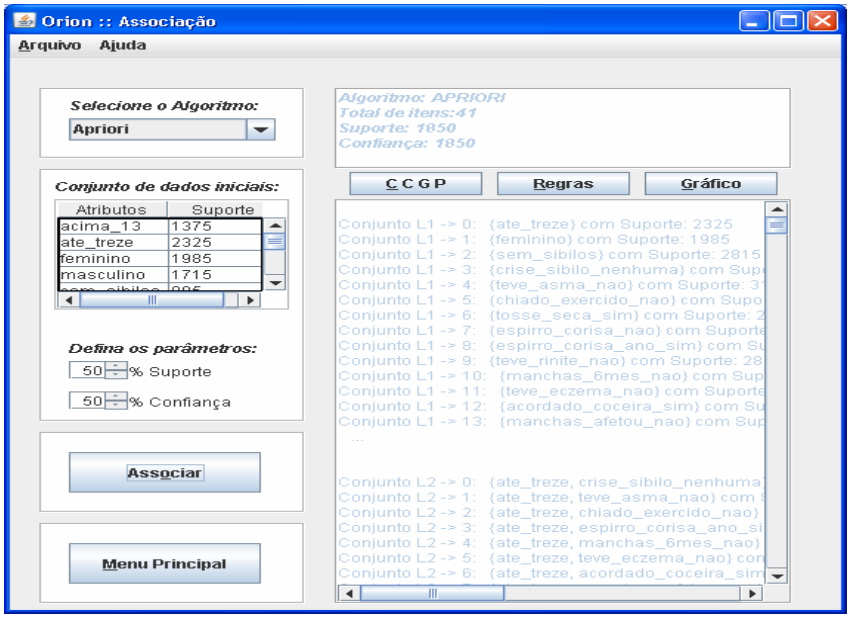


Figura 9. Módulo de associação pelo algoritmo Apriori
Fonte: CASAGRANDE, D. (2005)

No módulo de clusterização (Figura 10) tem-se implementado o algoritmo de Kohonen (BORTOLOTTO, 2007).

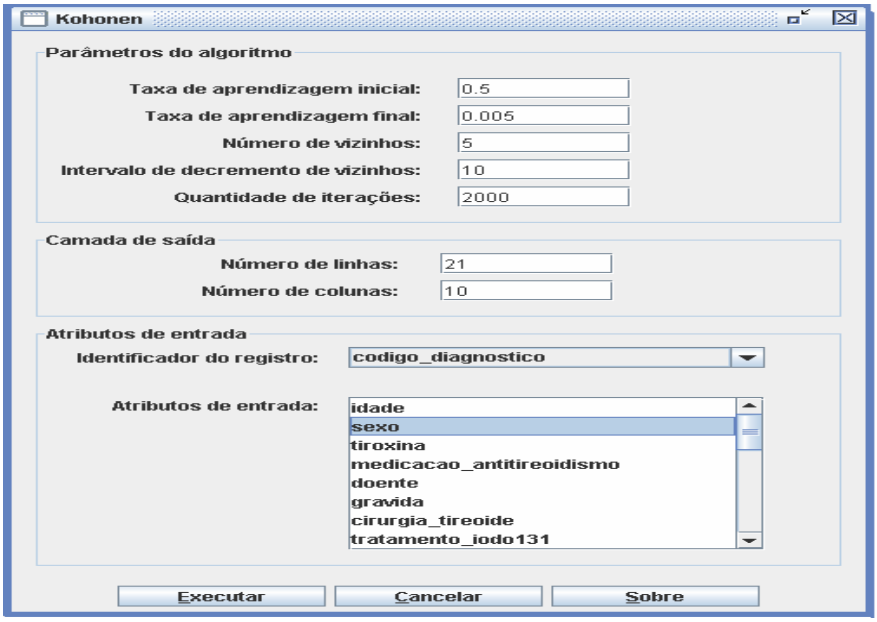


Figura 10. Módulo de clusterização pelo algoritmo de Kohonen
Fonte: BORTOLOTTO, L. (2007)

De acordo com Goldschmidt e Passos (2005) este algoritmo é pertencente à classe das redes neurais, tendo-se os conjuntos de informações como padrões de entradas divididos em grupos similares conforme o aprendizado da rede.

Na *Shell* os resultados obtidos pelo algoritmo de Kohonen podem ser visualizados por meio de gráficos, pelos grupos gerados, arquivos em *sql* e disponibiliza também um relatório (Figura 11).

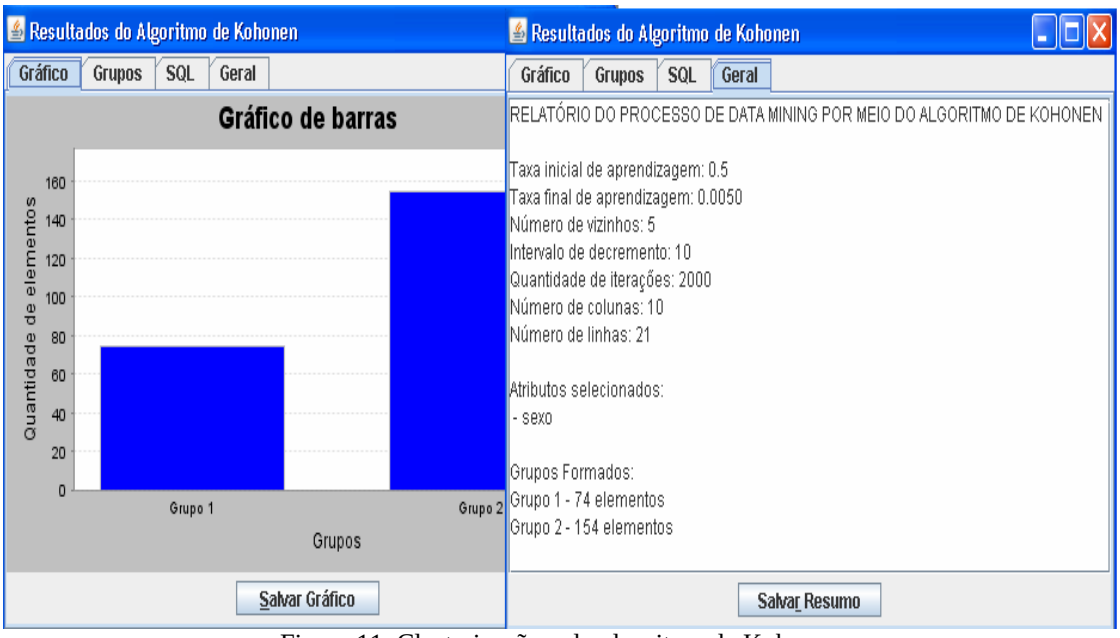


Figura 11. Clusterização pelo algoritmo de Kohonen
Fonte: BORTOLOTTO, L. (2007)

Ainda no módulo de clusterização tem-se o algoritmo K-Means implementado (MARTINS, 2007). Segundo Goldschmidt e Passos (2005) o K-Means consiste em selecionar de forma aleatória *k* atributos da base de dados que serão elementos centrais dos *clusters*. Este algoritmo faz um cálculo da distância das outras informações em relação aos *clusters* centrais e agrupa os dados que possuem maior similaridade.

Na Figura 12 tem-se os resultados da clusterização pelo algoritmo K-Means na *Shell Orion Data Mining Engine*.

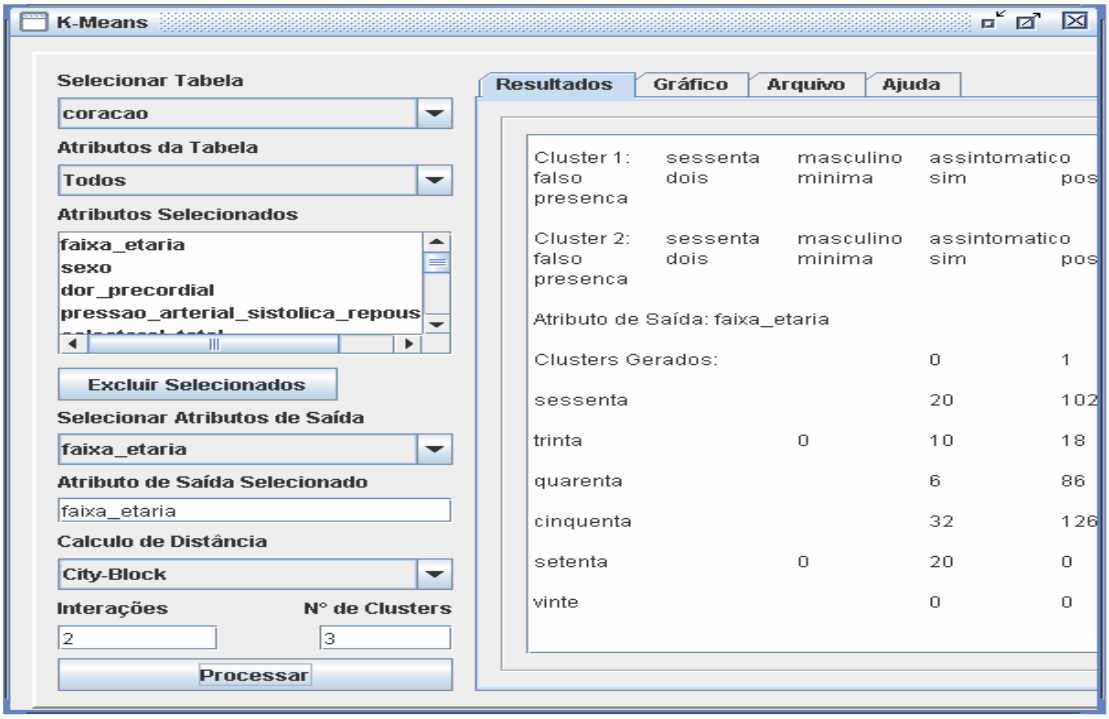


Figura 12. Clusterização pelo algoritmo de K-Means
Fonte: MARTINS, D. (2007)

No módulo de clusterização tem-se ainda implementado o método de lógica fuzzy por meio do algoritmo de Gustafson-Kessel desenvolvido (Figura 13) (CASSETARI JÚNIOR, 2008).

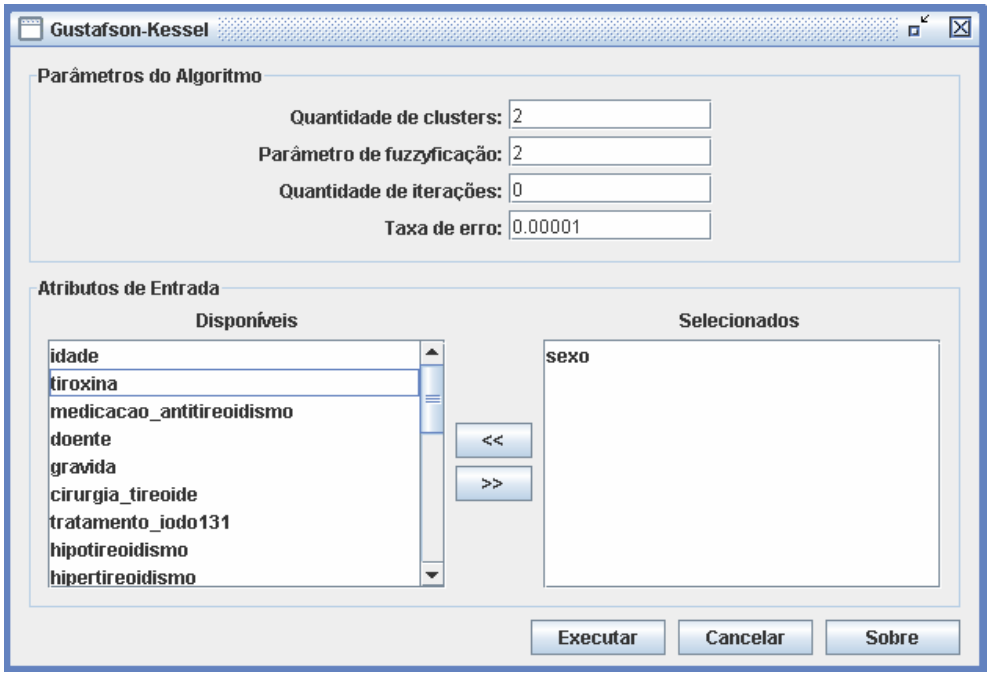


Figura 13. Módulo de clusterização pelo método fuzzy
Fonte: CASSETARI JUNIOR, J.(2008)

O Algoritmo Gustafson-Kessel tem a finalidade de agrupar as informações de uma base de dados de forma precisa, contudo apenas atributos numéricos podem ser clusterizados (GUSTAFSON; KESSEL, 1979). Na *Shell* são pedidos alguns parâmetros para a realização deste algoritmo: quantidade de *clusters*, parâmetro de fuzzyficação, quantidade de interações e taxa de erro. Na Figura 14 tem-se a clusterização por meio deste algoritmo.

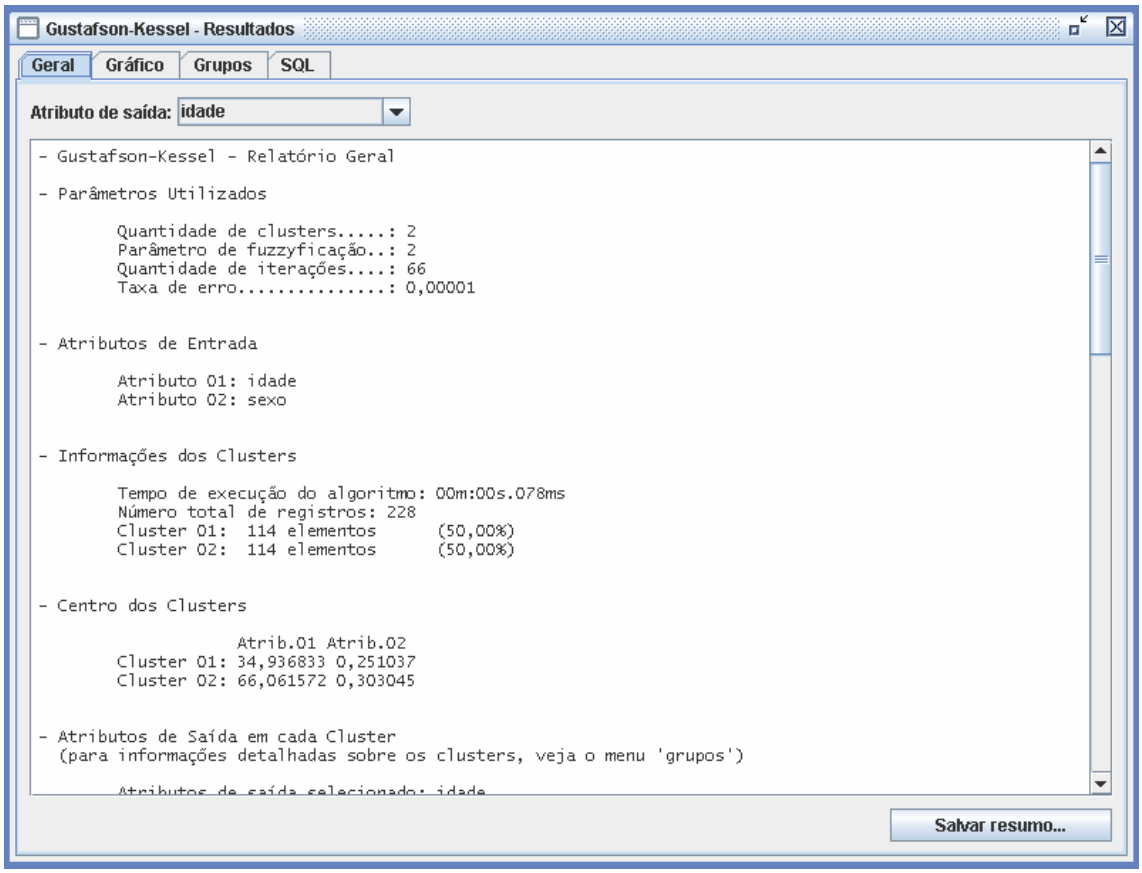


Figura 14. Clusterização pelo método fuzzy por meio do algoritmo de Gustafson-Kessel
Fonte: CASSETARI JUNIOR, J.(2008)

Ainda na clusterização pelo o método de lógica fuzzy por meio do algoritmo de Gustafson-Kessel na Shell Orion, pode-se visualizar o gráfico gerado (Figura 15).

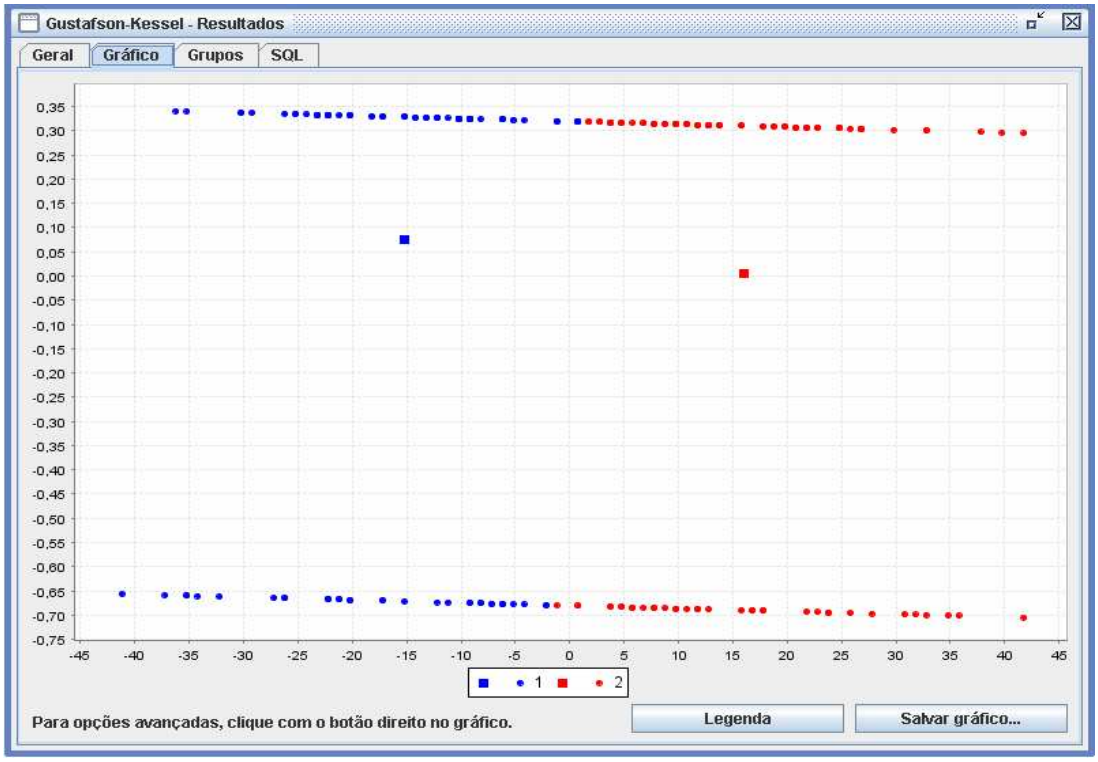


Figura 15. Gráfico gerado por meio do algoritmo de Gustafson-Kessel
Fonte: CASSETARI JUNIOR, J.(2008)

Finalizando-se os estudos sobre o processo de DCBD, em especial a etapa de *data mining*, responsável pela busca de informações relevantes em grandes bases de dados, tem-se a necessidade de representar o conhecimento obtido. No entanto, para isso podem-se utilizar algumas formas que apresentam essas informações de modo compreensível. Assim se faz necessário o estudo sobre representação do conhecimento.

3 REPRESENTAÇÃO DO CONHECIMENTO

A representação do conhecimento é um dos objetivos base da Inteligência Computacional (IC), que tenta compreender o comportamento humano e construir sistemas computacionais (BARRETO, 2001).

As formas de representação do conhecimento podem ser entendidas como procedimentos que junto com as estruturas de informações buscam modelar de forma hábil os conhecimentos adquiridos e disponibilizá-los por meio de sistemas inteligentes permitindo o acesso dos usuários (BUENO, 2005).

Rezende (2005) diz que a representação do conhecimento é uma forma ordenada de estruturar e codificar o conhecimento sobre uma determinada aplicação. Entretanto, deve possuir as seguintes características:

a) ser compreensível ao ser humano;
b) abstrair-se dos detalhes de como funciona;
c) ser robusta, isto é, permitir sua utilização mesmo que não aborde todas as situações possíveis;
d) possa ser atribuída a diversas situações e interpretações.

Na representação do conhecimento várias técnicas foram desenvolvidas para solucionar problemas de eficiência, facilidade de uso, conhecimentos incertos e incompletos. No entanto, não existe uma forma geral de representá-lo (HINZ, 2006).

3.1 FORMAS DE REPRESENTAÇÃO DO CONHECIMENTO

As formas de representação do conhecimento são de suma importância para a compreensão do domínio da aplicação de forma clara e objetiva. Apresenta-se a seguir

algumas técnicas utilizadas para a representação do conhecimento, por serem as mais citadas na literatura: representação lógica, redes semânticas, *frames*, redes bayesianas e ontologias.

3.1.1 Representação Lógica

A representação lógica é uma linguagem formal, que identifica se uma seqüência de símbolos está de acordo com as regras de construção da linguagem de programação. Nesta etapa de verificação têm-se várias regras sintáticas de dedução automáticas, ou seja, formas de inferências dedutivas[17] a partir do formato sintético[18] das expressões da linguagem sem basear-se em idéias extras e intuitivas (LINHALIS, 2007).

A linguagem de programação PROLOG[19] é uma forma de representação lógica e fornece uma descrição do problema que se pretende computar. Esta representação utiliza uma coleção de *fatos* e de relações lógicas (*regras*) que representam o domínio do problema a resolver (REZENDE, 2005). Na Figura 16 tem-se um exemplo desta linguagem.

```
homem(Georges) ->;                    pai(x,z)  pai(z,f);
homem(Richard) ->;                    avohomem(x,f) ->
mulher(Mary) ->;                      mae(x,z)  pai(z,f);
mulher(Doralice) ->;                  avomulher(x,f) ->
pai(Georges,Charles) ->;              pai(x,z)  mae(z,f);
pai(Ane,Mark) ->;                     avomulher(x,f) ->
mae(George, Rose) ->;                 mae(x,z)  mae(z,f);
mae(Ane,Dora) ->;
avohomem(x,f) ->
```

Figura 16. Árvore genealógica
Fonte: BARRETO, J. (2001)

[17] Consiste na argumentação das hipóteses de um programa com base nos fatos e regras no intuito de provar as metas que o sistema deseja atingir (LINHALIS, 2007).
[18] Formato das regras da sintaxe do programa (LINHALIS, 2007).
[19] Linguagem de programação que utiliza teoremas onde se tem fatos e regras (BITTENCOURT, 2001).

3.1.2 Redes Semânticas

As redes semânticas são notações gráficas rotuladas formadas por um conjunto de vértices (nós) que representam os objetos e por arestas (arcos) que são as relações entre estes objetos. O arco é rotulado com o nome da relação podendo existir vários, porém cada objeto pode ser representado apenas por um nó (RUSSEL; NORVIG, 2004). Na Figura 17 tem-se um exemplo de rede semântica onde João, Maria, 1 e 2 são os objetos. Mamíferos, Pessoas, Pessoas femininas e Pessoas Masculinas são as categorias. Por fim as relações são representadas por meio de arcos rotulados.

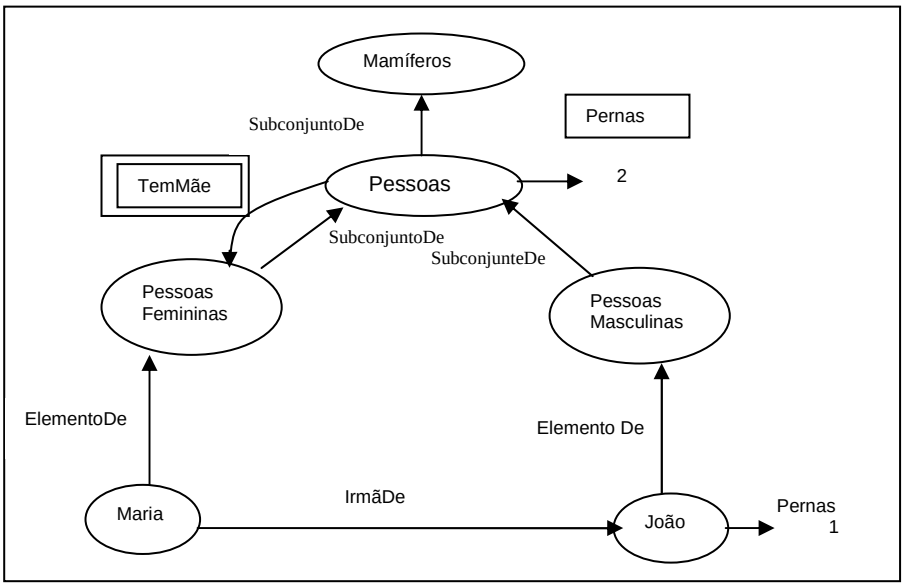


Figura 17. Rede semântica
Fonte: RUSSEL, S; NORVIG, P.(2004)

3.1.3 Frames

Frames (quadros) são estruturas utilizadas para designar conhecimentos sobre um grupo de objetos permitindo a demonstração das estruturas internas, bem

como mantendo a possibilidade de representar a herança existente entre os mesmos (LUGER, 2004).

Essas estruturas são expressas de forma organizada, possuem um nome que identifica o conceito definido e um grupo de propriedades, chamados de *slots* (REZENDE, 2005). Além disso, também apresentam estruturas pré-definidas para especificar restrições de tipo, domínio, valor *default* e cardinalidade sobre o conjunto de propriedades (BARRETO, 2001). Na Figura 18 tem-se um exemplo de *frames* descrevendo o quarto de hotel e seus componentes, cada *frame* pode ser visto como uma estrutura de dados.

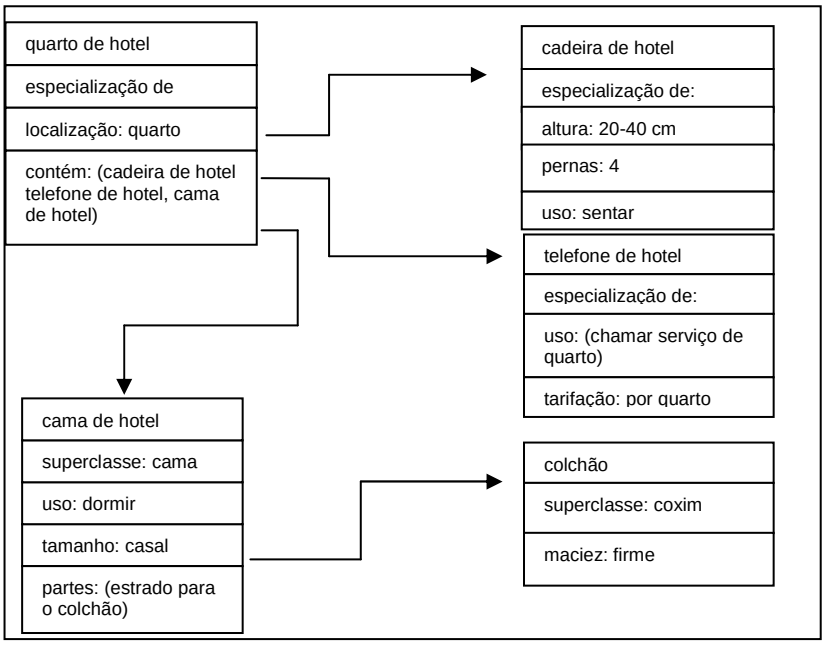


Figura 18. Exemplo de um Frame
Fonte: LUGER, F. (2004)

3.1.4 Redes Bayesianas

Redes Bayesianas são representações das relações de probabilidades entre sentenças de um problema. Abordam o raciocínio probabilístico e modelam sistemas

onde há presença de incerteza por aleatoriedade[20], utilizando-se para isso do teorema de Bayes[21] (LAROSE, 2005, tradução nossa).

Uma rede bayesiana é um grafo orientado em que cada nó é identificado com as informações das relações de causalidade das variáveis de um sistema. Para a aplicação do teorema de Bayes três termos são necessários: uma probabilidade condicional e duas incondicionais (LUGER, 2004).

Na Figura 19 apresenta-se um exemplo de rede bayesiana: um alarme foi instalado contra ladrões. Este alarme é muito confiável, contudo, ele pode disparar caso ocorra um terremoto. Os vizinhos, João e Maria, ficaram de telefonar para o dono da casa no trabalho caso o alarme dispare. João sempre liga quando ouve o alarme, entretanto, algumas vezes confunde o alarme com o telefone e também liga nestes casos. Maria, por outro lado, gosta de ouvir música alta e às vezes não escuta o alarme.

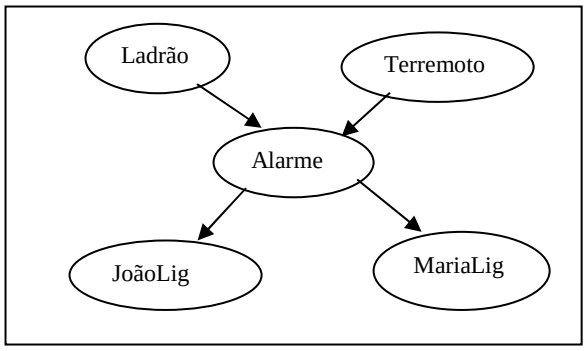


Figura 19. Exemplo de Rede bayesiana
Fonte RUSSEL, S; NORVIG, P (2004)

Neste caso, têm-se as probabilidades em que o alarme pode falhar e ainda, as condições em que João e Maria podem não estarem presentes, não ouvirem o alarme, entre outros. Assim, as probabilidades condicionais para cada caso podem ser criadas.

[20] Conhecimento incerto a cerca da resolução de um problema (RUSSEL; NORVIG, 2004).
[21] Fórmula para calcular as probabilidades do conhecimento incerto (RUSSEL; NORVIG, 2004).

3.1.5 Ontologias

A palavra ontologia teve origem na Filosofia com o objetivo de estudar a existência do ser, mais precisamente na análise das categorias de coisas que existem em algum domínio (MAEDCHE, 2002, tradução nossa).

Na Inteligência Computacional é utilizada como forma de representação de um domínio a partir de seus conceitos abstratos e como se relacionam (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa). Na Figura 20 tem-se um exemplo de ontologia sobre o domínio doenças sexualmente transmissíveis.

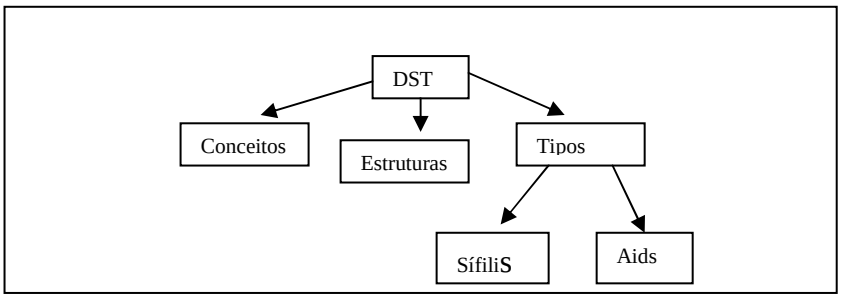


Figura 20. Exemplo de uma ontologia da Doença Sexualmente Transmissíveis
Fonte: FARIAS, R. (2006)

A ontologia pode ser entendida como uma descrição de conceitos de um domínio com suas classes[22], relações[23], funções[24], axiomas[25] e instâncias[26]. Além disso, é organizada por meio de hierarquias dos conceitos (STAAB; STUDER, 2004, tradução nossa). Contudo, por ser objeto de estudo desta pesquisa será abordada com maior ênfase na próxima seção.

[22] Representam os conceitos (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).
[23] Tipos de iterações entre os conceitos do domínio (ALMEIDA; BAX, 2003).
[24] São casos especiais de relações entre os conceitos (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).
[25] Modelam sentenças que são sempre verdadeiras estruturais ou semi estruturais (ALMEIDA; BAX, 2003).
[26] São os conhecimentos da ontologia desenvolvida (STAAB; STUDER, 2004, tradução nossa).

3.2 ONTOLOGIAS

O termo Ontologia originou-se na Filosofia, idealizada por Aristóteles na metafísica[27] um ramo que visa à organização da realidade e a descoberta das características comuns entre os seres (MAEDCHE, 2002, tradução nossa).

Aristóteles apresenta ontologia como sendo o estudo das categorias que classificam entidades em domínios com propriedades distintas (ALMEIDA; BAX, 2003).

Contudo, recentemente passou a ser utilizado também pela IC, na *web* semântica, gestão de conhecimento, representação do conhecimento, bancos de dados, recuperação, organização e descoberta de informações, no sentido de estabelecer conceitos e termos que podem ser usados para descrever ou desenvolver alguma área da descoberta de conhecimento (MCGUINNESS, 2002, tradução nossa).

Na gestão do conhecimento as organizações buscam por meio das ontologias a estruturação de suas informações, organização, e principalmente a representação do conhecimento adquirido para descobrir novas relações ajudando na tomada de decisão (GAŠEVIĆ; DJURIĆ; DEVEDŽIĆ, 2006, tradução nossa).

Na *web* semântica, devido ao fato da Internet possuir um grande volume de informações tem-se a necessidade de transformá-los em conhecimento e associar cada conceito a um grupo. Assim a ontologia tem sido utilizada na *web* para a obtenção de uma organização semântica das páginas nos *sites* (HINZ, 2006).

Na IC o conceito de ontologia foi adaptado e definido que seria responsável pela estruturação e organização de uma base de conhecimento, com o intuito de facilitar a sua compreensão (NIGRO; CÍZARO; XODO, 2008, tradução nossa).

---

[27] Ramo da filosofia que trata da natureza da existência do ser estuda o mundo como ele é ou a realidade da verdade e do conhecimento (MAEDCHE, 2002).

Segundo Staab, Studer (2004, tradução nossa p.172), as ontologias “constituem uma conceitualização formal de um domínio de interesse particular que é compartilhado por um grupo pessoas”.

De acordo com essa definição descreve-se conceitualização como sendo basicamente um modelo abstrato de algum aspecto do mundo, fazendo uma definição das propriedades dos conceitos e seus relacionamentos.

De acordo com Noy e McGuinness (2001, tradução nossa) o uso de ontologias proporciona:

a) **compartilhamento da mesma estrutura de conhecimento entre áreas de interesses comuns**: sistemas interligados podem permitir o compartilhamento de uma mesma ontologia agregando assim mais informações e conhecimento;
b) **eficiência na estruturação do conhecimento**: a forma clara com que defini os conceitos e restrições possibilita a maior validação do conhecimento;
c) **facilidade de manutenção e documentação das ontologias**: a estruturação do conhecimento por meio de ontologias de forma genérica facilita seu suporte;
d) **reuso do conhecimento ou domínio**: a ontologia criada pode ser reutilizada ou integrada a outras;

No entanto, segundo Menzies (1999, tradução nossa), o reuso pode ter um alto custo mediante a adequação dos sistemas e o tempo necessário para que se aprenda a utilizar essa ontologia.

Estruturalmente as ontologias não apresentam sempre os mesmos formatos, mas possuem na maioria características e componentes comuns. Contudo, mesmo apresentando propriedades distintas, possuem tipos bem definidos (HINZ, 2006).

Existem quatro tipos propostos de classificação de ontologias (GUARINO, 1998, apud MAEDCHE, 2002, p. 23, tradução nossa):

a) **ontologias de alto nível**: descrevem conceitos de forma mais abrangente e definem apenas termos gerais como o espaço, tempo, evento, entre outros. Estes servem de base para outras ontologias independentes de um problema ou domínio específico. Na Figura 21 tem-se um exemplo de ontologia de alto nível onde observa-se a classe Universidade com suas subclasses Professor e Aluno .

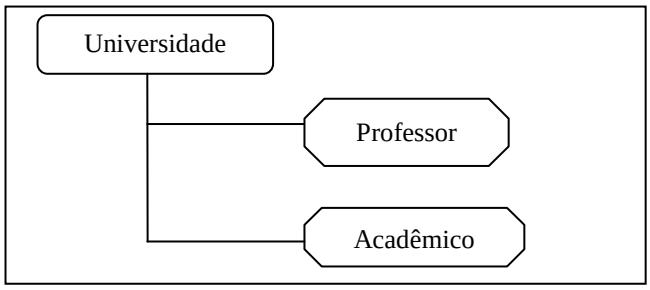


Figura 21. Ontologia de alto nível

b) **ontologias de domínio**: descrevem o conjunto de termos relacionados a um domínio genérico, por meio da especialização de conceitos, bem como a conceitualização de um domínio particular. Na Figura 22 tem-se um exemplo desta ontologia, observa-se a classe Disciplina e suas subclasses Filosofia e Matemática;

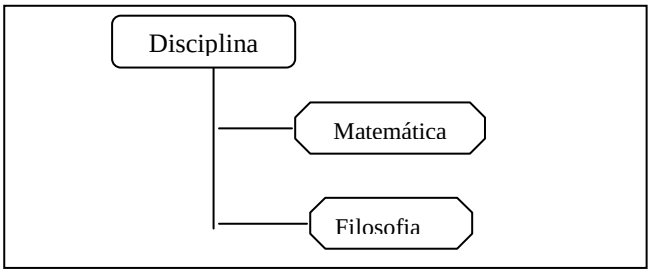


Figura 22. Ontologia de domínio

c) **ontologias de tarefas**: buscam descrever um vocabulário de termos relacionado a uma tarefa ou atividade genérica por especialização de uma ontologia de alto nível para a resolução de problemas, que podem ou não ser do mesmo domínio. Na Figura 23 tem-se um exemplo de ontologia de tarefa, pois a classe Universidade e suas subclasses conceitualizam a ontologia de alto nível;

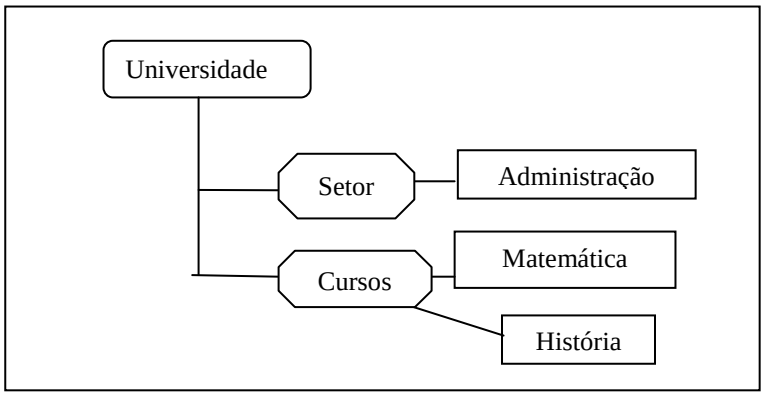


Figura 23. Ontologia de tarefa

d) **ontologias de aplicação**: procuram solucionar problemas específicos e descrevem os conceitos dos domínios da ontologia de tarefa em uma determinada aplicação, por possuírem esta característica têm a menor capacidade de reuso. Na Figura 24 têm-se um exemplo de ontologia de aplicação, pois conceitualiza a ontologia de tarefa Universidade.

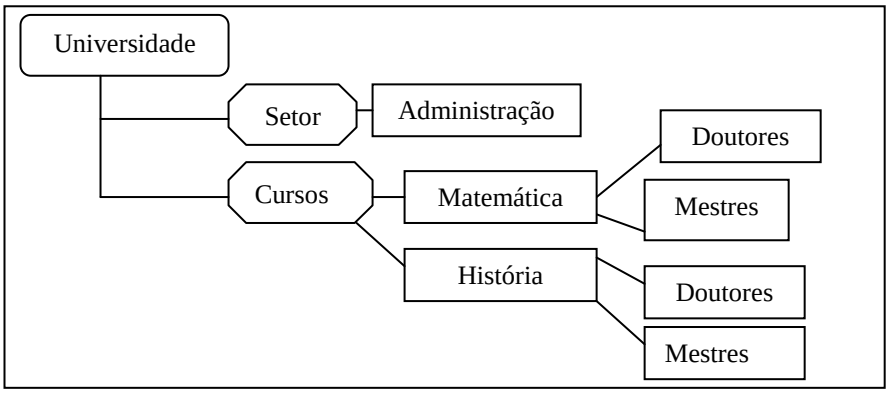


Figura 24. Ontologia de aplicação

Na construção de uma ontologia se faz necessário identificar o tipo, domínio a ser desenvolvido e a compreensão dos componentes que a constituem. O domínio de uma ontologia é caracterizado por meio da representação dos seguintes componentes (NOY; MCGUINNESS, 2001, tradução nossa):

a) **conceitos**: apresenta-se organizados em taxionomias, trazendo relações hierárquicas entre seus conceitos, estabelecendo relacionamentos entre e classes, subclasses. Na Figura 25 tem-se um exemplo do conceito pessoa subconceito homem;

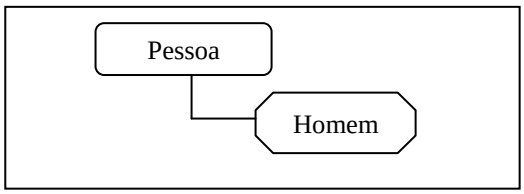


Figura 25. Exemplo de conceitos

b) **relações**: representam os tipos de interações entre os conceitos de um domínio específico. A Figura 26 mostra um exemplo das relações entre conceitos;

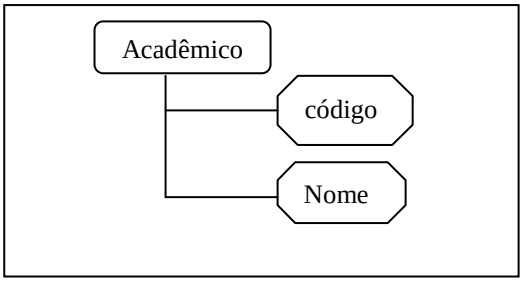


Figura 26. Exemplo de relações

c) **funções**: casos especiais para a representação de relacionamentos, não são empregados de forma geral, mas como uma relação única com um outro elemento. Definem as propriedades de uma relação e restringi seus valores tipo (inteiro, decimal, entre outros). Na Figura 27 tem-se um exemplo de função onde o atributo nome é caractere e limita-se a 200;

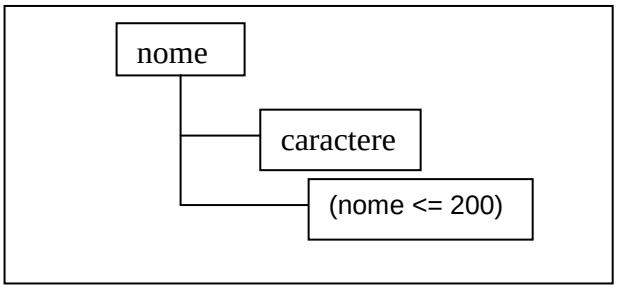


Figura 27. Exemplo representação de função

c) **axiomas**: modelam sentenças sempre verdadeiras, são aplicados normalmente para representar o conhecimento definindo restrições sobre as relações e classes, bem como para a verificação de correções. Na Figura 28 observa-se um exemplo de axioma;

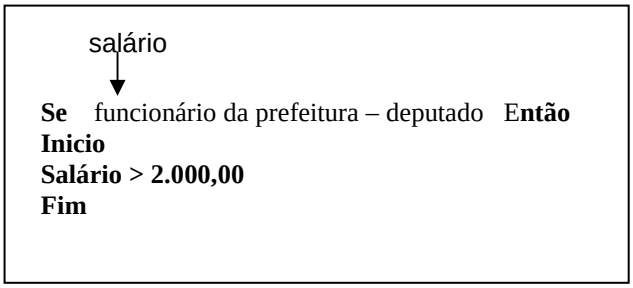


Figura 28. Representação de axioma

d) **instâncias**: são usadas para representar os elementos específicos de um domínio em uma ontologia, ou seja, é o próprio conhecimento que existe na ontologia a ser representada. Na Figura 29 observa-se exemplos de instâncias;

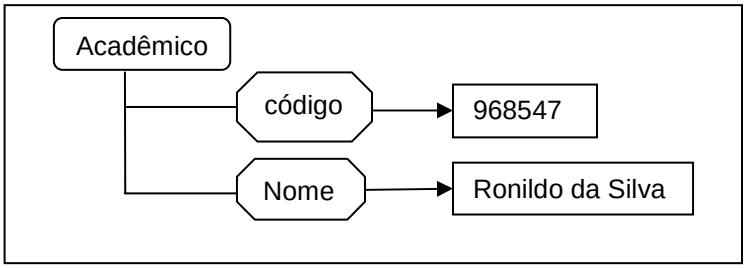


Figura 29. Representação de uma instância

Compreendidas as ontologias e suas definições, tipos, componentes, no entanto para o desenvolvimento das mesmas nas bases de dados se faz necessário as metodologias disponíveis.

3.2.1 Metodologias para Construção de Ontologias

O processo de construção de ontologias encontra-se pouco desenvolvido não existindo um consenso sobre uma metodologia por isso muitos desenvolvedores acabam utilizando suas próprias normas para a construção de ontologias (MARTIMIANO, 2006).

Devido a essa falta de padronização muitos problemas têm ocorrido durante o processo de desenvolvimento de ontologias como: os modelos conceituais ficam subentendidos na implementação e acabam gerando dificuldades na construção de ontologias complexas (HINZ, 2006).

Isso deve-se ao fato de que muitos desenvolvedores passam da fase de aquisição de conhecimento diretamente para a fase de implementação sem um planejamento adequado (FARIAS, 2006).

Contudo, com a intenção de organizar a construção e manipulação de ontologias, têm-se desenvolvido metodologias com a finalidade de reduzir as dificuldades encontradas no seu desenvolvimento e facilitar seu reuso (LÓPEZ, 2000, tradução nossa).

Metodologias para a construção de ontologias existem várias, a seguir tem-se uma breve descrição sobre as mais citadas pela literatura: Uschold e King's , Gruninger e *F*ox, Kactus, On-to-knoweledge e Methontology (ALMEIDA; BAX,

2003). No entanto estas metodologias encontram-se de forma mais detalhadas no Apêndice A.

A metodologia Uschold e King's foi desenvolvida por Uschold e King em 1995, com o intuito de desenvolver uma ontologia voltada para a modelagem de negócios empresariais baseada na experiência do projeto Enterprise Ontology[28] (ESBÍZARO, 2006). As atividades para o desenvolvimento das ontologias oferecidas por esta metodologia são: identificar o propósito da ontologia, construí-la, avaliá-la e documentá-la (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).

Gruninger e Fox foi a metodologia empregada no projeto Toronto Virtual Enterprise (TOVE)[29], desenvolvido pelo Enterprise Integration Laboratory, da Universidade de Toronto, em 1995. Volta-se para as atividades de negócio e basicamente envolve a construção de um modelo lógico do conhecimento que especifica as carências da ontologia (ESBÍZARO, 2006). Este método considera possíveis aplicações da ontologia por isso, depende parcialmente de aplicação. O processo de construção não é desenvolvido diretamente primeiramente uma descrição informal é elaborada da especificação da ontologia e após isto a descrição é formalizada (LINHALIS, 2007).

A metodologia *KACTUS* é um projeto criado por Amaya Berneras em 1996 e seus colegas, dentro do *KACTUS project* que visa o desenvolvimento de uma metodologia para reutilização dos conhecimentos adquiridos por meio das ontologias em todas as suas fases de construção (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa). Esta metodologia utiliza a mesma base de

[28] Projeto de ontologias possui uma coleção de termos e definição dos domínios empresariais (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).

[29] Projeto para o desenvolvimento de ontologias na modelagem de atividades empresariais (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).

conhecimento na construção da ontologia do domínio definido podendo ser reutilizada ou integrada a outras em diferentes áreas.

Na metodologia On-To-Knowledge o objetivo é auxiliar na análise dos processos empresariais e os diferentes papéis do conhecimento nas organizações, identificando metas e utilizando contribuições dos provedores e clientes da organização (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).

A Methontology foi proposta com a finalidade do desenvolvimento de ontologias a partir de um ponto inicial, reutilizando o conhecimento por reengenharia. Seu ciclo de vida baseia-se na evolução de protótipos, permitindo que suas definições possam ser adicionadas, removidas ou modificadas conforme sua necessidade (LINHALIS, 2007). Esta metodologia, contudo por ter sido utilizada nesta pesquisa será detalha na seção 3.2.1.1.

3.2.1.1 Methontology

A Methontology foi desenvolvida pelo Laboratório de Inteligência Artificial da Universidade Politécnica de Madrid (Espanha) e possibilita a construção de ontologias no nível de conhecimento (ESBÍZARO, 2006).

Esta metodologia tem suas atividades principais identificadas pelo processo de desenvolvimento de software do Institute of Electrical and Electronics Engineers (IEEE). Seu ciclo de vida é baseado em: evolução de protótipos, pois permite adicionar, mudar ou remover termos em uma nova versão da mesma ontologia; administração das técnicas de cada atividade e suporte (LINHALIS, 2007).

As fases de construção de ontologias por meio desta metodologia são (FERNÁNDEZ-LÓPEZ, GÓMEZ-PÉREZ, JURISTO, 1997):

a) **especificação**: nessas fases, são realizadas todas as tarefas que envolvem a definição do propósito da ontologia objetivando a elaboração de um documento com todos os requisitos;

b) **aquisição do conhecimento**: visa a aquisição do conhecimento, apesar de ser um dos passos iniciais está presente em todo o processo de desenvolvimento de ontologias, tendendo a diminuir ao longo do mesmo;

c) **conceitualização**: nesta etapa o conhecimento adquirido é estruturado em um modelo conceitual que descreve os conceitos, propriedades e restrições do seu domínio. Esta etapa pode ser considerada a mais importante;

d) **formalização**: objetiva formalizar o modelo conceitual da etapa anterior por meio de uma linguagem formal, entretanto no desenvolvimento de ontologias tem-se ferramentas que implementam o modelo conceitual por meio de linguagens de descrição, por isso essa etapa não é obrigatória;

e) **integração**: tem como objetivo a integração de ontologias em fase de desenvolvimento com as já existentes;

f) **implementação**: consiste em implementar o modelo conceitual por meio de uma linguagem de programação;

g) **avaliação**: fase onde a ontologia construída passa pelo processo de verificação e validação do conhecimento, por meio de técnicos, desenvolvedores e usuários;

h) **documentação** : tem o objetivo de auxiliar na manutenção, bem como facilitar a reutilização da ontologia, por isso todo o processo deve ser documentado;

i) **manutenção**: é nesta etapa onde as alterações necessárias são realizados a fim de que ocorra uma adequação da ontologia para sua reutilização.

De acordo com esta metodologia as fases de aquisição do conhecimento, avaliação e documentação acontecem durante todo o ciclo de vida de uma ontologia (LINHALIS, 2007). Além disso, a característica principal da Methontology é a geração dos componentes no desenvolvimento das ontologias na fase de conceitualização (Figura 30).

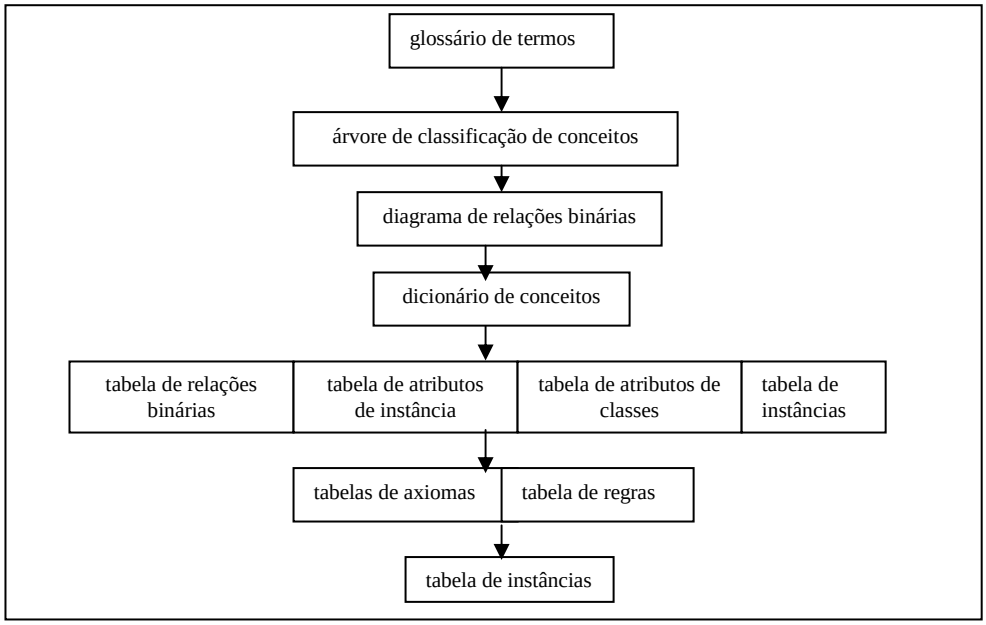


Figura 30. Componentes gerados na fase de conceitualização da methontology
Fonte: Adaptado de GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, (2004)

A seguir tem-se um detalhamento dos componentes observados anteriormente (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa):

a) **glossário de termos**: todos os conceitos e relações do domínio da ontologia criada e as suas descrições são incluídos no glossário;

b) **árvore de classificação de conceitos**: os termos do glossário são representados por meio de árvores (conceitos e subconceitos);

c) **diagrama de relações binárias**: estabelece os relacionamentos entre os conceitos sobre ontologias do mesmo domínio ou diferentes;

d) **dicionário de conceitos**: apresenta as relações, funções e axiomas, do domínio da ontologia. Cada árvore de classificação deve ter um dicionário de conceitos;

e) **tabela de relações binárias**: por meio de tabelas apresenta as relações binárias da ontologia, tais como: o nome da relação, os nomes dos conceitos de origem e destino, cardinalidade (valores que pode assumir), entre outros;

f) **tabela de atributos de instâncias**: detalham os atributos das instâncias incluídos no dicionário de conceitos, tais como: nome, faixa de valores aceitos, tipo (caractere, inteiro, decimal), entre outros;

g) **tabela de atributos de classe**: informam os atributos de classes do dicionário de conceitos que apresentam os mesmos valores para todas as instâncias do conceito. Cada linha da tabela de atributo de classe contém uma descrição detalhada do atributo de classe;

h) **tabela de axiomas**: este componente para ser desenvolvido deve identificar os axiomas formais da ontologia e os apresentar de forma minuciosa. Cada axioma tem que ser definido com nome, descrição em linguagem natural, atributos usados, conceitos, referências e as regras que definem o axioma formalmente;

i) **tabela de constantes**: neste componente para cada constante são especificados: nome, descrição em linguagem natural, valor constante, tipo (caractere, inteiro, decimal, entre outros);

j) **tabela de fórmulas**: são informados todas as fórmulas das tabelas de atributos de instância, especificando-se: nome da fórmula, expressão matemática, atributos presentes, atributos e constantes utilizados no cálculo e descrição em linguagem natural;

k) **tabelas de instâncias**: tendo o modelo conceitual da ontologia criada, tem-se que descrever os exemplos das instâncias do domínio, por meio dos nomes e valores dos atributos de instância.

Esses componentes tornam a documentação da ontologia mais completa, facilitando todo o processo de construção, uso, manutenção e ainda reutilização. Por isso, optou-se por utilizar a Methontology nesta pesquisa.

Compreendidas as metodologias e suas etapas, o próximo passo nesta pesquisa é a construção da ontologia.

A construção de ontologias é uma tarefa complexa e demorada, contudo existem ferramentas que auxiliam no desenvolvimento, manutenção e posterior utilização. Estes ambientes proporcionam aos desenvolvedores facilidades como: interfaces que permitem visualizar todo o processo de desenvolvimento da ontologia, disponibilizam o compartilhamento de ontologias com o objetivo de um possível reuso, entre outros (LINHALIS, 2007).

Neste estudo estudaram-se várias ferramentas para o desenvolvimento de ontologias (Apêndice B). Contudo a Protégé foi utilizada nesta pesquisa por ser disponibilizada gratuitamente e possuir uma interface de fácil utilização, sendo abordada na próxima seção.

3.2.2 Protégé

Essa ferramenta foi desenvolvida, em 1987, na Universidade de Standford pelo Departamento de Informática Médica (*Stanford Medical Informatics* – SMI) (GÓMEZ-PÉEREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa). Encontra-se disponível gratuitamente para *download* no *site* http://www.smi.stanford.edu/projects/protege/.

Protégé é implementada em Java e oferece uma interface gráfica para o desenvolvimento de ontologias de forma interativa, bem como uma arquitetura para criação de ferramentas baseadas em conhecimento (NOY, 2000, tradução nossa).

Suas características principais são (STAAB; STUDER, 2004, tradução nossa):

a) modelo de conhecimento ampliável permitindo aos usuários redefinirem o domínio da ontologia;
b) adapta os arquivos para qualquer linguagem formal[30];
c) interface de fácil entendimento para os usuários;
d) integração com outras aplicações por meio dos *plugins*.

Tendo em vista a arquitetura baseada em *plugins* a Protégé possibilita a interação com outras ferramentas, bem como proporciona mais funcionalidades em seu uso. A seguir têm-se alguns exemplos de *plugins*: (GÓMEZ-PÉEREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa):

a) tab *plugins*: adiciona uma aba no editor de ontologias de forma que o usuário possa ter acesso as outras funções;

[30] Linguagem para a representação do conhecimento de uma ontologia (STAAB; STUDER 2004, tradução nossa).

b) *slot widgets*: interface de extensão para os usuários exibirem e editarem os valores dos *slots*;

c) *backends*: permite aos usuários habilitarem as ontologias para serem exportadas e importadas em diferentes formatos: RDF Schema, XML, XML Schema, entre outros.

A interface da Protégé (Figura 31) apresenta as seguintes áreas de visualização (*views)* para o desenvolvimento das ontologias: Classes (onde as classes são criadas), *Slots* (relações entre as ontologias), *Forms* (visualização dos formulários), *Instances* (inserção do conhecimento nas instâncias) e *Queries* (consultas).

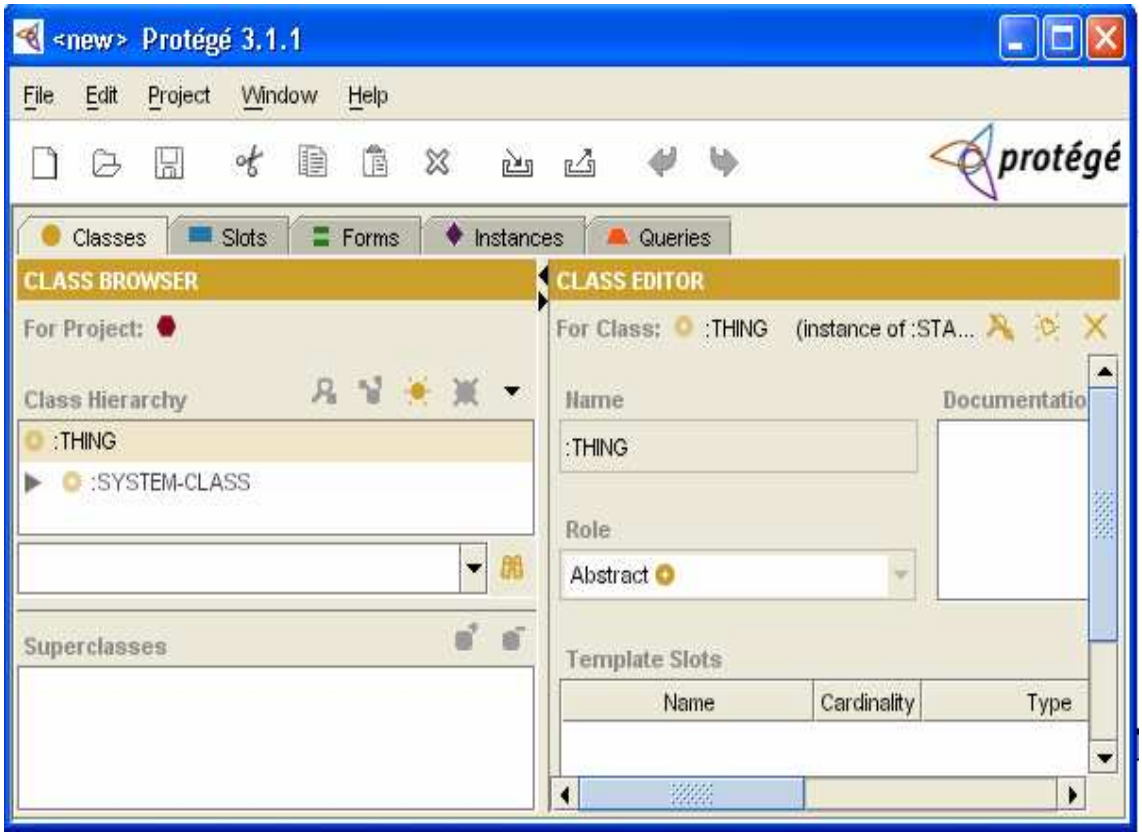


Figura 31. Interface principal do Protege
Fonte: Stanford Medical Informatics (2007)

Na Protégé o modelo de conhecimento é baseado em *frames* e lógica de primeira ordem[31], tendo-se no momento da construção os seguintes passos: definir um

[31] Permite descrever e raciocinar sobre objetos e predicados que especificam propriedades de objetos ou relacionamentos entre objetos de um domínio (LINHALIS, 2007).

esquema com as classes (*class*), subclasses (*subclass*), propriedades e relações (*slots*) referentes ao domínio que se deseja modelar.

Apesar de não apresentar uma metodologia específica, a Protégé suporta por meio de seus *plugins*, as várias fases da maioria das metodologias para a construção de ontologias, como por exemplo, a Methontology.

Neste capítulo foram abordadas a representação do conhecimento em IC, as técnicas e principalmente ontologias onde se abordou: suas definições, tipos, metodologias e a ferramenta escolhida (Protégé) para sua construção na base de dados têm-se a necessidade de entendimento da técnica de *data mining* em conjunto com ontologias no processo de DCBD para a descoberta de conhecimento sendo essencial para o restante deste estudo.

4 ONTOLOGIAS E DATA MINING

A base de dados deve ser estruturada e organizada de forma que seja detectado padrões úteis, previamente desconhecidos, visando uma melhor compreensão do problema, para isso tem-se o processo já descrito nos itens anteriores denominado de DCBD onde existem as etapas de pré-processamento, *data mining* e pós-processamento.

Češpivová et al (2006, tradução nossa) abordam o uso de ontologias nas fases do DCBD compreendendo:

a) **entendimento do domínio da aplicação**: é de suma importância entender o domínio das ontologias antes da escolha de uma técnica específica;

b) **compreensão dos dados**: as ontologias possibilitam a identificação de atributos que devem ser acrescentados, bem como observar a redundância de informações presentes na base;

c) **preparação**: após a definição e entendimento do domínio da ontologia, bem como das tarefas e métodos de *data mining* a serem utilizados tem-se a fase de preparação que consiste na identificação de vários grupos de atributos e/ou valores de acordo com critérios semânticos;

d) **modelagem**: em grandes bases de dados pode-se utilizar as ontologias na estruturação dos conhecimentos;

e) **avaliação**: os conhecimentos descobertos podem ser estruturados em conceitos (previamente organizados em atributos) a fim de serem interpretados por meio de ontologias;

f) **desenvolvimento**: o conhecimento obtido pode ser modelado novamente pelas ontologias.

Desta forma, entende-se que as ontologias podem ser aplicadas na etapa de *data mining*, que se constitui no núcleo do processo de DCBD, de diferentes maneiras. No *data mining* as informações são exploradas por meio de tarefas e métodos. Contudo, muitas vezes estes dados podem estar no formato inadequado para esta fase ou apresentarem variadas denominações (sinônimos) para um mesmo atributo.

Neste contexto, as ontologias vêm sendo utilizadas pela IC para estruturar os dados na fase de pré-processamento, bem como para organizar os conhecimentos descobertos pelo *data mining* na etapa de pós-processamento. Objetiva-se com isso, obter uma representação mais relevante dos dados.

As ontologias na IC são utilizadas para a representação do conhecimento e definem modelos conceituais que especificam os domínios e as relações entre eles (GRUBER, 1993, tradução nossa). A motivação para a aplicação de ontologias é a forma eficiente com que representam o conhecimento sobre os domínios (CIMIANO et al., 2004, tradução nossa).

Assim, podem ser aplicadas em conjunto com a técnica de *data mining* da seguinte forma:

a) **ontologias para *data mining*:** incorpora-se o conhecimento no processo pelo uso de ontologias anteriormente ao *data mining*, utilizando-se por exemplo para a interpretação e validação de conhecimentos (BERNSTEIN et al, 2005, tradução nossa);
b) ***data mining* para ontologias:** usa-se as ontologias para a representação e análise dos resultados. Tradicionalmente, empregam-se nas áreas de Medicina, Biologia e Dados Espaciais, como por exemplo, em aplicações médicas que possuam domínio específico e representações

genéticas (BREZANY et al, 2004; CANNATARO et al, 2003, tradução nossa).

Dessa forma, quando se representa e inclui conhecimento no processo por meio de ontologias, efetivamente pode-se transformar o *data mining* (mineração de dados) em *knowledge mining* (mineração de conhecimento) (CÍSARO; NIGRO; XODO, 2008, tradução nossa).

Este processo pode ser observado na Figura 32 onde tem-se, segundo Císaro, Nigro e Xodo (2008, tradução nossa), o ciclo do *knowledge mining* que é composto basicamente pelos metadados das ontologias, domínio da ontologia e ontologias para o processo de *data mining*.

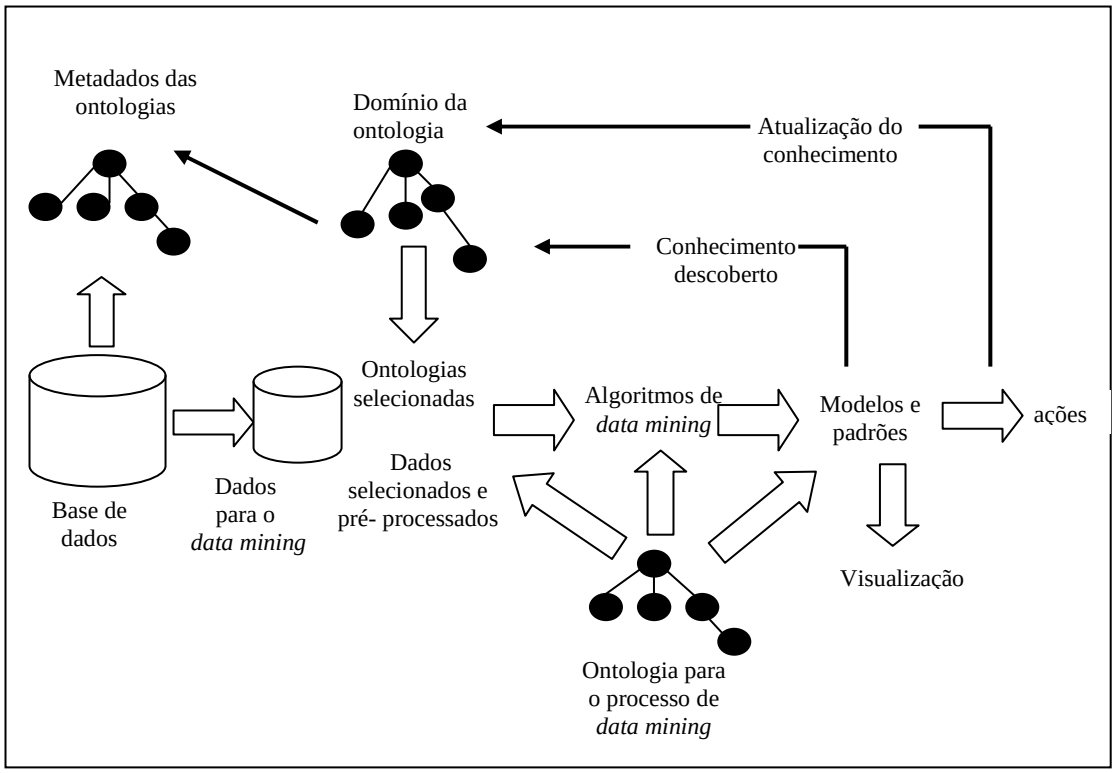


Figura 32. *Data mining* e ontologias
Fonte: Adaptado de CÍSARO, S.; NIGRO, H.; XODO, D. (2008)

4.1 METADADOS DAS ONTOLOGIAS

No processo de *data mining* uma característica interessante indica a integração entre metadados e ontologias, pois as linguagens para descrição de ontologias podem ser utilizadas como metadados. Assim, a engenharia de ontologias por meio dessas linguagens pode prover vocabulários de metadados para facilitar a gestão, descoberta e recuperação da informação. Nos metadados o prefixo *meta* significa um nível de descrição mais alto, assim pode-se dizer que consiste na descrição dos dados, tendo-se, portanto dados sobre os dados (GAŠEVIĆ; DJURIĆ; DEVEDŽIĆ, 2006, tradução nossa).

Neste modelo identificam-se vários conhecimentos ocultos sobre o domínio dos dados a fim de construir as ontologias e melhorar o processo de descoberta do conhecimento (GOTTGTROY et al, 2003, tradução nossa).

No entanto, os modelos de dados representam a estrutura e integridade destes elementos que são desenvolvidos para uma aplicação específica. Conseqüentemente, a conceitualização e o vocabulário do modelo, a princípio, não permitem seu compartilhamento por outras aplicações (GOTTGTROY et al, 2003, tradução nossa).

As ontologias e os modelos de dados são similares, pois ambos dependem do contexto da informação para a especificação de seus termos e questões, além disso, possuem tarefas intensivas na aquisição do conhecimento. Contudo, na construção das ontologias não existe uma separação bem definida entre o conhecimento genérico e específico.

Devido as bases de dados não terem o conhecimento semântico exigido para a construção de ontologias teve-se o desafio de desenvolver formas de se obter e representar os domínios específicos .

4.2 DOMÍNIO DA ONTOLOGIA

Muitos cientistas utilizam modelos de diagramas de causa e efeito[32] para a representação do conhecimento do domínio da ontologia. Estes modelos possuem regras e são usados para predizer ou explicar o comportamento em situações específicas.

Diagramas de causa e efeito podem reproduzir ontologias por meio de mapas conceituais que estão discriminados em taxonomias organizadas em: conceitos centrais, principais, secundários e específicos (BRIDEWELL; LANGLEY, 2006, tradução nossa).

Segundo Hinz (2006) taxonomias são sistemas de classificação que agrupam e organizam um conjunto de conhecimentos num domínio de forma hierárquica por meio de herança simples e/ou múltipla (Figura 39).

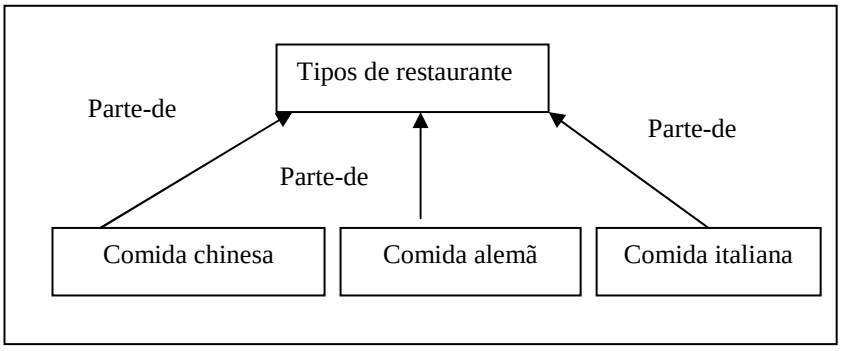


Figura 39. Exemplo de taxonomia
Fonte: HINZ, V. (2006)

Na representação do conhecimento do domínio da ontologia e ao aplicar o DCBD deve-se ter a participação de um especialista nos termos da área, a fim de

[32] Têm o objetivo de estabelecer relações entre o efeito e todas as causas de um processo (RODRIGUES; AMORIN, 1995).

contribuir com informações claras e precisas, proporcionando a atualização e refinamento da base de dados. Em particular na etapa de *data mining* pode-se ter diferentes formas de representação e visualização do conhecimento, como por exemplo, regras, árvores de decisão e *clusters*[33] (BRIDEWELL, 2006; LANGLEY, 2000).

4.3 ONTOLOGIAS PARA *DATA MINING*

Ontologias para *data mining* estão sendo utilizadas devido ao fato que o processo de DCBD passou a discutir a importância de obter-se conhecimento útil e específico sobre o domínio, simplificando assim este processo. Na área médica, por exemplo, onde por meio da experiência tem-se a atualização constante das informações, pode-se representar o conhecimento utilizando ontologias. Bernstein et al (2001, tradução nossa) afirma que a descoberta de conhecimento, quando o *data mining* e ontologias são utilizadas em conjunto possuem vantagens de:

a) validação do processo de *data mining*, na fase de pré-processamento contribuindo na seleção dos dados;
b) utilização de técnicas de *data mining* de acordo com o domínio específico;
c) ontologia permite ao usuário identificar e resolver problemas relacionados a estruturação do conhecimento antes ao *data mining*;
d) permite a reutilização de uma ontologia por outras;
e) possibilita uma melhor representação do conhecimento por meio das ontologias.

[33] Subconjunto de dados agrupados particionados de um registro de uma base de dados (GOLDSCHMIDT; PASSOS, 2005).

O processo de desenvolvimento de ontologias para *data mining* envolve a análise de detalhes como: determinação das chaves primárias e estrangeiras, dependências de inclusão, determinando assim as relações entre as entidades principais identificadas. Pode-se observar que a construção da ontologia é essencialmente baseada em um modelo Entidade Relacionamento (ER) (TRINKUNAS; VASILECAS, 2007 tradução nossa).

Concluindo-se, quando se pode representar e incluir conhecimento no processo de *data mining* por meio de ontologias, consegue-se transformar os dados analisados em conhecimento compreensível e útil.

Assim, conhecendo-se a aplicação de ontologias juntamente com a etapa de *data mining*, a seguir têm-se alguns exemplos da utilização destas duas técnicas em diversas áreas conhecimento.

5 TRABALHOS CORRELATOS

Muitas pesquisas vêm sendo realizadas na área de ontologias, isto deve-se ao fato de que estas podem ser utilizadas para tratar a estruturação e padronização do conhecimento em bases de dados de forma que os padrões encontrados sejam úteis, reutilizáveis e de fácil compreensão. A seguir têm-se de algumas pesquisas relacionadas ás ontologias, e outras referentes a técnica de *data mining* em conjunto com ontologias.

5.1 INTEGRAÇÃO SEMÂNTICA DE DADOS ATRAVÉS DE FEDERAÇÃO DE ONTOLOGIAS

Dissertação de mestrado desenvolvida no ano de 2006, no Programa de Pós-Graduação do Departamento de Informática do Centro Técnico e Científico da Pontifica Universidade Católica do Rio de Janeiro (DIAS, S. 2006).

Este trabalho apresenta uma possível solução por meio de métodos merge[34] para os problemas de construção de uma federação de ontologias (disponibilizar várias ontologias do mesmo domínio em um mesmo lugar). A aplicabilidade da proposta foi verificada pela implementação por meio da ferramenta *Protégé*, de um estudo de caso referente ao modelo de telefonia que trata da relação dos produtos fornecidos para os clientes (DIAS, S., 2006).

Os resultados estabeleceram que a integração dos dados, por meio da federalização das ontologias, permite usar apenas a representação dos termos no nível lógico do conhecimento independente da implementação dos mesmos. Observaram também que houve uma facilidade na manutenção do modelo desenvolvido, bem como a sua reutilização (DIAS, S., 2006).

[34] Integração de duas ou mais ontologias resultando em apenas uma (DIAS, S., 2006).

5.2 ARQUITETURA PARA UTILIZAÇÃO DE ONTOLOGIAS EM SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÃO

Dissertação de Mestrado desenvolvida em 2005, no Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina. Essa pesquisa descreve a construção de um módulo para integrar ontologias em um sistema de busca (GUÉRIOS, 2005).

O método proposto foi desenvolvido por meio da adição dos termos da ontologia relacionados ao vetor de busca. Os itens recuperados foram ordenados segundo a similaridade entre estes vetores de busca (GUÉRIOS, 2005).

A fim de analisar a viabilidade desta proposta a arquitetura desenvolvida foi empregada em um *site* de buscas sobre currículos de pesquisadores da Rede ScienTI[35]. Utilizou-se como exemplo a ontologia Descritores em Ciências da Saúde (DeCS[36]) usada para a indexação de artigos, livros e outros materias, e comparou-se com os resultados obtidos por um sistema tradicional de recuperação de informação (GUÉRIOS, 2005).

Dessa forma observou-se que a expansão da consulta por meio de ontologias incrementa o contexto semântico do vetor de consulta e, juntamente com o cálculo de similaridade vetorial[37], promove uma melhor classificação dos documentos. Assim recuperou-se um maior número de currículos (GUÉRIOS, 2005).

[35] Rede internacional de Fontes de Informação e Conhecimento para a Gestão da Ciência, Tecnologia e Inovação (www.scienti.net).
[36] Biblioteca virtual em saúde (http://decs.bvs.br/).
[37] Cálculo feito por meio da medida do co-seno sobre a distância entre cada vetor, e a freqüência em que ocorrem os termos consultados (GUÉRIOS, 2005).

5.3 ONTOLOGIA PARA A GESTÃO DO CONHECIMENTO EM SAÚDE POR MEIO DA METODOLOGIA METHONTOLOGY

Esta pesquisa foi desenvolvida em 2006 como Trabalho de Conclusão do Curso de Ciência da Computação da UNESC (FARIAS, 2006).

O trabalho consistiu na estruturação de uma base de conhecimento de dados médicos referente a Doenças Sexualmente Transmissíveis (DST), utilizando a metodologia *Methontology* para a construção de ontologias e a ferramenta *Protégé* para o desenvolvimento das mesmas (FARIAS, 2006).

A estruturação do domínio das DST por meio das ontologias facilitou sua compreensão, bem como representou uma forma de tratar o problema da falta de padronização na base de conhecimento. Observou-se também, que a estruturação por meio das ontologias proporcionou a realização de consultas mais objetivas (FARIAS, 2006).

5.4 DATA MINING PARA A CONSTRUÇÃO DE ONTOLOGIAS

Esta pesquisa foi desenvolvida no Programa de Engenharia de Sistemas e Computação, linha de pesquisa em Banco de Dados na Universidade Federal do Rio de Janeiro (BRAGA; JUNIOR, 2003).

Este trabalho consistiu no estudo da aplicação de técnicas de *data mining* no desenvolvimento de ontologias e implementação de uma ferramenta protótipo baseada na tarefa de classificação pelo algoritmo ID3 (BRAGA; JUNIOR, 2003).

O protótipo apresenta três módulos: engine, uma simples combinação de entrada de dados, um engenho de mineração de dados com um *paser* do modelo gerado

pelo engenho e um builder para a construção final da ontologia (BRAGA; JUNIOR, 2003).

A ferramenta apresenta as seguintes etapas para geração de ontologias: representação dos dados em linguagem XML, conversão para o formato do programa WEKA por meio de uma tarefa de *data mining*, aplicação do algoritmo no WEKA, exportar o resultado obtido para um formato de uma linguagem de descrição de ontologias, no caso desta pesquisa utilizou-se a RDF (BRAGA; JUNIOR, 2003).

Após as análises observou-se que o desenvolvimento manual de ontologias é uma tarefa difícil, principalmente quando os relacionamentos entre os atributos são complexos assim quando se utilizam algoritmos de *data mining* na construção de ontologias pode-se consideravelmente auxiliar os projetos de desenvolvimentos de ontologias (BRAGA; JUNIOR, 2003).

Observou-se ainda que a disponibilização pela ferramenta do modelo inferido usando uma linguagem de descrição de ontologias na forma de árvore de decisão ou na forma de sub-classes possibilitou que outras aplicações utilizassem o conhecimento mesmo sem darem suporte as tarefas de *data mining* (BRAGA; JUNIOR, 2003).

5.5 CONSTRUINDO ONTOLOGIAS, MAPAS PARA *DATA MINING* E DESCOBERTA DE CONHECIMENTO EM INFORMÁTICA BIOMÉDICA

Este artigo foi publicado no Simpósio Brasileiro de Matemática e Biologia Computacional em 2003 no Rio de Janeiro. Foi desenvolvido por Paulo Gottgtroy, Nik Kasabov e Stephen MacDonell do Instituto de Descoberta e Engenharia do Conhecimento da Universidade de Tecnologia Auckland, Nova Zelândia.

Este estudo teve como objeto de analisar como ontologias e *data mining* podem facilitar na análise de dados biomédicos, bem como a construção de uma ontologia para compartilhar conhecimentos de diferentes experiências empreendidas por pesquisas na área de ciências (GOTTGTROY; KASABOV; MACDONELL, 2003, tradução nossa).

Mediante isso, analisou-se a relação da descoberta de conhecimento em bancos de dados biomédicos por meio de ontologias na aplicação do *data mining* em na bases de dados por meio de uma ferramenta de mineração (GOTTGTROY; KASABOV; MACDONELL, 2003, tradução nossa).

A ontologia em dados biomédicos foi desenvolvida sobre o domínio Mapa de Infogene[38], evolvendo: genes, conceitos, entidades e processos de conhecimentos médicos, foi implementada na ferramenta Protégé. Os conhecimentos descobertos foram aplicados na ferramenta de *data mining* Neucom[39] (GOTTGTROY; KASABOV; MACDONELL, 2003, tradução nossa).

Nos resultados obtidos observaram-se grandes descobertas em dados biomédicos, na modelagem da base de dados por ontologias consegui-se extrair conhecimento útil e reutilizável, bem como facilitou a técnica de DATA MINING. O Mapa Infogene contribui para a descoberta biomédica, pois reuniu diferentes dimensões e diversas perspectivas para usuários como: investigadores, doutores, enfermeiras e farmacêuticos (GOTTGTROY; KASABOV; MACDONELL, 2003, tradução nossa).

Mediante as pesquisas apresentadas, a seguir tem-se uma tabela comparativa dos principais aspectos das pesquisas relatadas (Tabela 2). Na Tabela 3 tem-se a descrição dos trabalhos apresentados.

---

[38] Mapa que contém todos os genes que compõem a cadeia do DNA (GOTTGTROY; KASABOV; MACDONELL, 2003, tradução nossa).

[39] http://www.aut.ac.nz/neucom/NeuComStudent_Setup.exe

Tabela 2. Particularidades técnicas das pesquisas relatadas.

Pesquisas	Particularidades técnicas
Integração semântica de dados através de Federação de Ontologias	Métodos merge para a federação de ontologias. Implementada na ferramenta Protege. Estudo de caso referente ao modelo de Telefonia.
Arquitetura para utilização de Ontologias Em sistemas de recuperação de informação	Método proposto adicionar termos da ontologia relacionados ao vetor de busca. Arquitetura desenvolvida foi empregada em um site de buscas sobre currículos de pesquisadores da Rede ScienTI.
Ontologia para a Gestão do Conhecimento Em saúde por meio da metodologia Methontology	Estruturou-se por meio de ontologias a base de dados referente a Doenças Sexualmente Transmissíveis. Implementada na ferramenta Protege.
Data mining para a construção de Ontologias	Desenvolveu-se uma ferramenta protótipo baseada na tarefa de classificação pelo algoritmo ID3, para o desenvolvimento de ontologias. A ferramenta apresenta as seguintes etapas para geração de ontologias: representação dos dados em linguagem XML, conversão para o formato do programa WEKA por meio de uma tarefa de *data mining*, aplicação do algoritmo no WEKA, exportar o resultado obtido para um formato de uma linguagem de descrição de ontologias.
Construindo ontologias mapas para *Data mining* e DCBD	Construiu-se uma ontologia sobre o domínio Mapa de Infogene, implementada na ferramenta Protégé. Os conhecimentos descobertos foram aplicados na ferramenta de data mining Neucon.

Tabela 3. Descrições dos trabalhos correlatos

Pesquisas	Descrição
Integração semântica de dados através de Federação de Ontologias	Representação no nível do conhecimento Independente da aplicação Facilidade na manutenção e reutilização
Arquitetura para utilização de Ontologias Em sistemas de recuperação de informação	Consulta por meio de ontologias incrementa o contexto semântico Promoveu uma melhor classificação dos documentos Proporcionou um maior número de currículos recuperados
Ontologia para a Gestão do Conhecimento Em saúde por meio da metodologia Methontology	Facilitou a compreensão do domínio DST A estruturação da base de dados por meio de ontologias proporcionou consultas mais objetivas
Data mining para a construção de Ontologias	A utilização de algoritmos de *data mining* auxiliou na construção de ontologias Ontologia descrita por meio de uma linguagem na forma de árvore de decisão possibilitou a outras aplicações a utilização do conhecimento sem darem suporte a tarefas de *data mining*
Construindo ontologias mapas para *Data mining* e DCBD	Facilitou a técnica de *data mining* Proporcionou grandes descobertas biomédicas

Concluindo-se a apresentação de algumas pesquisas referente as ontologias e a técnica de *data mining* em conjunto com ontologias, tem-se no próximo capítulo o trabalho desenvolvido.

6 APLICAÇÃO DE ONTOLOGIAS E DATA MINING PARA A DESCOBERTA DO CONHECIMENTO

Esta pesquisa consistiu na análise dos benefícios proporcionados pela utilização de *data mining* e ontologias, para isso realizaram-se basicamente duas etapas: no primeiro momento desenvolveu-se a ontologia para estruturação e padronização da base de dados e aplicou-se a técnica de *data mining*, enquanto no segundo momento, realizou-se o *data mining* para a descoberta de conhecimento e se desenvolveu uma ontologia das relações identificadas.

Na execução da pesquisa empregaram-se ferramentas desenvolvidas em ambiente acadêmico e disponibilizadas gratuitamente. A construção das ontologias realizou-se com a Protégé, enquanto a aplicação do *data mining* ocorreu por meio da Shell Orion Data Mining Engine. Além da definição das ferramentas que foram empregadas também foi preciso escolher uma base de dados para ser utilizada no desenvolvimento do trabalho. Assim, optou-se por uma base de dados na área da saúde referente ao diagnóstico de doença coronariana disponível gratuitamente no repositório UC Irvine Machine Learning Repository no *site* http://www.ics.uci.edu/~mlearn/MLRepository.html. Neste repositório existem várias bases de dados usadas por estudantes e pesquisadores para análises dos mesmos por meio de algoritmos de aprendizado de máquina.

6.1 A BASE DE DADOS

A base de dados utilizada denomina-se *Heart Disease Database* e está disponível para *download* no *site*: http://archive.ics.uci.edu/ml/datasets/Heart+Disease.

Esta base tem como objetivo predizer problemas do coração considerando o diâmetro das suas artérias. Assim, se alguma das quatro artérias do coração apresenta diâmetro reduzido em 50% a pessoa é considerada portadora de doença coronariana. Os fatores julgados importantes vão desde informações objetivas como idade, sexo e hábitos de fumante até informações subjetivas, como descrição dos sintomas de dores nos diversos pacientes.

A *Heart Disease Database* apresenta dados originados de: Cleveland Clinic Foundation, Hungarian Institute of Cardiology Budapest, Veterans Affairs Medical Center (Long Beach, Califórnia) e University Hospital (Zurich). Nesta pesquisa utilizou-se a base de Cleveland, pois apresentou maior número de dados completos. Esta base é composta por 303 registros e 76 atributos, entretanto foram utilizados apenas 13, pois existiam muitos dados incompletos. Além disso, estes foram os atributos considerados de maior relevância para o diagnóstico de doença coronariana (DDC) pelo especialista em cardiologia.

Estas adequações foram realizadas de acordo com alguns livros e orientações do especialista do domínio de aplicação o professor e cardiologista do curso de Medicina da UNESC Miguel Moretti.

Assim, os atributos utilizados foram:

a) faixa etária: vinte a setenta anos;
b) sexo: masculino ou feminino;
c) dor precordial: dor ocasionada pelo fornecimento insuficiente de sangue ao coração (COWAN, 2000). Assumindo os valores de: típica para angina[40] (presença de dor característica de angina); atípica para angina

[40] Dor no peito de origem cardíaca (MICHIELIN, 2003).

(presença de dor não característica para angina) e assintomático (não apresenta dor anginosa);

d) pressão arterial sistólica em repouso: pressão no interior das artérias no momento em que o coração bombeia o sangue (NOBRE; SERRANO JÚNIOR, 2005). Classificada em: baixa ou hipotensão (menor 90mmHg[41]); ótima (entre 90mmHg a 120mmHg); normal (entre 120mmHg a 130mmHg); limítrofe (entre 130mmHg a 140mmHg) e alta (maior e igual 140mmHg);

e) colesterol total no soro sanguíneo: níveis de gordura no sangue composto pelos triglicérides e lipídios formados por: HDL e LDL (FARRET, 2005). Níveis de colesterol: desejável (abaixo de 200 mg/dl[42]); limítrofe (entre 200 mg/dl e 239 mg/dl) e aumentados (acima de 240 mg/dl);

f) glicemia em jejum > 120 mg/dl: quantidade de açúcar no soro sangüíneo mas precisamente no plasma (NOBRE; SERRANO JÚNIOR, 2005). Assumiu os valores de: verdadeiro e falso;

g) eletrocardiograma em repouso: exame que faz uma análise da freqüência, ritmo e eixo do coração por meio dos impulsos elétricos (HESS, 2003). Valores assumidos: zero (normal, ritmo sinusal[43], medidas de segmentos e intervalos dentro dos limites da normalidade); um (anormal, elevação ou infradesnível do ST-T[44] maior que 0,05 milivolts (mV)) e dois (probabilidade ou provável hipertrofia ventricular esquerda);

h) freqüência cardíaca máxima (fcmax) alcançada durante teste ergométrico: consiste em uma prova de esforço físico onde o paciente

[41] Unidade de medida de pressão chamada milímetro de mercúrio.
[42] Unidade de medida chamada miligramas por decilitro.
[43] Ritmo dos batimentos cardíacos considerados normais (AZEVEDO, 1999).
[44] Segmento observado no eletrocardiograma produzido pelas ondas elétricas do coração (AZEVEDO, 1999).

exercita-se na bicicleta ou esteira rolante até manifestar algum desconforto. Neste teste observa-se a freqüência cardíaca máxima alcançada durante o exercício (COWAN, 2000). O cálculo para saber quantos batimentos por minuto pode-se alcançar é realizado subtraindo-se 220 da idade. De acordo com essa fórmula tem-se as seguintes classificações (MICHIELIN, 2003):

- mínima: atingiu 60% da fcmax calculada (220 - Idade * 60%),
- atingida: atingiu 100% da fcmax calculada (220 - Idade);
- superada: atingiu mais do fcmax calculada;
- máxima: atingiu 85% da fcmax calculada (220 - Idade * 85%),
- baixa: atingiu 40% da fcmax calculada;

i) dor induzida pelo teste ergométrico (angina induzida pelo exercício): assume os valores sim ou não;

j) depressão do ST[45] induzido pelo teste ergométrico: assume os valores positivo (ST igual 1mm ou infradesnível de ST >= 1mm) e negativo (quando infradesnível de ST < 1mm ou quando for zero);

k) *slope*: refere-se a inclinação no pico do teste ergométrico, ou seja, do segmento ST no exercício para a fcmax estipulada para a idade. Pode assumir os valores: horizontal, para cima e para baixo;

l) cateterismo: exame para diagnosticar o grau de estreitamento das coronárias (número de vasos comprometidos) que pode assumir os valores zero, um, dois ou três (HESS, 2003);

m) diagnóstico: baseia-se no grau de estreitamento (estenose) da coronária observado no cateterismo cardíaco. Assume para o diagnóstico de doença

[45] Segmento observado no eletrocardiograma produzido pelas ondas elétricas do coração (AZEVEDO, 1999).

coronariana os valores de ausência (quando a coronária tem um estreitamento menor que 50%) e presença (quando a coronária tem um estreitamento maior que 50%).

6.2 METODOLOGIA

A metodologia aplicada para o desenvolvimento desta pesquisa compreendeu, inicialmente, a etapa de: levantamento bibliográfico que constituiu no estudo dos termos relacionados ao trabalho como *data mining*, ontologias, ferramentas e metodologias para a construção de ontologias, dentre outros.

Além desta etapa realizou-se também: desenvolvimento da ontologia para *data mining;* aplicação do *data mining* para ontologia e a análise referente a aplicabilidade de uma forma de organização e padronização do conhecimento no processo de DCBD.

6.2.1 Ontologia para Data Mining

Nesta seção tem-se a descrição da primeira etapa da pesquisa, apresentando-se o desenvolvimento de uma ontologia de aplicação para a base de dados de doença coronariana. Posterior a esta fase realizou-se a aplicação do *data mining* por meio da tarefa de classificação e do algoritmo ID3 para indução de árvores de decisão na *Shell* Orion Data Mining Engine.

6.2.1.1 Desenvolvimento da Ontologia Referente a Doença Coronariana

A estruturação da base de dados referente ao DDC, por meio de ontologia, foi desenvolvida utilizando-se da metodologia Methontology para a construção da mesma. A methontology foi aplicada realizando-se as seguintes etapas: especificação, aquisição do conhecimento, conceitualização, formalização, integração, implementação, avaliação, documentação e manutenção.

6.2.1.1.1 Especificação

Na especificação define-se o tipo de ontologia a ser utilizada e o domínio de conhecimento. Nesta pesquisa desenvolveu-se uma ontologia de aplicação tendo como domínio o DDC.

A ontologia apresenta definições, diagnósticos, tipos de exames: físicos, laboratoriais e complementares. Cada classe da ontologia contém dados, como uma subclasse, apresentando definições e informações. Por exemplo, considerando-se a classe exames físicos tem-se as subclasses: dor precordial e pressão alta que apresentam, respectivamente, as suas definições e informações.

6.2.1.1.2 Aquisição do Conhecimento

Esta etapa refere-se à aquisição do conhecimento sobre a área específica do domínio para a formação da base de conhecimento, para isso utilizaram-se livros da área

da saúde relacionados a Cardiologia46; informações presentes no repositório de onde retirou-se a base de dados (http://archive.ics.uci.edu/ml/datasets/Heart+Disease) e entrevistas desestruturadas realizadas com o especialista do domínio de aplicação o médico Cardiologista Miguel Moretti.

6.2.1.1.3 Conceitualização

A conceitualização compreendeu a estruturação do conhecimento a respeito do DDC por meio de um modelo conceitual[47]. Dessa forma, alguns problemas como desorganização e ambigüidades, presentes na base, foram resolvidos.

Esta etapa é uma das principais da metodologia Methontology, sendo responsável por descrever os termos do domínio identificado na atividade de especificação.

O modelo conceitual desenvolvido nesta fase utilizou-se dos seguintes componentes: glossários de termos, árvore de classificação de conceitos, dicionário de conceitos, tabela de atributos de instâncias e tabela de instância.

6.2.1.1.3.1 Glossário de Termos

Os termos do domínio da ontologia criada e as suas descrições são incluídos no glossário, sendo de suma importância, pois é nele que se descrevem todos os conceitos e relações da ontologia.

[46] Ramo da medicina que se ocupada do tratamento e diagnósticos de doenças que acometem o coração (NOBRE; SERRANO JÚNIOR, 2005).

[47] Descreve os problemas e soluções em termos do vocabulário para o domínio da ontologia identificado durante a fase de especificação (FERNÁNDEZ-LÓPEZ, GÓMEZ-PÉREZ, JURISTO, 1997, tradução nossa).

Na Tabela 4 tem-se um exemplo do glossário de termos desenvolvido para a ontologia DDC, expondo-se todas as conceituações utilizadas bem como as suas descrições. No entanto, todas as tabelas dos termos a serem descritos estão apresentadas no Apêndice C.

Tabela 4. Descrição dos conceitos utilizados na ontologia DDC

Conceitos	Descrição
Diagnóstico de doença coronariana	Primeiro e principal conceito da hierarquia da ontologia.
Conceitos de diagnóstico de doença coronariana	Classe contendo a conceitualização do termo diagnóstico de doença coronariana.
Diagnóstico	Classe contendo alguns diagnósticos de ausência e presença de doença coronariana.
Tipos de exames	Classe contendo alguns tipos de exames: físicos, laboratoriais e complementares necessários para o diagnóstico de doença coronariana.

Na Tabela 5 apresentam-se as relações existentes na ontologia DDC que descrevem a interação entre os seus conceitos.

Tabela 5. Descrição das relações utilizadas na ontologia DDC

Relação	Descrição
Descrição	Descrição sobre o conceito relacionado, neste caso diagnóstico de doença coronariana
Definição	Definição sobre a descrição do conceito relacionado
Sexo	Sexo a qual se refere à definição relacionada
Observação	Observação sobre o conceito relacionado
Fonte	Fonte das informações obtidas
Faixa etária	Faixa etária a qual se refere à definição relacionada

6.2.1.1.3.2 Árvore de Classificação de Conceitos

Nesta etapa todos os conceitos relacionados no glossário de termos são representados por meio de árvores, permitindo a visualização de todas as classes com suas respectivas subclasses. Na Figura 33 tem-se a hierarquia da ontologia DDC.

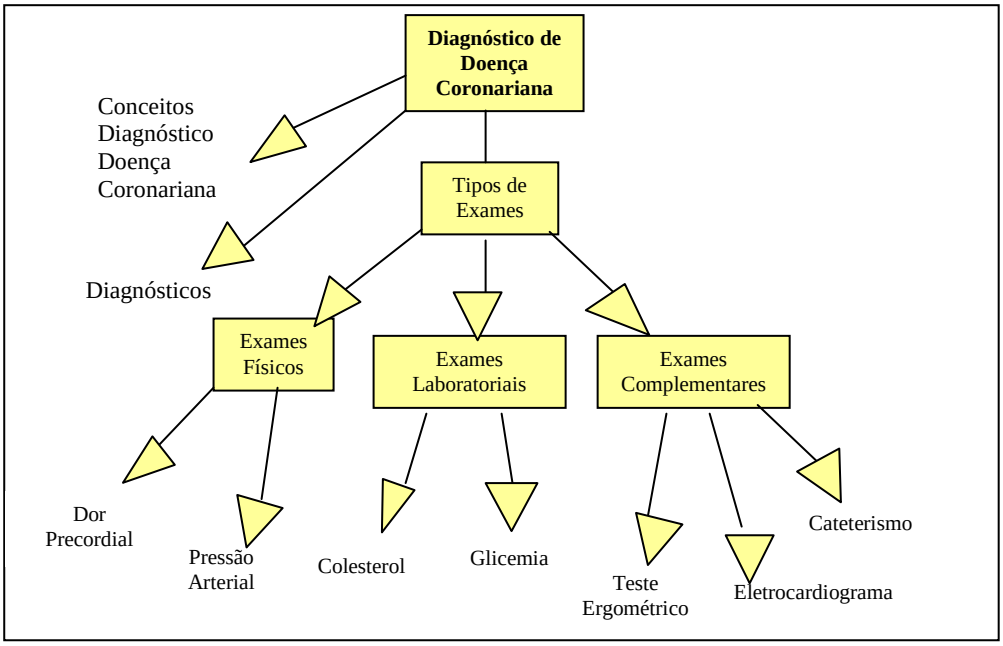


Figura 33. Hierarquia da ontologia diagnóstico de doença coronariana

Os conceitos Dor Precordial, Pressão Arterial, Colesterol, Glicemia, Teste Ergométrico, Eletrocardiograma e Cateterismo possuem subconceitos. Um exemplo é o conceito de dor precordial que tem os subconceitos: conceitos e características (Figura 34). As demais árvores de classificação dos outros conceitos estão apresentadas no Apêndice D.

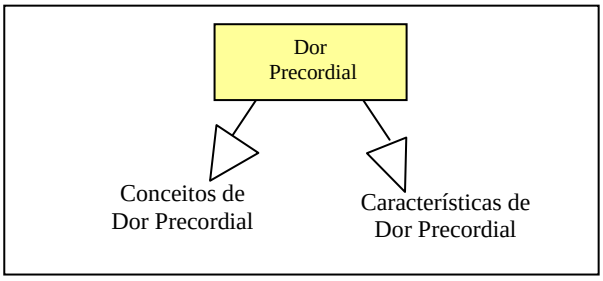


Figura 34. Árvore de classificação do conceito Dor Precordial

Finalizando-se a apresentação das árvores de classificação geradas observa-se a importância desta etapa na visualização gráfica da hierarquia completa da ontologia descrita no glossário de termos.

6.2.1.1.3.3 Dicionário de Conceitos

Descrevem todos os conceitos utilizados no domínio e apresentados nas árvores de classificação como: tipo, hierarquia e relação. O tipo do conceito pode ser classificado em: concreto, ou seja, pode ser instanciado e abstrato que aparece na hierarquia, mas não é instanciado.

Nesta fase a ontologia (DDC) tem seus conceitos detalhados por meio de tabelas. A Tabela 6 apresenta o detalhamento do conceito DDC, enquanto na Tabela 7 tem-se as relações deste conceito.

As diversas tabelas dos conceitos estão apresentadas no Apêndice E.

Tabela 6. Árvore de classificação de conceitos e subconceitos de DDC

Conceito	Tipo	Hierarquia
Diagnóstico de Doença Coronariana	Abstrato	Conceito Principal
Conceitos Diagnóstico de Doença Coronariana	Concreto	Subconceito de DDC
Diagnósticos	Concreto	Subconceito de DDC
Tipos de Exames	Abstrato	Subconceito de DDC

Tabela 7. Árvore de classificação das relações do conceito DDC

Conceito	Relação
Conceitos Diagnóstico de Doença Coronariana	Descrição Definição Fonte
Diagnósticos	Descrição Definição Sexo Faixa etária Fonte
Tipos de Exames e Todos os subconceitos de: Exames Físicos, Exames Laboratoriais Exames Complementares	Descrição Definição Observação Fonte

Logo, o dicionário de conceitos apresenta de forma detalhada a ontologia a ser desenvolvida facilitando o processo de construção da mesma.

6.2.1.1.3.4 Tabela de Atributos de Instância

Fornece informações sobre os atributos e os valores utilizados nas instâncias. Detalha as relações dos atributos descritos no dicionário de conceitos por meio de uma tabela de atributo de instâncias, apresentando-se nome do conceito, tipo (inteiro, caractere e símbolo), cardinalidade (quantidades de valores mínimos e máximos), entre outros. Na Tabela 8 têm-se as informações dos atributos das instâncias existentes na ontologia DDC.

Tabela 8. Atributos das instâncias da ontologia DDC

Conceito	Relação	Função
Conceitos Diagnóstico de Doença Coronariana	Descrição	Caractere
	Definição	Caractere
	Fonte	Caractere
Diagnósticos	Descrição	Caractere
	Definição	Caractere
	Faixa etária	Caractere
	Sexo	Símbolos, valores: feminino, masculino ou ambos
	Fonte	Caractere
Tipos de Exames, Exames Físicos, Exames Laboratoriais e Exames Complementares	Descrição	Caractere
	Definição	Caractere
	Observação	Caractere
	Fonte	Caractere

6.2.1.1.3.5 Tabelas de Instâncias

As tabelas de instâncias representam o conhecimento da ontologia desenvolvida possuindo: nome do conceito, relação e instância. A ontologia DDC apresenta várias instâncias (Apêndice F), porém, nas Tabelas 9 e 10 tem-se alguns exemplos.

Tabela 9. Instância do conceito de DDC – subconceito Conceitos de DDC

Conceito	**Conceitos Diagnósticos Doença Coronariana**
Relação	**Instância**
Descrição	Diagnóstico de Doença Coronariana
Definição	O processo de investigação é o método pelo qual o médico pode adicionar ou descartar uma hipótese e prescrever os exames apropriados para o diagnóstico de uma doença coronariana.
Fonte	HESS, Michael L. **Doenças cardíacas: primeiros cuidados.** São Paulo: Monole, 2002.

Tabela 10. Instância do conceito de DDC – subconceito Diagnósticos

Conceito	**Diagnósticos**
Relação	**Instância**
Descrição	Diagnóstico de Ausência de Doença Coronariana
Definição	Em 61,31 % registros foi constatado ausência de doença coronariana.
Faixa etária	Vinte a Setenta anos.
Sexo	Masculino
Fonte	*Heart Disease Database* (http://archive.ics.uci.edu/ml/datasets/Heart+Disease)

Na tabela de instâncias são apresentadas as informações adquiridas na fase de aquisição do conhecimento, por meio de livros, pesquisas e entrevistas com o especialista da área. Esta fase é responsável pela formação da base de conhecimento da ontologia.

6.2.1.1.4 Formalização

Esta etapa da metodologia objetiva a formalização do modelo conceitual por meio de uma linguagem formal. Há ferramentas de desenvolvimento de ontologias, como por exemplo a Protégé, que automaticamente implementam o modelo conceitual

em diferentes linguagens, como: RDF[48], XML[49], entre outras. Porém, de acordo com FERNÁNDEZ-LÓPEZ, GÓMEZ-PÉREZ e JURISTO (1997) a formalização não é uma fase obrigatória na atividade de construção de ontologias.

No caso desta pesquisa a formalização não foi realizada, pois o objetivo não era a criação de uma ontologia por meio de uma linguagem, mas sim a estruturação na forma de uma base de dados.

6.2.1.1.5 Integração

Esta fase visa a integração da ontologia desenvolvida com definições de outras existentes, ou seja, objetiva o reuso de contextualizações do mesmo domínio a fim de adicionar informações. Porém, considerando-se que não se teve conhecimento de algum domínio de DDC e também por ser uma ontologia de aplicação esta etapa não foi necessária.

6.2.1.1.6 Implementação

Nesta fase realizou-se a implementação da ontologia por meio da ferramenta Protégé 3.1**.**

O desenvolvimento de uma nova ontologia inicia-se pela escolha do tipo de projeto desejado, na Protégé tem-se: *Protégé Database* voltada para banco de dados; *Experimental XML File* produz arquivos em XML; OWL *Database* consiste em uma

[48] Linguagem de descrição de ontologias para a representação do conhecimento da web semântica (GAŠEVIĆ; DJURIĆ; DEVEDŽIĆ, 2006, tradução nossa).
[49] Linguagem para marcação de documentos na web (GAŠEVIĆ; DJURIĆ; DEVEDŽIĆ, 2006, tradução nossa).

linguagem de web semântica; OWL Files que gera arquivos na linguagem OWL; RDF *Files* produz um arquivo na linguagem para descrição de ontologias RDF.

Nesta pesquisa utilizou-se o *Protégé Files* para a construção da ontologia, como pode-se visualizar na Figura 35.

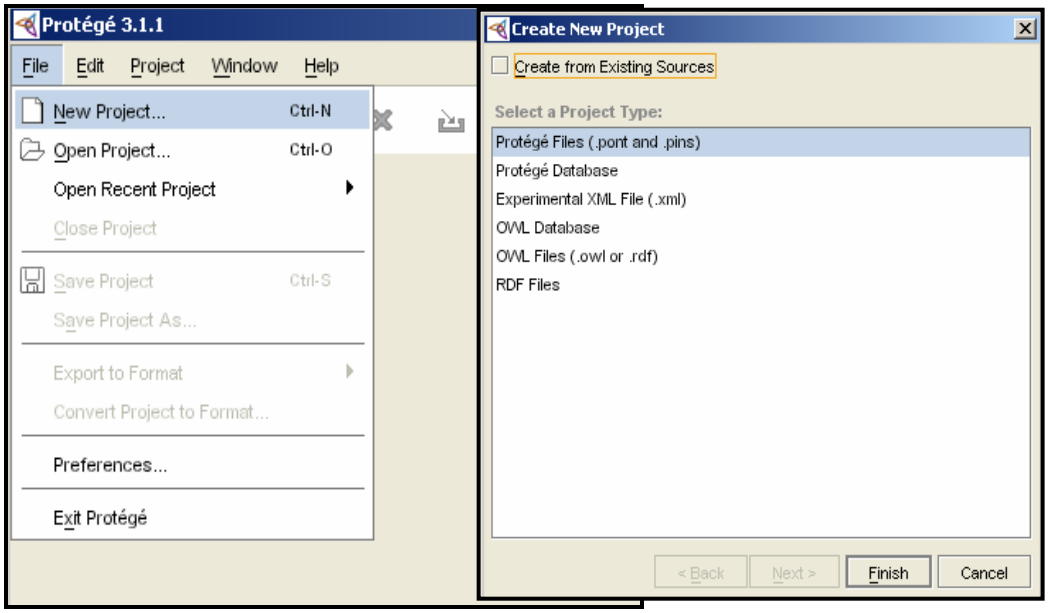


Figura 35. Construção de um novo projeto na ferramenta Protégé.

A seguir realizou-se a inserção dos conceitos e subconceitos presentes na ontologia DDC definidos na etapa de conceitualização (Figura 36).

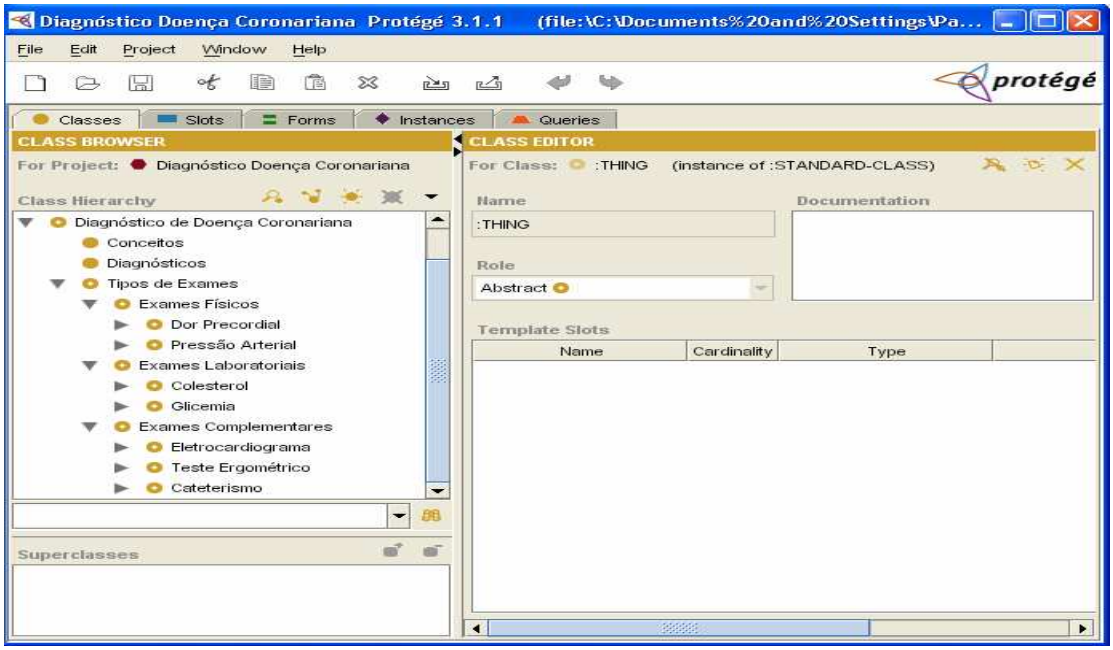


Figura 36. Criação dos conceitos e subconceitos da ontologia

Posteriormente, foram criadas as relações que estão especificadas no glossário de termos que consta de: tipo, valor, cardinalidade e domínio. Na Protégé as relações são definidas por *slots*, tendo-se na Figura 37 um exemplo da criação do *slot* definição.

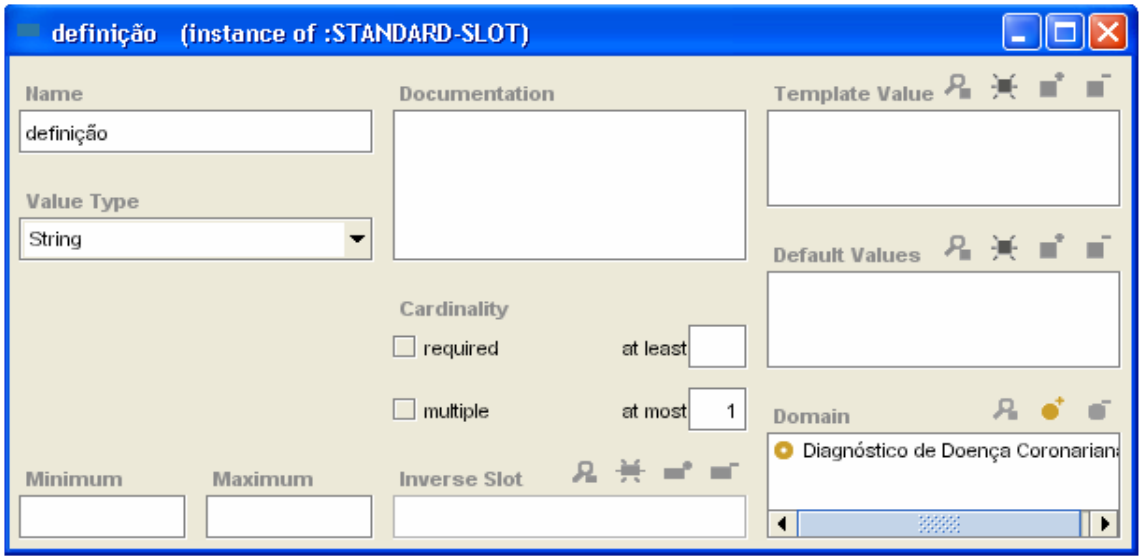


Figura 37. Criação da relação ou *slot* definição

Esse processo foi repetido para todos os conceitos da ontologia DDC de acordo com o dicionário de conceitos e as características apresentadas na tabelas de atributos de instâncias. Salienta-se que os conceitos e subconceitos da ontologia herdam os *slots* criados na classe principal, isso se deve ao fato de possuir a herança múltipla.

Considerando-se, por exemplo, a relação *definição* que foi criada na classe DDC esta poderá ser utilizada pelos demais conceitos da ontologia. No entanto, existem relações particulares que são criadas apenas para um determinado conceito, como por exemplo: o *slot* faixa etária pertencente apenas ao conceito diagnósticos.

Finalizadas estas etapas realizou-se a inserção de conhecimento na ontologia criada, ou seja, a inclusão de instâncias. Na Figura 38 têm-se um exemplo de instância no conceito dor precordial.

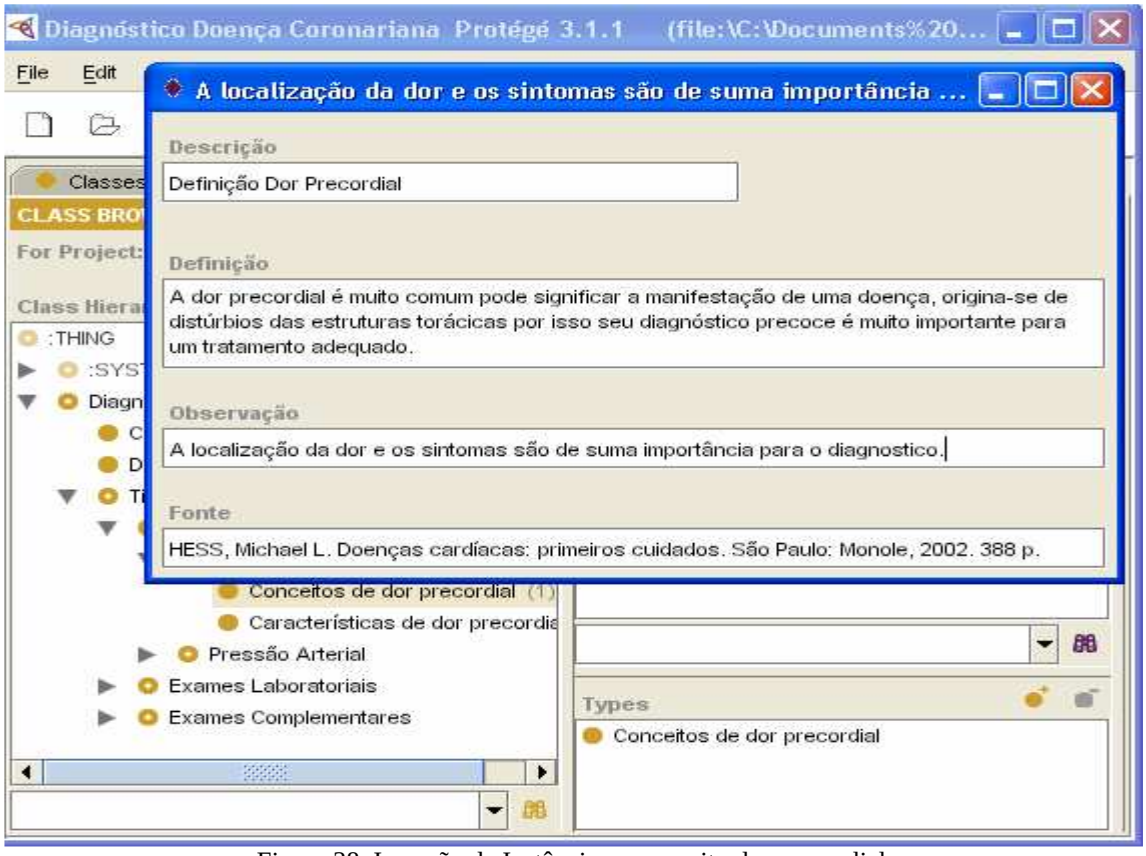


Figura 38. Inserção da Instância no conceito dor precordial

Concluída a inclusão das instâncias nos vários conceitos da ontologia DDC, tem-se esta totalmente implementada. A Figura 39 ilustra a sua apresentação final.

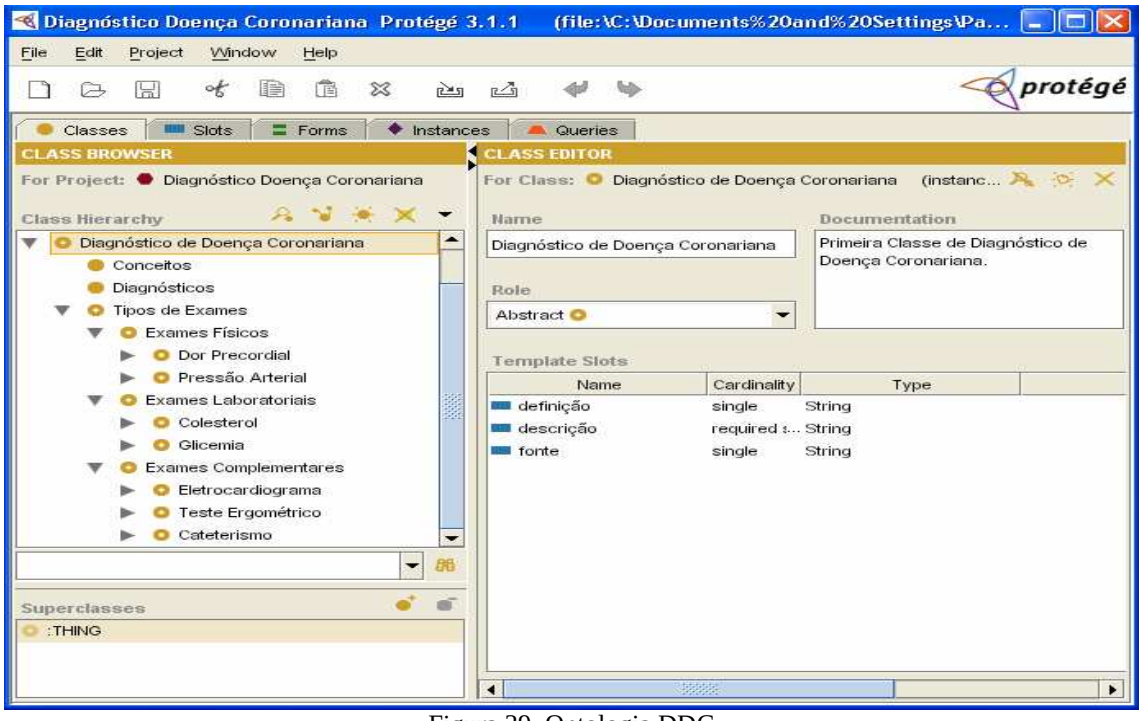


Figura 39. Ontologia DDC

6.2.1.1.7 Avaliação

A avaliação indica a análise da ontologia desenvolvida por meio dos benefícios obtidos, bem como a validação da consistência da ontologia na representação do conhecimento do domínio identificado. Assim, a avaliação é utilizada para verificar por meio de métricas de ontologias uma medida quantitativa da sua evolução.

No caso desta pesquisa não se aplicou uma metodologia específica para avaliação, pois se tinha a limitação de tempo e a ausência de dados para a realização de testes comparativos. Porém, quando se finalizou a construção da ontologia realizou-se uma verificação junto ao especialista na área de cardiologia, Miguel Moretti, onde se observou que os conhecimentos representados estavam corretos e confirmavam as informações da área.

6.2.1.1.8 Documentação

A documentação da ontologia se faz necessária para informar o processo de construção da mesma para sua posterior utilização, apresentando todas as informações geradas. A documentação referente a ontologia DDC está disponível no desenvolvimento da pesquisa, bem como nos Apêndices.

6.2.1.1.9 Manutenção

Esta fase de manutenção ficou impossibilitada de ser realizada, visto que se tinha pouco tempo para realização da pesquisa, pois para isso depende do conhecimento do especialista nesta área. Concluindo-se a estruturação da base de dados por meio da

ontologia tem-se na próxima seção a etapa de realização da técnica de *data mining* na base estruturada.

6.2.1.2 Aplicação do *Data Mining*

Na realização da etapa de *data mining* tem-se a necessidade da fase de pré-processamento onde os dados são preparados e organizados. Neste caso, o conhecimento foi previamente estruturado pela ontologia. Mediante isso, desenvolveu-se tabelas, a partir da ontologia criada anteriormente, para posterior desenvolvimento no sistema gerenciador de banco de dados PostgreSQL 8.2[50].

As tabelas construídas estão apresentadas no Apêndice G e são constituídas pelo nome do campo (especificado como chave primária) e tipo onde consta a definição do valor assumido. Estes foram definidos por *character varying*, pois o algoritmo ID3, apenas processa valores nominais. Exemplos são dados nas Tabelas 11 e 12, onde se tem faixa etária e sexo.

Tabela 11. Faixa etária

Campo	Tipo de campo
Faixa_etaria	Character varying (200)
Descrição	Character varying (200)

Tabela 12. Sexo

Campo	Tipo de campo
Descricao_sexo	Character varying (200)
Descrição	Character varying (200)

Todas as tabelas foram assim desenvolvidas, logo duas tabelas distintas foram criadas separando-se os casos de ausência com 210 registros (Tabela 13) e presença com 93 registros de doença coronariana (Tabela 14).

[50] Banco de dados disponível para download em http://www.postgresql.org/.

Tabela 13. Ausência de doença coronariana

Campo	Tipo de campo
Saudável	Character varying(200)
Faixa_etaria	Character varying(200)
Sexo	Character varying(200)
Dor_precordial	Character varying(200)
Pressão_arterial_sistolica_repouso	Character varying(200)
Colesterol	Character varying(200)
Glicemia_jejum	Character varying(200)
Ecg_repouso	Character varying(200)
Fcmax	Character varying(200)
Dor_induzida	Character varying(200)
DepressaoST_induzida	Character varying(200)
SlopeST	Character varying(200)
Cateterismo	Character varying(200)

Tabela 14. Presença de doença coronariana

Campo	Tipo de campo
Doente	Character varying(200)
Faixa_etaria	Character varying(200)
Sexo	Character varying(200)
Dor_precordial	Character varying(200)
Pressão_arterial_sistolica_repouso	Character varying(200)
Colesterol	Character varying(200)
Glicemia_jejum	Character varying(200)
Ecg_repouso	Character varying(200)
Fcmax	Character varying(200)
Dor_induzida	Character varying(200)
DepressaoST_induzida	Character varying(200)
SlopeST	Character varying(200)
Cateterismo	Character varying(200)

A construção da tabelas baseou-se nos diagramas de entidades relacionamentos. Definidas as tabelas o próximo passo foi criá-las com comandos SQL e (Figura 40) e carregá-las no PostgreSql 8.2.

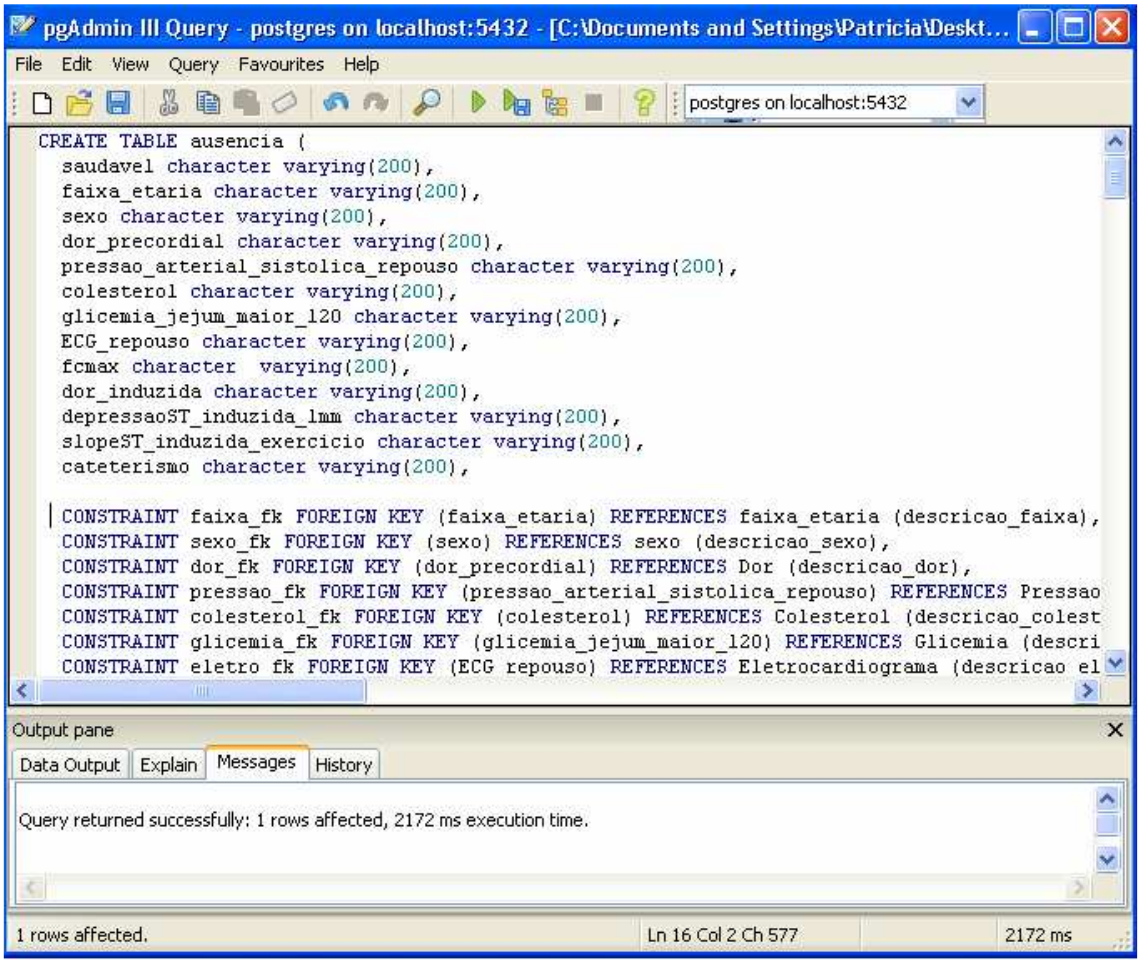


Figura 40. Criação da tabela ausência

Construída a base de dados realizou-se a etapa de *data mining* aplicando-se a tarefa de classificação pelo algoritmo ID3 na Shell Orion. Na execução do *data mining* realizou-se a geração das regras referente aos casos de ausência e presença de doença coronariana.

Na Figura 41 tem-se as regras obtidas na tabela de ausência, com o atributo de saída cateterismo e nível de profundidade da árvore 7, ou seja, o algoritmo seleciona atributos que possuam o maior ganho de informação, ou seja, identifica aquele atributo que faz a diferença na classificação dos dados para a saída escolhida.

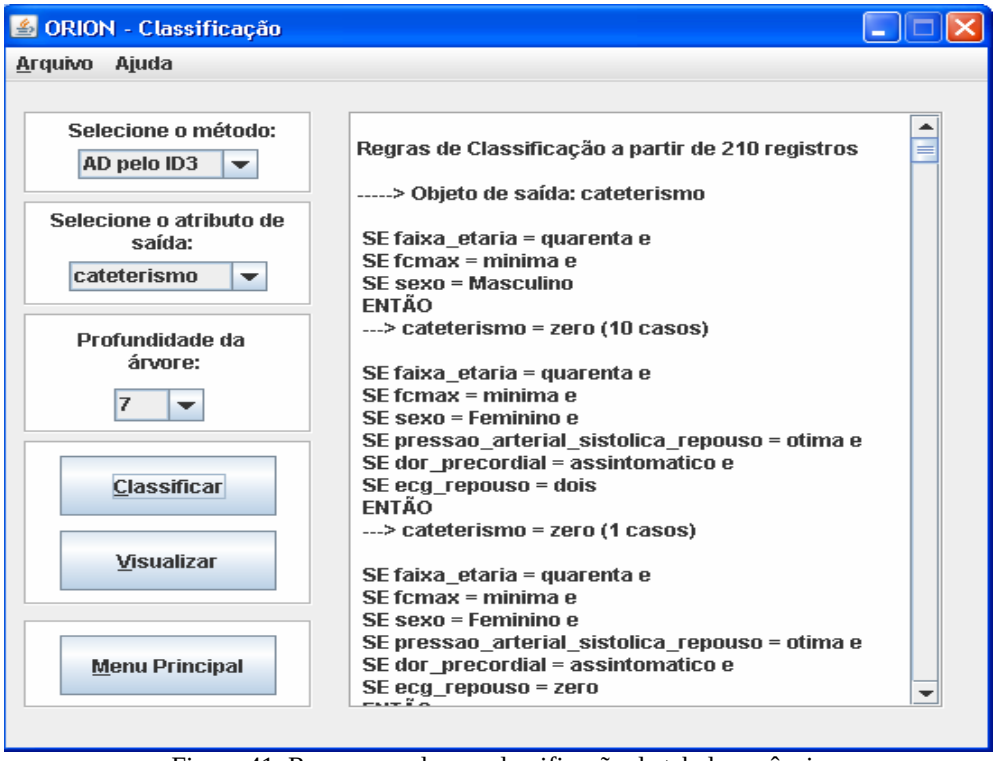


Figura 41. Regras geradas na classificação da tabela ausência

Conforme se observa na Figura acima, foram analisados todos os atributos existentes na base de dados, gerando-se 115 regras a partir de 210 registros existentes. Percebeu-se que as informações relevantes no caso de cateterismo foram: faixa etária, freqüência cardíaca máxima alcançada no teste ergométrico, pressão arterial sistólica em repouso, dor precordial, dentre outros. Na Figura 42 tem-se a árvore de decisão gerada pelo algoritmo ID3 na tarefa de classificação.

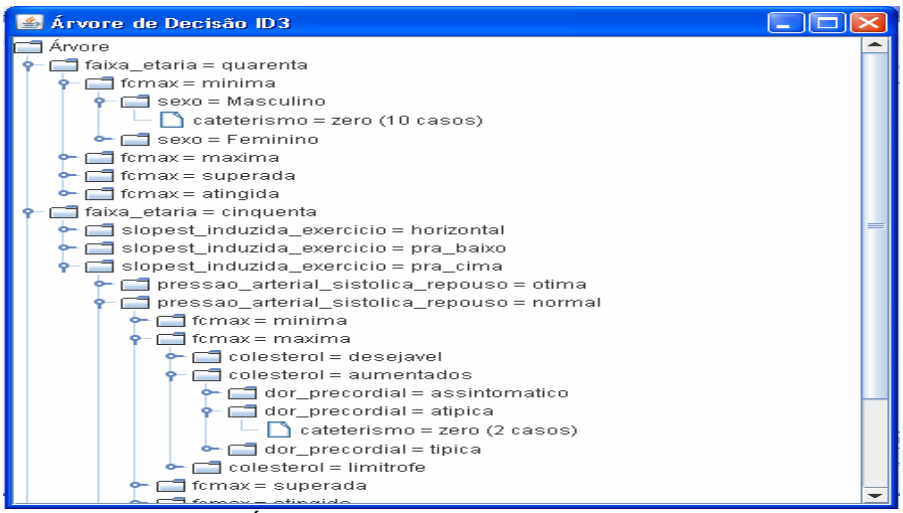


Figura 42. Árvore gerada na classificação da tabela ausência

Realizando a mesma análise na tabela presença (Figura 43), obteve-se 82 regras de 93 registros, onde o exame cateterismo teve seus valores de 1 a 3, significando o número de artérias comprometidas. Nas regras geradas pode-se perceber que a faixa etária varia entre cinqüenta à setenta anos, com maior incidência em pacientes na faixa etária dos sessenta anos.

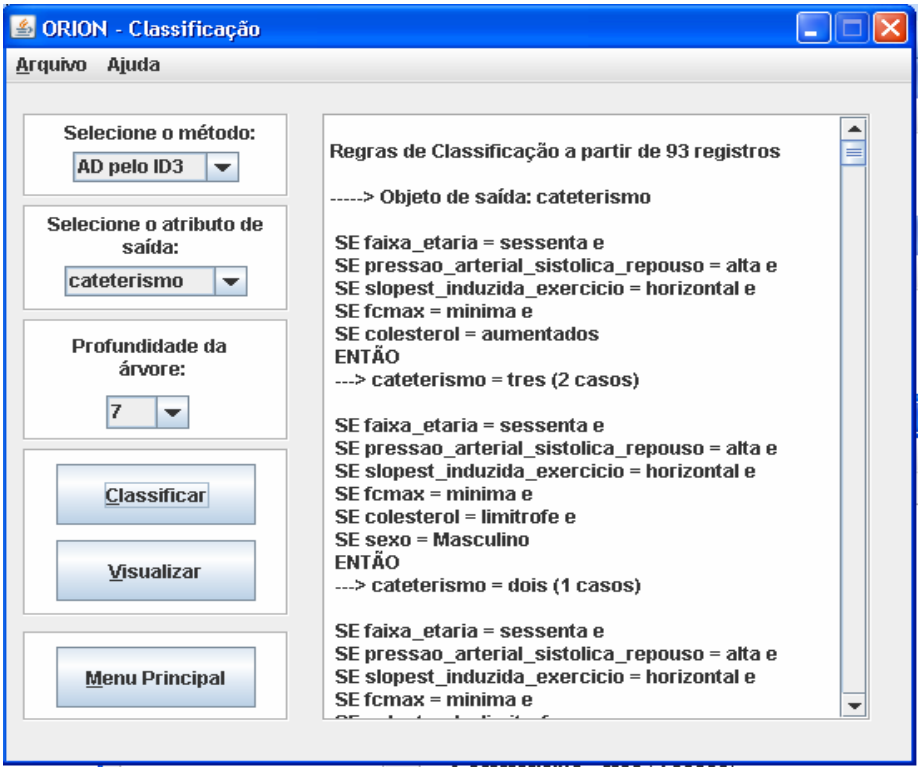


Figura 43. Regras geradas na tabela presença

6.2.2 Data mining para Ontologia

Nesta seção apresenta-se a segunda fase desta pesquisa que compreendeu a realização do *data mining* na base de dados referente a doença coronariana, identificando-se relações e padrões que foram utilizados para a construção de uma ontologia na etapa de pós-processamento do processo de DCBD.

6.2.2.1 Aplicação do *Data Mining*

A base de dados de doença coronariana foi estruturada no sistema gerenciador de banco de dados PostgreSqL 8.2 por meio do desenvolvimento de uma única tabela. Neste momento realizaram-se as atividades de pré-processamento que consistem na seleção, limpeza, normalização, exclusão dos dados ruidosos e transformação dos dados de numéricos para nominais, pois era uma das necessidades do algoritmo ID3.

Na Tabela 15 tem-se os campos criados, todos foram definidos como caractere (*character varying (200*)).

Tabela 15. Campos criados na tabela coração

Campo
Faixa_etaria
Sexo
Dor_precordial
Pressao_arterial_sistolica_repouso
Colesterol_total
Glicemia_jejum_maior_120
Eletrocardiograma_repouso
Frequencia_cardiaca_maxima
Dor_induzida_teste_ergometrico
DepressaoST_induzido_teste_ergometrico
SlopeST_pico_teste_ergometrico
Cateterismo_vasos_comprometidos
Diagnostico

A tabela denominada de coração consta de 303 casos e foi implementada no por meio de comandos *sql* e executada no gerenciador PostgreSqL 8.2, conforme mostra a Figura 44.

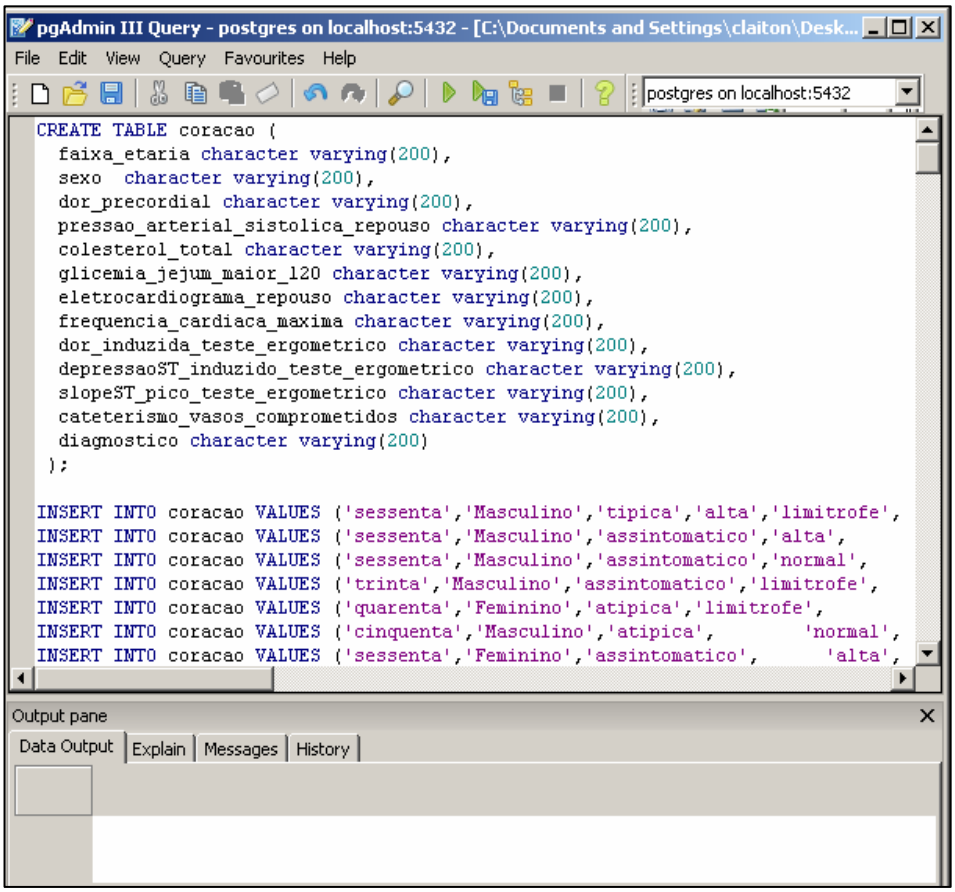


Figura 44. Criação da tabela coração no Postgresql

A seguir realizou-se a etapa de *data mining* pelo algoritmo ID3 na Shell Orion Data Mining Engine. As regras obtidas informando como atributo de saída cateterismo e profundidade da árvore 7 podem ser visualizadas na Figura 45, enquanto tem-se na Figura 46 a árvore de decisão construída.

As regras geradas mostram que para os casos de diagnóstico de presença de doença coronariana há comprometimento de alguma artéria. Enquanto isso, nos casos de ausência observa-se que muitos têm fatores de risco e estão propensos a terem problemas de coração, no entanto são considerados saudáveis, pois no exame de cateterismo não apresentam artérias comprometidas.

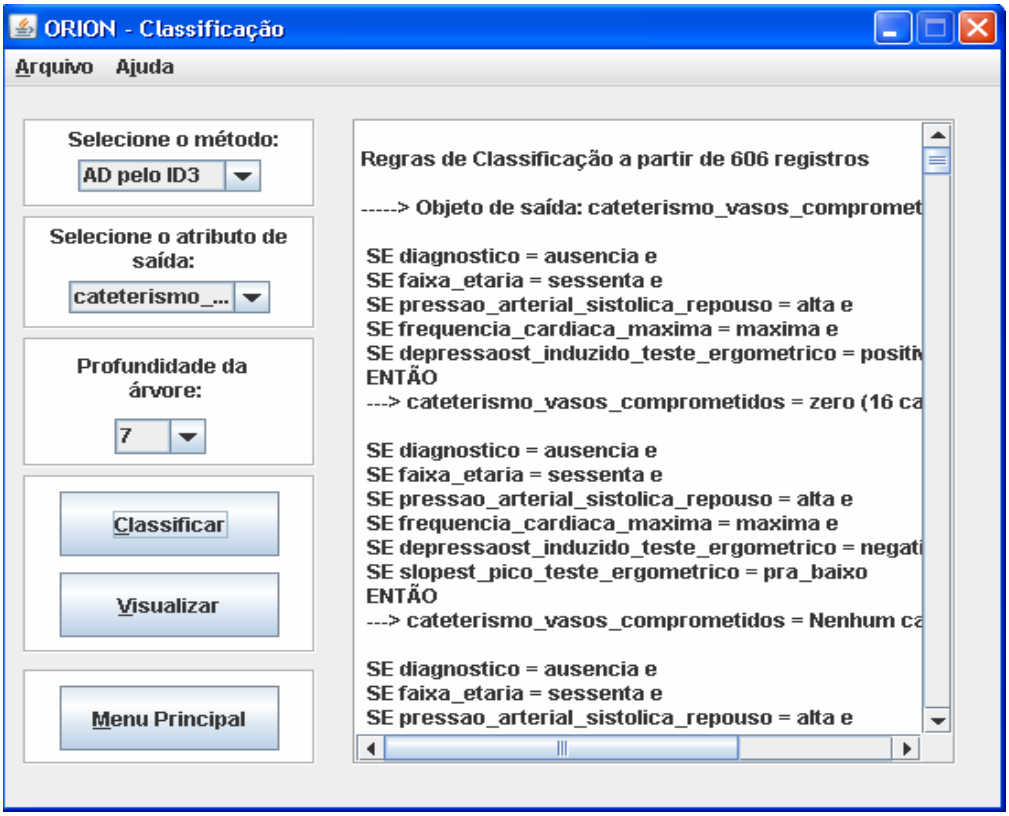


Figura 45. Regras geradas na tabela coração

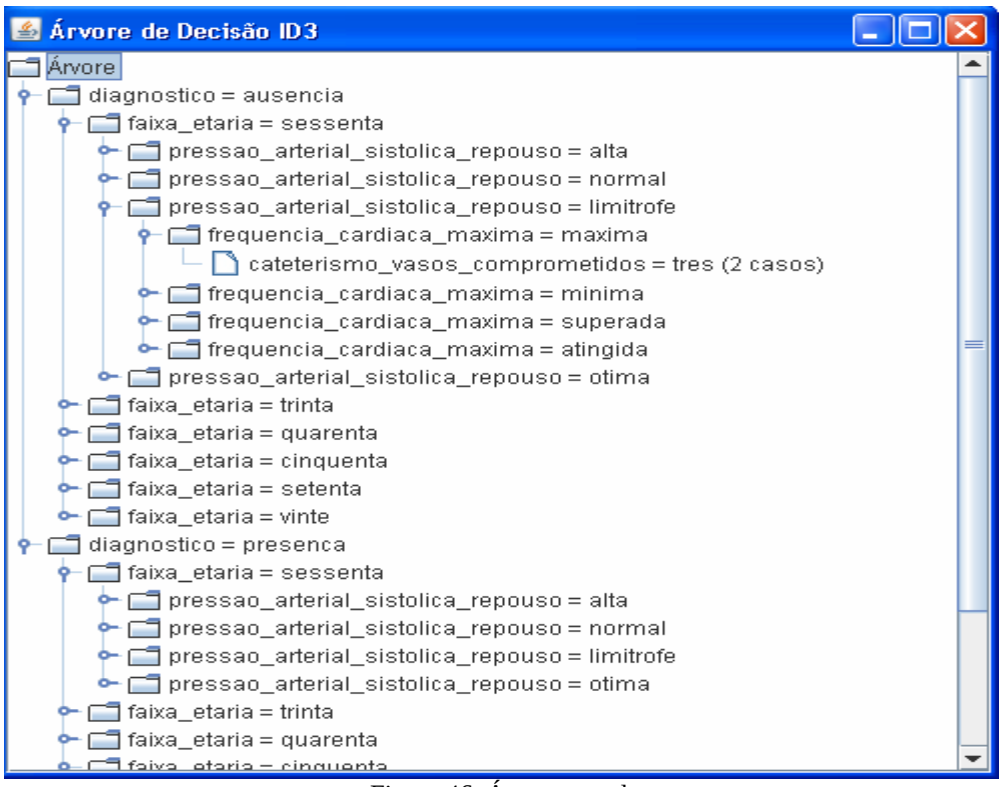


Figura 46. Árvore gerada

Aplicado o *data mining* na base de dados de doença coronariana, a próxima etapa da pesquisa compreendeu o pós-processamento que foi realizado por meio da construção de uma ontologia.

6.2.2.2 Construção da Ontologia

As relações descobertas durante o *data mining* foram representadas por meio de uma ontologia, utilizando-se a metodologia Methontology já mencionada na seção 3.2.1.1 desta pesquisa, seguindo-se as etapas de: especificação, aquisição do conhecimento, conceitualização, formalização, integração, implementação, avaliação, documentação e manutenção.

As etapas de especificação e aquisição do conhecimento, onde o tipo de ontologia e o domínio são definidos, continuaram com as mesmas especificações anteriores no que se refere a diagnóstico de doença coronariana, porém as classes foram organizadas de forma diferente, considerando-se a análise, pelo especialista do domínio de aplicação, das regras identificadas no *data mining*.

Devido a base de dados ter sido estruturada por uma única tabela, gerou-se várias regras na *Shell* Orion Data Mining Engine dificultando o entendimento dos conhecimentos descobertos.

Assim, retirou-se a classe exames para só exames e acrescentou-se a classe sintomas deixando esse diagnóstico mais claro. Mediante isso pode-se observar uma melhor forma de representar as relações descobertas, referente ao diagnóstico de doença coronariana.

Mediante isso a ontologia apresentou como classes definições, diagnósticos, sintomas (dor precordial) e exames (glicemia, colesterol, pressão arterial,

eletrocardiograma, teste ergométrico e cateterismo). Cada classe conteve as suas caracterizações na forma de subclasses denominadas de definições e informações. Exemplo: classe dor precordial subclasses conceitos e características.

Assim, no decorrer desta pesquisa serão apresentadas somente aquelas etapas da metodologia Methontology que tiveram alguma alteração em relação a primeira fase do trabalho (6.2.1.1).

6.2.2.2.1 Conceitualização

Na conceitualização os componentes gerados: glossários de termos, árvore de classificação de conceitos, dicionário de conceitos, tabela de atributos de instâncias e tabelas de instâncias, em sua maioria apresentam informações distintas aos componentes da fase anterior (6.2.1.1.3), pois a hierarquia da ontologia foi estruturada de maneira diferente.

6.2.2.2.1.1 Glossário de Termos

Neste componente foram descritos todos os termos da ontologia, conceitos (Apêndice H) e relações (Tabela 17). Na Tabela 18 apresenta-se um exemplo dos conceitos da ontologia DDC.

Tabela 16. Descrição dos conceitos da ontologia

Conceitos	Descrição
Sintomas	Classe contendo sintomas para diagnóstico de doença coronariana
Dor Precordial	Classe contendo conceitos e características de dor precordial
Conceitos de Dor Precordial	Definição sobre o conceito de dor precordial
Características de Dor Precordial	Apresenta algumas características de dor precordial

Tabela 17. Descrição das relações da ontologia

Relação	Descrição
Descrição	Descrição sobre a informação do conceito relacionado, neste caso diagnóstico de doença coronariana
Definição	Definição sobre o conceito relacionado
Observação	Observação sobre o conceito relacionado
Sexo	Sexo a qual se refere à definição relacionada
Fonte	Fonte das informações obtidas
Tipo	Tipo do conceito relacionado
Valores	Valores dos conceitos relacionados

6.2.2.2.1.2 Árvore de Classificação de Conceitos

Na árvore de classificação de conceitos aconteceram mudanças na hierarquia da ontologia, conforme se pode observar na Figura 47, pois de acordo com a análise dos resultados gerados pelo *data mining* constatou-se que esta nova hierarquia proporciona uma forma mais objetiva para a visualização do diagnóstico de doença coronariana.

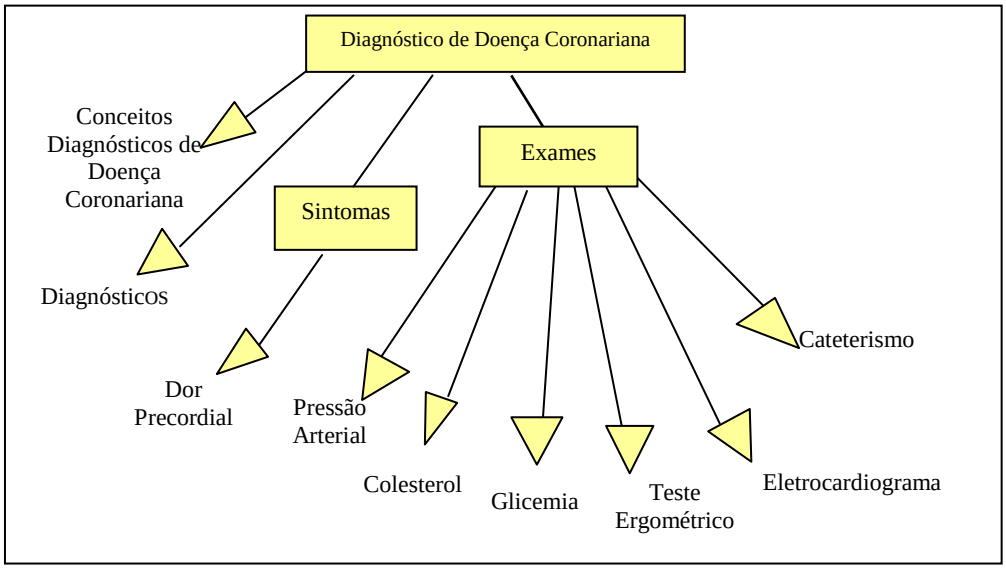


Figura 47. Hierarquia da ontologia diagnóstico de doença coronariana

Na Figura 48 tem-se uma árvore de classificação do conceito de cateterismo. As demais árvores estão apresentadas no Apêndice D e permanecem iguais àquelas geradas na primeira etapa deste estudo.

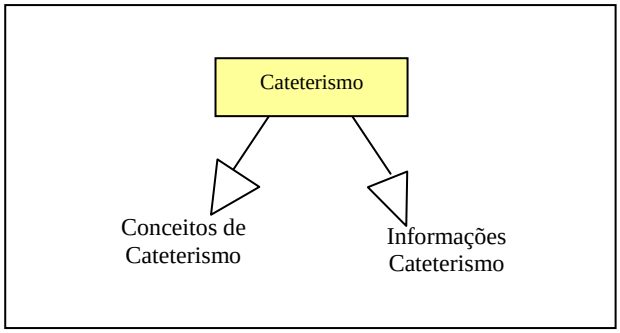


Figura 48. Hierarquia do conceito cateterismo

6.2.2.2.1.3 Dicionário de Conceitos

Nesta etapa os conceitos da ontologia, tipos e relações são descritos de acordo com a árvore de classificação (Apêndice I). Na Tabela 19 tem-se um exemplo do dicionário de conceitos e na Tabela 20 das relações.

Tabela 18. Árvore de classificação conceitos - subconceitos de DDC

Conceito	Tipo	Hierarquia
Diagnóstico de Doença Coronariana	Abstrato	Conceito Principal
Conceitos Diagnósticos de Doença Coronariana	Concreto	Subconceito de DDC
Diagnósticos	Concreto	Subconceito de DDC
Sintomas	Abstrato	Subconceito de DDC
Exames	Abstrato	Subconceito de DDC

Tabela 19. Árvore de classificação conceitos – relações de DCC

Conceito	Relação
Conceitos Diagnósticos de Doença Coronariana	Descrição Definição Fonte
Diagnósticos	Descrição Definição Sexo Fonte Faixa etária
Eletrocardiograma Teste Ergométrico Cateterismo	Descrição Definição Fonte Observação
Dor Precordial Pressão Arterial	Descrição Definição Observação Tipo Fonte
Glicemia Colesterol	Descrição Definição Observação Valores Fonte

6.2.2.2.1.4 Tabela de Atributos de Instância

Nesta tabela estão descritos os atributos das instâncias, detalhando as relações e seus valores: tipo, cardinalidade, entre outros (Apêndice J). Um exemplo é apresentado na Tabela 21.

Tabela 20. Atributos de instâncias utilizados na ontologia

Conceito	Relação	Função
Conceitos Diagnósticos de Doença Coronariana	Descrição	Caractere
	Definição	Caractere
	Fonte	Caractere
Dor Precordial	Descrição	Caractere
	Definição	Caractere
	Tipo	Caractere
	Observação	Caractere
	Fonte	Caractere

6.2.2.2.1.5 Tabelas de Instâncias

Os conhecimentos inseridos nas tabelas de instâncias durante a primeira fase do trabalho (6.2.1.1.3.5) mantiveram-se inalterados, porém sofreram alterações no que se refere as relações (*slots*).

Nas Tabelas 22 e 23 têm-se alguns exemplos de instâncias, considerando-se as novas relações, enquanto no Apêndice K pode-se visualizar na íntegra as instâncias dos demais conceitos da ontologia.

Tabela 21. Instância do conceito do dor precordial – subconceito características de dor precordial

Conceito	**Características de dor precordial**
Relação	**Instância**
Descrição	Classificação Dor Precordial
Definição	Pode ser causada por: desconforto ou dor retroesternal, desencadeada pelo exercício, ou estresse emocional;
Tipo	Angina Típica
Observação	Pode ser aliviada com repouso
Fonte	NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** Barueri, SP: Manole, 2005.

Tabela 22. Instância do conceito dor precordial – subconceito características de dor precordial

Conceito	**Características de dor precordial**
Relação	**Instância**
Descrição	Classificação Dor Precordial
Definição	A angina manifestar-se por uma sensação de opressão, a dor não é intensa, contudo causa desconforto.
Tipo	Angina Típica
Observação	A localização da dor também é um fator importante para o diagnóstico.
Fonte	COWAN, J. Campbell. **Cardiologia.** 6. ed. São Paulo: Santos, 2000.

6.2.2.2.2 Implementação

O processo de implementação da ontologia deu-se de forma bastante semelhante com a apresentada no item 6.2.1.1.6, porém a construção da sua estrutura ficou diferente em função das alterações sofridas na fase de contextualização.

Nas Figuras 49 e 50, por exemplo pode-se observar, respectivamente, a implementação das novas classes e subclasses, bem como das relações (*slots*).

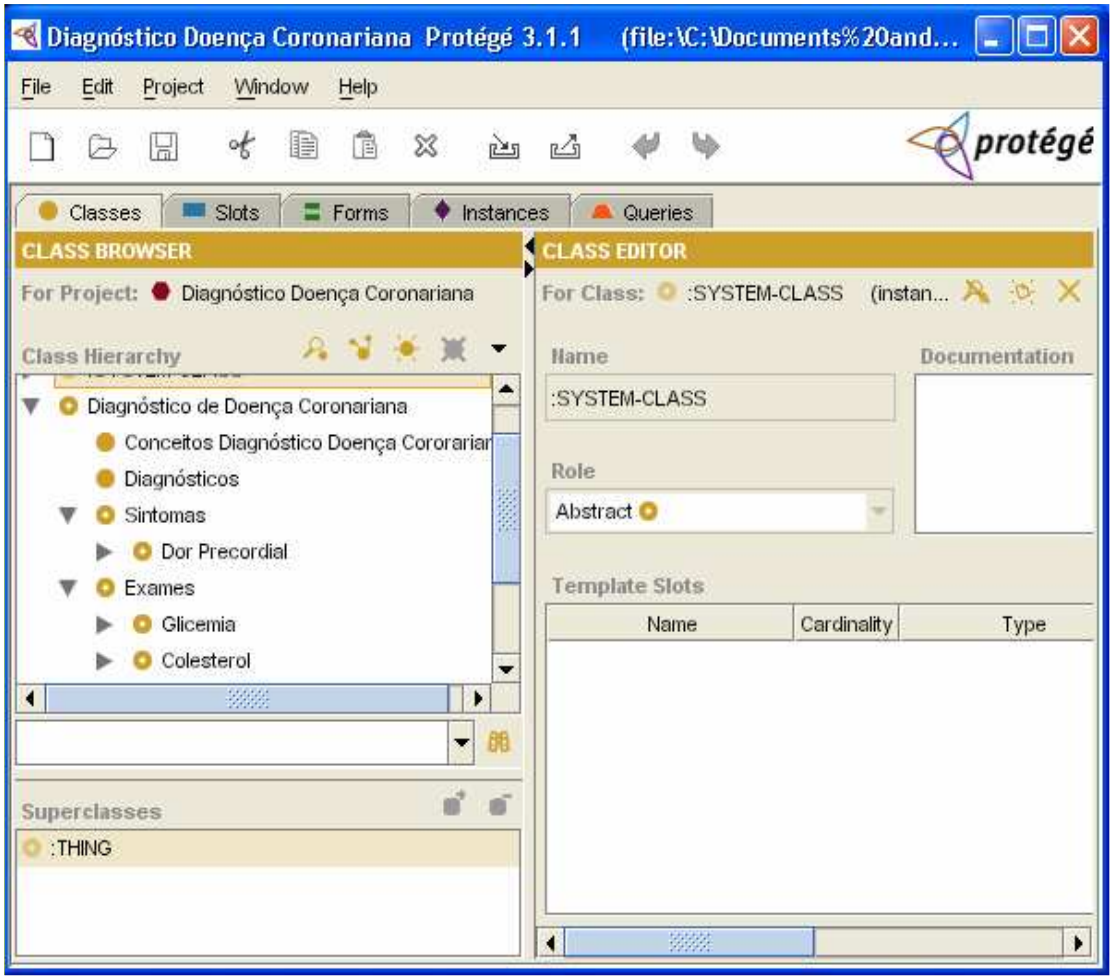


Figura 49. Criação dos conceitos e subconceitos

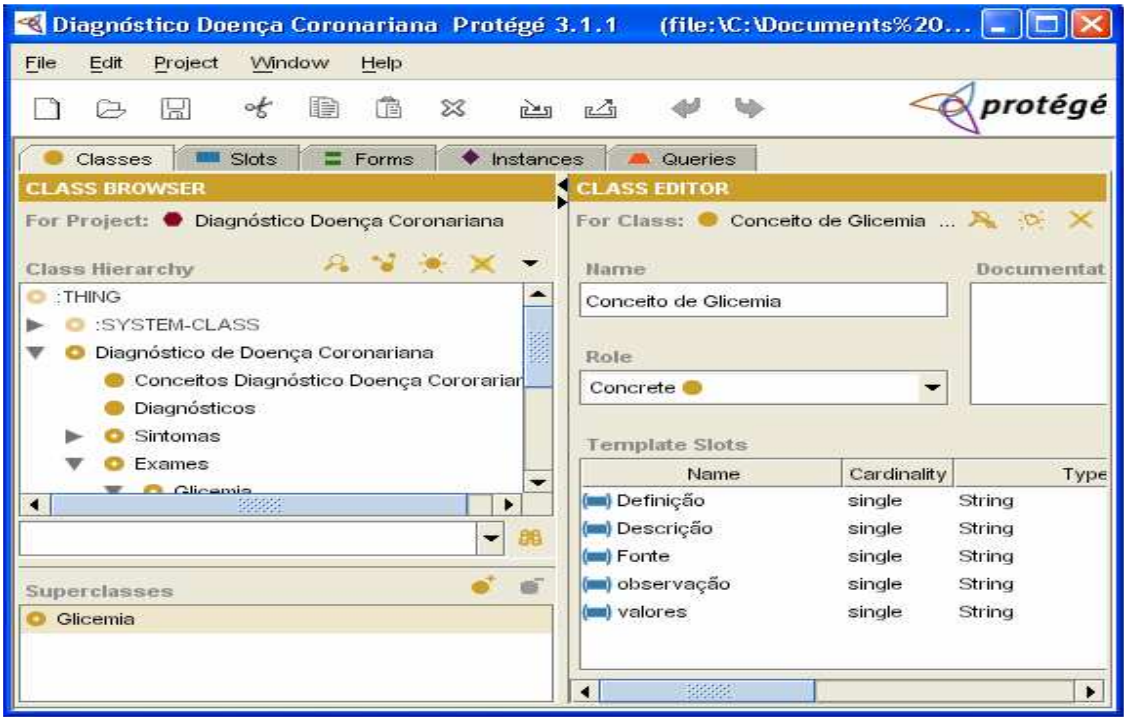


Figura 50. Criação dos *slots* do conceito glicemia

Finalmente, na Figura 51, tem-se a ontologia DDC completamente desenvolvida, considerando-se as relações entre o conhecimento identificadas na etapa de *data mining*, o que auxiliou na sua construção.

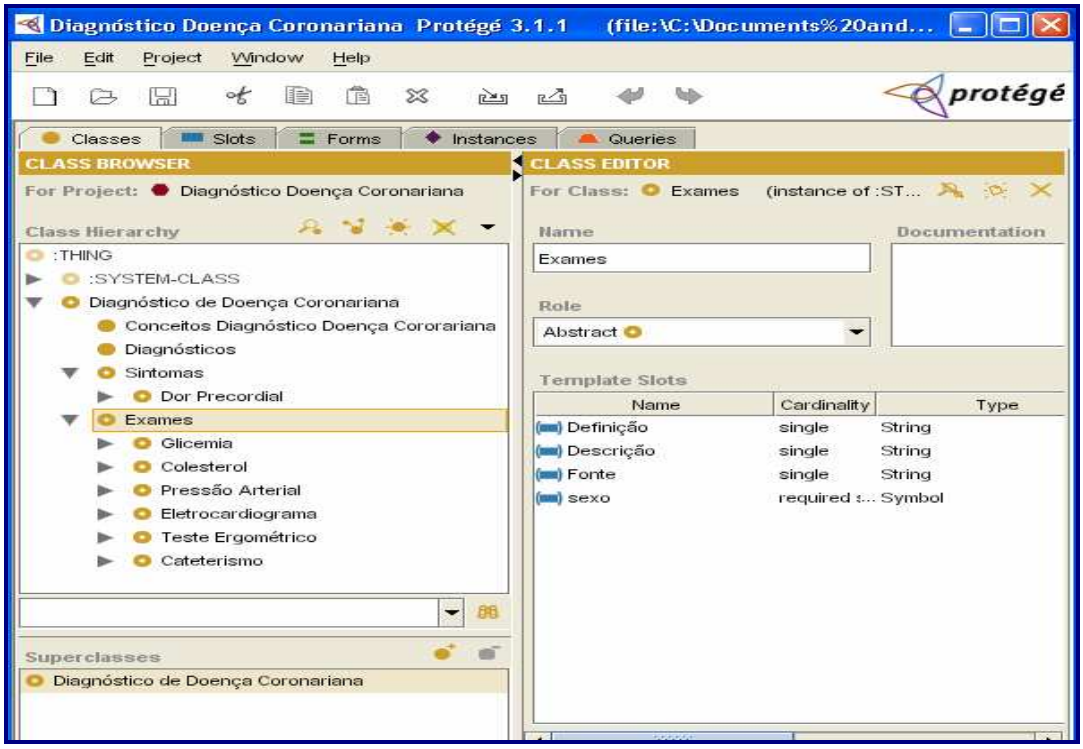


Figura 51. Ontologia Diagnóstico de Doença Coronariana

6.3 RESULTADOS OBTIDOS

Os resultados obtidos por esta pesquisa consistiram na análise da aplicação de ontologias anterior e posteriormente ao processo de DCBD para a descoberta de conhecimento em uma base de dados, neste caso, de doença coronariana. Realizando-se também uma interpretação, pelo especialista do domínio de aplicação, das relações e padrões encontrados.

Nesta interpretação dos resultados constatou-se a confirmação dos conhecimentos médicos na área, não trazendo nenhuma nova descoberta. Na base de dados foram identificados padrões para um diagnóstico de pessoas que estão propensas a terem doença coronariana ou não, são os chamados saudáveis ou doentes. Entretanto, várias pessoas possuem todos os fatores de risco, porém apesar disso são consideradas saudáveis. Isso ocorre devido ao fato do exame de cateterismo ser primordial para o fechamento do diagnóstico.

No cateterismo observa-se o percentual de comprometimento das artérias, assim define-se quantas estão comprometidas. Se o paciente apresentar uma ou mais artérias com estenose ele é considerado doente, caso contrário é considerado saudável (zero).

Na primeira etapa desta pesquisa, de ontologias para *data mining*, observou-se que a estruturação da base na fase de pré-processamento permitiu uma melhor preparação dos dados para a aplicação da tarefa de classificação. Isso se deve a minimização das inconsistências presentes nos dados. Além disso, proporcionou um melhor entendimento do diagnóstico da doença coronariana já que identificou a importância da relação do exame de cateterismo para o diagnóstico de ausência ou presença deste tipo de doença.

Ainda nesta etapa, o especialista apontou que com a aplicação do *data mining* após a ontologia teve-se uma forma simplificada para o entendimento das relações encontradas no diagnóstico de doença coronariana. Assim, confirmou-se a afirmação de Cízaro, Nigro e Xodó (2007, tradução nossa), de que quando se incorpora o conhecimento no processo de DCBD pelo uso de ontologias, anteriormente a etapa de *data mining* tem-se uma interpretação e validação do conhecimento no pré-processamento.

Na segunda etapa da pesquisa, de *data mining* para ontologias, o especialista do domínio de aplicação apontou a dificuldade no entendimento dos resultados, por exemplo, pelas pessoas leigas, visto que havia muitas informações na base e resultavam em mais regras do que na análise realizada na primeira fase. Considerando-se isso, percebeu-se que a geração das ontologias no pós-processamento resultou em uma melhor forma de visualização do conhecimento. Desta forma, confirma-se o que a bibliografia da área aborda, pois de acordo Brezany et al (2004, tradução nossa) com a aplicação de *data mining* para ontologias proporcionam benefícios na representação e análise dos resultados em áreas como da Saúde, no caso, por exemplo, de aplicações médicas que possuam um domínio bem específico.

Concluindo-se este estudo comprovou que no processo de DCBD para a descoberta de conhecimento a aplicação da técnica de *data mining* e o uso de ontologias no pós-processamento demonstram uma maior eficiência na interpretação dos conhecimentos gerados, de acordo com Cannataro et al (2003, tradução nossa), a validação por meio de ontologias dos conhecimentos descobertos no *data mining* proporcionam uma forma melhor na compreensão do domínio da aplicação.

Mediante os resultados obtidos, tem-se uma comparação (Tabela 24) entre a primeira etapa que consistiu em: construção da ontologia na base de dados diagnóstico

de doença coronariana e aplicação do *data mining*; estruturação pelo do *data mining* e a construção da ontologia nos conhecimentos descobertos.

As ontologias diferiram apenas em sua hierarquia e em algumas relações inseridas na segunda etapa. Enquanto que na primeira fase deste estudo tinha-se classe tipos de exames: físicos, laboratoriais e complementares, na segunda optou-se por uma classe chamada exames não tenho nenhuma divisão.

Tabela 23. Quadro com as principais alterações das ontologias

Etapas	**Fases**	**Principais alterações das ontologias em relação ao *data mining***
Ontologia para *Data Mining*	Construção da Ontologia	Nesta etapa a ontologia permitiu uma melhor visualização no desenvolvimento das tabelas para a etapa de *data mining*, pois estruturou o conhecimento sobre o diagnóstico de doença coronariana.
	Aplicação do *Data Mining*	Procurou-se construir uma tabela para cada conceito da ontologia utilizando-se do diagrama de entidade relacionamento.
Data Mining para Ontologia	Aplicação do *Data Mining*	Na estruturação da base de dados pela etapa de *data mining* construiu-se uma única tabela.
	Construção da Ontologia	Nesta etapa realizou-se por análise das regras geradas onde o especialista relatou o fato de as informações observadas estarem de difícil compreensão. Mediante isso a ontologia construída foi desenvolvida de uma forma geral, contendo a classe exames, sintomas e diagnósticos.

CONCLUSÃO

O processo de descoberta de conhecimento passou de uma simples tarefa de busca de informações úteis em grandes bases de dados para uma importante questão discutida entre os especialistas, pois é de importante encontrar metodologias que proporcionem uma maneira eficiente na realização deste processo.

Nesta pesquisa foram utilizadas ontologias no pré-processamento e pós-processamento do processo de DCBD para estruturar a base de dados referente ao diagnóstico de doença coronariana no intuito de analisar os benefícios.

No desenvolvimento apresentou-se o processo de DCBD, etapas, técnicas, *data mining* e a ferramenta para sua realização (*Shell* Orion Data Mining Engine).

Esta pesquisa demonstrou a conceitualização de ontologias, tipos, componentes, metodologias e a ferramenta para a sua construção (Protégé), realizou a aplicação de uma ontologia anterior e posterior a técnica de *data mining.*

No desenvolvimento da ontologia DDC foi utilizada a metodologia Methontology que por meio de suas etapas possibilita o desenvolvimento da ontologia de forma detalhada.

Apesar da conclusão deste estudo foram encontradas algumas dificuldades ao desenvolvê-lo. Ao realizar a revisão bibliográfica observaram-se poucos livros nacionais e internacionais sobre ontologias com conteúdo mais detalhado e principalmente em conjunto com a técnica de *data mining*. Na parte prática encontrou-se dificuldades na compreensão da forma para a aplicação de ontologias no processo de DCBD.

O desenvolvimento deste estudo foi realizado por meio de duas etapas distintas, primeiramente teve-se ontologias para *data mining,* que consistiu na

estruturação da base de dados referente ao diagnóstico de doença coronariana pela ontologia e aplicação do *data mining*.

No segundo momento a etapa de *data mining* para ontologias onde realizou-a tarefa de classificação pelo algoritmo ID3 *Shell Orion Data Mining Engine* na base de dados, e a construção da ontologia de aplicação DDC na ferramenta Protégé dos conhecimentos descobertos, com isso atingiu-se os objetivos desta pesquisa em relação a compreensão da aplicação de ontologias no processo de DCBD para a descoberta de conhecimento.

Após a conclusão desta pesquisa, pode-se afirmar que das duas formas a aplicação de ontologias traz benefícios no DCBD, a estruturação da base sobre o diagnóstico de doença coronariana pela ontologia na fase de pré-processamento contribui para resolver problemas como falta de padronização dos dados, bem como facilitou a etapa de *data mining* em relação a criação das tabelas no sistema gerenciador de banco de dado PostgreSQL 8.1. Na fase de pós-processamento a aplicação da ontologia permitiu a representação do diagnóstico de doença coronariana de uma forma objetiva.

Os conhecimentos gerados a respeito do diagnóstico de doença coronariana conforme o Médico especialista foram confirmados. Observou-se que muitos pacientes têm vários fatores de risco, no entanto são consideradas saudáveis isso mediante ao fato do exame de cateterismo ser primordial para o diagnóstico final deste tipo de doença. Além disso, concluiu-se que a ontologia DDC permitiu uma forma melhor de compreensão de diagnóstico de doença coronariana.

A partir desta pesquisa pode-se dar continuidade a utilização das ontologias no processo de DCBD por meio de algumas sugestões de trabalhos futuros:

a) empregar a abordagem da pesquisa em outras bases de dados a fim de analisar e poder retirar outras conclusões;

b) aplicar outras metodologias para a construção de ontologias;

c) utilizar outras:

d) tarefas e métodos de *data mining;*

e) ferramentas na construção de ontologias.

f) utilizar métricas para a avaliação de ontologias;

g) implementar mecanismos para consulta das ontologias criadas.

REFERÊNCIAS

ALMEIDA, M.B.; BAX, M. P. **Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção**. Ciência da Informação, v.32, n.3, p.7-20, set./dez. 2003. Disponível em:<http://ibict.br/cienciada-informacao/viewarticle.php?id=36&layout=abstract>. Acesso em: 10 de fev. 2007.

AZEVEDO, Decio Faraco de. **Iniciação à eletrocardiografia.** Porto Alegre: Artmed, 1999. 197 p.

BARRETO, Jorge Muniz. **Inteligência artificial no limiar do século XXXI.** 3.ed Florianópolis: Duplic, 2001. 392 p.

BECHHOFER, S.; et.al.**OilEd: a reasonable ontology editor for the Semantic** Web. In COFERENCE ON ARTIFICIAL INTELLIGENCE. Spring – Verlag, Berlin, Germany. 2001. Disponível em:<http://citeseer.ist.psu.edu/bechhofer01oiled.html>. Acesso em 10 de mar. 2007.

BERNSTEIN, A., Provost, F., & Hill, S. (2005). Towards intelligent assistance for the data mining process: An ontology-based approach for cost/sensitive classification. In IEEE Transactions on Knowledge and Data Engineering, 17(4), 503-518

BITTENCOURT, Guilherme. **Inteligência artificial**: ferramentas e teorias. 2.ed Florianópolis: UFSC, 2001.

BORTOLOTTO, Leandro Sehnem. **O método de redes neurais pelo algoritmo de Kohonen para clusterização** na **Shell de Data Mining Orion Engine.** Trabalho de Conclusão de Curso – Ciência da Computação, Universidade do Extremo Sul Catarinense, Criciúma, Santa Catarina. 2007.

BOVO, Alessandro Botelho. **Um Método de Tradução de Fontes de Informação em um Formato padrão que viabilize A Extração de Conhecimento por meio de Link Analysis e Teoria dos Grafos**. 2004. 102 f. Dissertação (Mestrado em Engenharia de Produção)- Programa de Pós-Graduação em Engenharia da Produção, Universidade Federal de Santa Catarina. Florianópolis. 2004. Disponível em <http://teses.eps.ufsc.br/Resumo.asp?5798>. Acesso em 29 mai. 2007.

BRAGA**,** Bruno da Rocha; JR, José D´Almeida Nogueira. **Data Mining para Construção de Ontologias**. 2003. Universidade Federal do Rio de Janeiro. Programa de Engenharia de Sistemas e Computação. Linha de Pesquisa em Banco de Dados. Disponível em: <http://www.cos.ufrj.br/~brunorb/docs/dmonto.pdf >. Acessado em: 20 de abril. 2007.

BREZANY, P., Janciak, I., Woehrer, A., & Tjoa, A.M. *GridMiner: **A framework for knowledge discovery on the Grid from a vision to design and implementation*.** Cracow Grid Workshop. Cracow,Poland: Springer. 2004.

BRIDEWELL, W., LANGLEY, P.**An Interactive environment for the modeling on discovery of scientific knowledge.** *International Journal of Human-Computer Studies*, 64, 1009-1014.2006.

BUENO, Tânia Cristina D'Agostini . **Engenharia da Mente: Uma Metodologia de Representação**. **Do Conhecimento Para a Construção de Ontologias Em Sistemas Baseados em Conhecimento**. 2005. 174 f. Tese (Doutorado em Engenharia)- Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina. Florianópolis. 2005. Disponível em < http://teses.eps.ufsc.br/Resumo.asp?6164>. Acesso em 1 mar. 2007

CÂMERA, G.; Egenhofer, M.; Fonseca, F.; Monteiro, A. M. V.**What's In An Image?**, in Conference on Spatial Information Theory, Santa Barbara, 2001.

CANNATARO, M., & Comito, C. (2003). ***A data mining ontology for Grid programming*.** Paper apresentado no I Workshop Sobre Semântica E Computação de Grid. Budapeste. Disponível em: <Http://www.isi.edu/~stefan/SemPGRID> Acessado em 03 set. 2008.

CARVALHO, Luís Alfredo Vidal de. **Datamining:** A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração. 2. ed. São Paulo: Érica Ltda, 2002.

CARVESAN, Fábio Lopes; ANDRADE, Marco Túlio Carvalho de. **Exploração de Relações Isomórficas Entre Técnicas Simbólicas e Conexionistas da Inteligência Computacional**. 2005.8f. Disponível em: < http://www.dcc.ufla.br/infocomp/artigos/v4.1/art03.pdf>. Acessado em: 07 nov. 2008

CASAGRANDE, Diego Paz. **O Módulo da Tarefa de Associação pelo Algoritmo Apriori no Desenvolvimento da Shell de Data Mining Orion**. 2005. 79 f. Trabalho de Conclusão de Curso – Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense, Criciúma, Santa Catarina, 2005.

CASTOLDI, André Vinícius**. Uma ontologia para Enlaces de Unidades de Informação em Plataformas de Governo Eletrônico.** 2003. Dissertação (Mestrado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina**.** Florianópolis. 2003. Disponível em: <http://teses.eps.ufsc.br/defesa/pdf/6847.pdf>. Acesso em 10 fev. 2008.

CASSETARI JUNIOR, José Marcio. **Ontologia para gestão do conhecimento em saúde por meio da metodologia methondology.** Trabalho de Conclusão de Curso – Ciência da Computação, Universidade do Extremo Sul Catarinense, Criciúma, Santa Catarina. 2008.

CIMIANO, P., Stumme, G., Hotho, A., & Tane, J. **Conceptual knowledge processing with formal concept analysis and ontologies.** In *Proceedings of The Second International Conference on FormalConcept Analysis (ICFCA 04)*.2004

CHAVES, Marcirio Silveira. **Gerenciamento e acesso a documentos na internet através de xml, RDF e ontologia.** Trabalho de Conclusão de Curso – Curso de

Informática, Universidade do Vale do Rio Dos Sinos, São Leopoldo. 2001. Disponível em: <http://www.inf.pucrs.br/~mchaves/pg_portugues/tc/mono_marcirio.pdf>. Acessado em: 13 out. 2006.

COWAN, J. Campbell. **Cardiologia.** 6. ed. São Paulo: Santos, 2000. 404 p.

DOMINGUE, J. **Tadzebao and webonto: discussing, browsing and editing ontologies on the web**. In: PROCEEDINGS OF THE 11TH BANFF KNOWLEDGE ACQUISITION WORKSHOP, 11., 1998, Alberta. Proceedings Alberta: [s.n.].

DIAS, Maria Madalena. **Um Modelo de Formalização do Processo de Desenvolvimento de Sistemas de Descoberta de Conhecimento em Banco de Dados**. 2001. 212 f. Tese (Doutorado em Engenharia de Produção)- Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal Santa Catarina. Florianópolis. 2001. Disponível em <http://teses.eps.ufsc.br/defesa/pdf/3469.pdf>. Acesso em 18 mar. 2007.

DIAS, Sandra Aparecida. **Integração Semântica de Dados.Através de Federação de Ontologias**. 2006. 79 f. Dissertação (Mestrado em Informática)- Programa de Pós-Graduação em Informática da PUC-Rio, Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro. 2006. Disponível em: <http://www.maxwell.lambda.ele.puc-rio.br/>. Acesso em 10 fev. 2007.

ESBÍZARO, André Luiz Dias**. Recuperação de Informações sobre *log*. de eventos apoiada em ontologias** 2006. 108 f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Braília, Brasília, 2006

FARIAS, Renan. **Ontologia para gestão do conhecimento em saúde por meio da metodologia methondology.** Trabalho de Conclusão de Curso – Ciência da Computação, Universidade do Extremo Sul Catarinense, Criciúma, Santa Catarina. 2006.

FARQUHAR, A; FIKES, R; RICE, J. The Ontolingua Server: A Tool for Collaborative Ontology Constuction. International Journal of Human Computer Studies. 1997.

FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005. 266 p.

FERNANDES, Anita Maria de Rocha. **Inteligência Artificial:** noções gerais**.** Florianópolis:Visual Books, 2003. 195 p.

FERNÁNDEZ-LÓPEZ, M; GÓMEZ-PÉREZ, A; JURISTO, N. **METHOTOLOGY: From Ontological Art Towars Ontological Engineering**. Spring Symposium on Ontological Engineering of AAAI. Stanford University, California, p. 33-40.

FERNÁNDEZ-LÓPEZ, M; GÓMEZ-PÉREZ, A; PAZOS, A; PAZOS, J. **Building a Chemical Ontology Using Methotology and the Ontology Design Environment**. IEEE Intelligent Systems & their Applications, [s. l.], p. 37-46, jan./feb. 1999.

Disponível em: <http://ieeexplore.ieee.org/iel4/5254/16144/00747904.pdf?arnumber=747904>. Acesso em: fev. 2007.

FORTE, Marcos. **Especificação de perfis e regras, baseada em ontologias, para adaptação de conteúdo na Internet**. 2006. 180 f. Dissertação (Programa de Pós-Graduação em Ciência da Computação)- Departamento de Computação. Universidade Federal de São Carlos. São Carlos. 2006. Disponível em: <http://www.bdtd.ufscar.br/tde_busca/arquivo.php?-codArquivo=1316>. Acesso em abr. 2007.

FROELICHER, Victor F.; MARCONDES, Gilberto D.; SETTINERI, Luiz Irineu. **Manual de teste ergométrico**. Porto Alegre: Artmed, 1992. 230 p.

GAŠEVIĆ, Dragan; DJURIĆ, Dragan; DEVEDŽIĆ, Vladan. **Model Driven Architecture and Ontology Development**. Verlag Berling Heidelberg: Springer, 2006. 315 p.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel Lopes. **Data mining:** uma guia prático : conceitos, técnicas, ferramentas, orientações e aplicações. Rio de Janeiro: Elsevier, 2005. 261 p.

GÓMEZ-PÉREZ, Asunción; FERNÁNDEZ-LÓPEZ, Mariano; CORCHO, Oscar. **Ontological engineering/** with examples from the areas of knowledge management, e-commerce and the Semantic Web. New York: Springer, 2004.

GRUBER, T. R. **Towards principles for the design of ontologies used for knowledge sharing.Formal ontology in conceptual analysis and knowledge representation**. *International Journal of Human-Computer Studies, 43,* 907-928. 1993.

GRUBER, T. **What is an ontology?** [S. l. : s. n.], 1996. Disponível em: <http://wwwksl.stanford.edu/ kst/ what-is-an-ontology.html>. Acesso em: 10 de fev. 2007

GUÉRIOS, Marlon Candido**. Uma Arquitetura para Utilização de Ontologias em Sistemas de Recuperação de Informação**. 2005. 108 f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2005.

GUSTAFSON, Donald E., KESSES, Willian **C. Fuzzy clustering with fuzzy covariance matrix.** Proceedings of the IEEE Control and Decision Conference, San Diego, p. 761-766, jan 1979.

HAN**,** Jiawei; KAMBER, Micheline**. Data mining**: concepts and techniques. San Francisco: Morgan Kaufmann Publishers, 2001.

HAND, David.; MANNILA, Heikki.; SMYTH, Padhraic**. Principles of Data Mining.** Massachusetts LondonEngland. A Bradford Book The MIT Press, 2001. 546 p.

HESS, Michael L. **Doenças cardíacas: primeiros cuidados.** São Paulo: Monole, 2002. 388 p.

HINZ, Verlani Timm. Proposta de Criação de uma Ontologia de Ontologias. 2006. 68 f. Dissertação (Pós-Graduação em Informática) – Programa de Pós-Graduação em Informática, Universidade Católica de Pelotas, Pelotas, 2006.

KANTARDZIC, Mehmed. **Data Mining: Concepts, Methods and Algorithms**. John Wiley & Sons. 2003. 385 p.

LANGLEY, P. **The computational support of scientific discovery**. *International Journal of Human-Computer Studies*. 2000.

LAROSE, Daniel, T. **Discovering Knowledge in data**: an introduction to data mining. New Jersey. John Wiley, 2005. 241 p.

_____**Data Mining Methods and Models**. New Jersey. John Wiley, 2005. 241 p.

LINHALIS, Flávia. **Mapeamento Semântico entre UNL e Componentes de Software para a Execução de Requisições Imperativas em Linguagem Natural**. 2007. 236 f. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional)- Instituto de Ciências Matemáticas e de Computação. ICMC-USP. São Carlos. 2007. Disponível em: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-18052007-103617/>. Acesso em jan. 2007.

LÓPEZ, Mariano Fernández. **A Suruey na Methodologies for developing, maintaining, evalvating and reengineering otologies**. Universidad politécnica de Madri. Madri. Spain. 2000. Disponível em:<http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/ontoweb Del 1-4.pdf>. Acesso em 10 jan. 2007.

LUGER, George F**. Inteligência Artificial:** estruturas e estratégias para a resolução de problemas complexos. 4. ed Porto Alegre: Bookman, 2004**.**

MCGUINNESS, D.L. (2002), “Ontologies come of age,” in *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, eds. D. Fensel, J. Hendler, H. Lieberman, & W. Wahlster, MIT Press, Boston, MA, pp. 1–18.

MAEDCHE, Alexander. **Ontology learning for the semantic web**. Boston: Kluwer Academic, 2002. 244 p.

MARTIMIANO, Luciana Andréia Fondazzi. **Sobre a estruturação de informação em sistemas. de segurança computacional: uso de ontologia**. 2006. 185 f. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional)- Instituto de Ciências Matemáticas e de Computação. ICMC-USP. São Carlos. 2006. Disponível em: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-02102006-091853/>. Acesso em jan. 2006.

MARTINS, Denis Piazza. **O Algoritmo de Particionamento *K-means* na Tarefa de Clusterização da *Shell* Orion *Data Mining Engine***. 2007. Trabalho de Conclusão de Curso – Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense, Criciúma, Santa Catarina, 2007.

MENZIES, T. **Cost Benefits of Ontologies**, **in Intelligence**, Fall 1999. p. 27-31.

MICHIELIN, Francisco**. Doenças do Coração**. São Paulo: Robe Editorial, 2003. 1395 p.

NIGRO, Hector Oscar; CISARO, Sandra Gonzalez; XODO, Daniel. **Data Mining with ontologies: implementations, findings and frameworks.** 2008. 289p.

NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** Barueri, SP: Manole, 2005. 1850 p.

NOY, N. F.; et.al. **The Knowledge Model of Protégé-2000**: combining interoperability and flexibility. 2000. Disponível em: <http://smi-web.stanford.edu/auslese/smi-web/reeports/SMI-2000-0830.pdf>. Acesso em: 10 de fev.2007.

NOY, Natalya Fridman and MCGUINNESS, Deborah L. Ontology Development 101:**A Guide to Creating Your First Ontology**. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001. Disponível em: <http://protege.stanford.edu/publications/-ontologydevelopment/ontology101.html>. Acessado em: 13 out. 2006.

OLIVEIRA, Rodrigo Réus de. **Uso de Data Mining para Obter perfis de clientes com maior lucratividade**. Universidade do Extremo Sul Catarinense. 2006.

PAL, Sankar K.; MITRA, Pabitra. **Pattern recognition algorithms for data mining:** scalability, knowledge discovery and soft granular computing. Florida: Chapman & Hall, c2004. 244 p.

PELEGRIN, Diana Colombo. **A tarefa de classificação e o algoritmo ID3 para a indução de árvores de decisão na Shell de Data Mining Orion. 2005.** Trabalho de Conclusão de Curso – Ciência da Computação, Universidade do Extremo Sul Catarinense, Criciúma, Santa Catarina, 2005.

PIZZI, Luciene Cristina. **Mineração de Dados Em Múltiplas Tabelas**: O Algoritmo GFP-Growth. 2006. 118 f. Dissertação (Mestrado em Ciências da Computação)- Programa de Pós-Graduação em Ciência da Computação, Universidade Federal São Carlos. São Carlos. São Paulo 2006. Disponível em <http://www.bdtd.ufscar.br/tde_busca/arquivo.php?codArquivo> Acesso em 18 mai. 2007.

PUNTAR, Sergio Gonçalves. **Métodos e Visualização de Agrupamentos de Dados**. 2003. 131 f. Tese (Mestrado em Ciências e Engenharia Civil)- Programa de Pós-Graduação em Engenharia, Universidade Federal do Rio de Janeiro. Rio de Janeiro.

2003. Disponível em <www.cin.ufpe.br/~tg/2006-2/tmas.pdf>. Acesso em 15 mai. 2007.

RAIMUNDO, Lidiane Rosso. **O Algoritmo CART na Tarefa de Classificação da *Shell* Orion *Data Mining Engine***. 2007. Trabalho de Conclusão de Curso – Curso de Ciência da Computação, Universidade do Extremo Sul Catarinense, Criciúma, Santa Catarina, 2007.

REZENDE**,** Solange Oliveira**. Sistemas inteligentes: fundamentos e aplicações. Barueri, SP: Manole**, 2005. 525 p.

RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência artificial**. Rio de Janeiro: Elsevier, 2004.

SILVA, Marcelino Pereira dos Santos. **Mineração de Imagens Usando Ontologias**. Monografia da Proposta de Tese apresentada ao Programa de Pós-Graduação em Computação Aplicada do Instituto Nacional de Pesquisas Espaciais de São José dos Campos. 2004.
Disponível em: <http://www.dpi.inpe.br/~mpss/docs/PropostaMarcelino.pdf>. Acessado em: 13 out. 2006.

STAAB, Steffen; STUDER, Rudi. **Handbook on ontologies.** Berlin: Springer, c2004.

SURE, York et al. **OtoEdit: Collaborative Ontology Engineering for the Semantic Web**. In: Proccedings of the International Semantic Web Conference 2002 (ISWC 2002), June 9-12 2002, Sardinia, Italia. Disponível em: www.aifd.uni-karlsruhe.de/WBS/yus/publications/2002 iswc ontoedit.pdf. Acesso em: 12 fev. 2007.

TRINKUNAS, Justas; VASILECAS, Olegas. **Building Ontologies from Relational Databases Using Reverse Engineering Methods**. International Conference on Computer Systems and Technologies - *CompSysTech'07*. 2007. Disponível em: Acessado em: 01 out. 2008.

WITTEN, I. H; FRANK, Eibe. **Data mining** : **practical machine learning tools and tecniques with Java implementations**. San Francisco: Morgan Kaufmann Publishers, 2000.

_____. **Data mining**: **practical machine learning tools and tecniques.** San Francisco: Morgan Kaufmann, 2005.

APÊNDICE A – METODOLOGIAS PARA A CONSTRUÇÃO DE ONTOLOGIAS

1. USCHOLD E KING'S

Esta metodologia foi criada com o intuito de desenvolver uma ontologia voltada para a modelagem de negócios empresarias chamada de Enterprise Ontology (ESBÍZARO, 2006).

O processo de construção da ontologia é totalmente separado do uso da mesma, ou seja, é independente de aplicação. Segundo Gómez-Pérez, Fernández-López e Corcho (2004, tradução nossa) as seguintes atividades fazem parte do procedimento de desenvolvimento:

a) a identificação o propósito: onde o objetivo visa esclarecer a razão pelo qual a ontologia esta sendo desenvolvida e intenção de uso;
b) criação: ocorre propriamente a construção da ontologia, e esta dividida em três fases :
    - capturar a ontologia: fazer a identificação dos conceitos chaves e relacionamentos do domínio de interesse eliminando-se às duplicidades para cada conceito,
    - codificação: envolve representar o conhecimento adquirido na fase de captura na forma de uma linguagem formal,
    - integração a ontologias existentes: visa as integrações das ontologias novas com as já existentes;
c) avaliação: onde as ontologias construídas são associadas a ambientes de desenvolvimento para a sua avaliação;

d) documentação: recomenda o estabelecimento de documentação de acordo com o tipo e propósito da ontologia desenvolvida.

2. GRUNINGER E FOX

Esta metodologia foi empregada no projeto *Toronto Virtual Enterprise* (TOVE), desenvolvido pelo *Enterprise Integration Laboratory*, da Universidade de Toronto, desenvolvida para as atividades de negócio, basicamente envolve a construção de um modelo lógico do conhecimento que especifica as carências da ontologia. Este método não é construído diretamente, primeiramente uma descrição informal é elaborada da especificação da ontologia e após isto a descrição é formalizada (ESBÍZARO, 2006).

A construção desta ontologia esta dividida nas seguintes fases (LÓPEZ, 2000, tradução nossa):

a) capturar cenários de motivação: é motivada por cenários específicos que destacam-se na aplicação. Buscam a resolução de problemas por meio dos cenários motivadores que fornecem um conjunto de possíveis soluções para as ontologias que não se adequaram;

b) formular questões de competência informal: as questões são baseadas nos cenários obtidos na fase anterior, estas expressam requerimentos, axiomas e definições para sua construção. As questões de competência são usadas para avaliar a ontologia devendo esta ser capaz de representar por meio de sua termologia, todas as suas operações.

c) especificação da terminologia utilizando uma linguagem formal: esta etapa esta dividida em duas: primeiramente quando as questões de competência informais estiverem disponíveis serão analisadas e os

termos utilizados serão extraídos; uma vez analisadas na segunda etapa pode-se propor uma nova ontologia ou até mesmo a extensão de uma ontologia existente, a terminologia então é especificada com o uso da lógica de primeira ordem;

d) formulação de questões de competência com uso da terminologia da ontologia: estabelecidas às definições das questões de competência informais bem como a terminologia se faz necessário então que as questões sejam definidas formalmente;

e) especificação dos axiomas e definições dos termos da ontologia com linguagem formal: os axiomas, que são definidos com a utilização da lógica de primeira ordem para os objetos das ontologias, devem prover a definição o significado dos termos e restringirem sua interpretação. O processo de desenvolvimento de axiomas é iterativo, por isso se estas não forem suficientes para representar as questões de competências formais e caracterizar as soluções das questões, devem ser adicionados objetos ou axiomas ate que sejam suficientes.

f) estabelecer condições para caracterizar a integridade da ontologia: estabelecidas às questões de competência declaradas formalmente, devem ser definidas as condições para que as soluções destas questões sejam completas.

3. KACTUS

É um projeto ESPRIT-iii europeu criado por Amaya Bernaras e seus colegas que visa o desenvolvimento de uma metodologia para reutilização dos conhecimentos

adquiridos por meio das ontologias em todas as suas fases de construção (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).

Para isto utilizam a mesma base de conhecimento para cada vez que um padrão seja construído a ontologia que representa o modelo definido pode reutilizada ou ser integrada a outras em diferentes áreas.

As fases de construção de ontologias por meio desta metodologia são: (STAAB; STUDER, 2004, tradução nossa):

a) a especificação da aplicação onde é fornece todos os requisitos para o desenvolvimento de um modelo de ontologia baseados nestes componentes. É o processo que lista os termos e as tarefas que serão desenvolvidas;
b) o projeto preliminar baseado nas categorias aqui são descritas baseadas no modelo pretendido as categorias, conceitos, relações e atributos;
c) a definição do refinamento nesta fase se a estruturação das ontologias ou seja são definidos a hierarquia a organização bem como a estruturação principais da ontologia desenvolvida.

4. ON-TO-KNOWELEDGE

Esta metodologia objetiva auxiliar a administração da construção de conceitos em organizações, identificando metas para as ferramentas de gestão do conhecimento e utilizando cenários e contribuições dos provedores e clientes de informação da organização realiza o estudo sobre ontologias existentes antes da construção de uma nova, valorizando o reuso do conhecimento.

As principais fases da construção de ontologias por esta metodologia são (STAABET et al., 2001, tradução nossa):

a) kick-of: onde os conceitos são capturados e especificados, questões de relacionamentos são identificados, ontologias reusáveis são estudadas e a partir disso é criada uma versão;
b) refinamento: é desenvolvida uma ontologia para aplicação;
c) avaliação: onde são analisados se os conceitos da ontologia estão de acordo com sua especificação;
d) manutenção: depois da avaliação se houver necessidade de alguma mudança é nesta fase que isso ocorre.

APÊNDICE B – FERRAMENTAS PARA O DESENVOLVIMENTO DE ONTOLOGIAS

1. THE ONTOLINGUA SERVER

Desenvolvida em 1990 pelo Knowledge Systems Laboratory (KLS) de Stanfort University com o objetivo de fornecer um ambiente colaborativo distribuído para criar, editar, modificar e utilizar ontologias entre grupos (FARQUHAR et al, 1997, tradução nossa).

Nesta ferramenta existem os módulos: Servidor Ontolingua (editor das ontologias), servidor (Open Knowledge Based Connectivity) OKBC e o Chimaera que faz a análise, o *merging* e a integração das ontologias (LINHALIS, 2007).

Ontolingua Server é implementada em Lisp e foi criada para trabalhar com a linguagem Ontolingua, mas possui tradutores para outras linguagens. O usuário pode utilizar a biblioteca existente para definir todo o processo de desenvolvimento, este ambiente está disponível no *site*: http://ontolingua.stanford.edu/ (GÓMEZ-PÉEREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).

Essa plataforma possibilita três diferentes formas de montar uma ontologia (FARQUHAR et al, 1997, tradução nossa):

a) **por inclusão**: cria-se uma nova ou pode-se utilizar a biblioteca existente e usar definições prontas para modelar uma ontologia;

b) **reutilização**: uma ontologia pode importar definições de uma outra e torná-las mais específica de acordo com o seu objetivo;

c) **refinamento de poliformismo**: toda uma ontologia pode ser redefinida por meio de importações de outros conceitos.

Estas características permitem visualizar gráficos periódicos da inclusão e modificação das ontologias. Na Figura 52, tem-se a construção da ontologia Voar, documentação e subclasse Viagem.

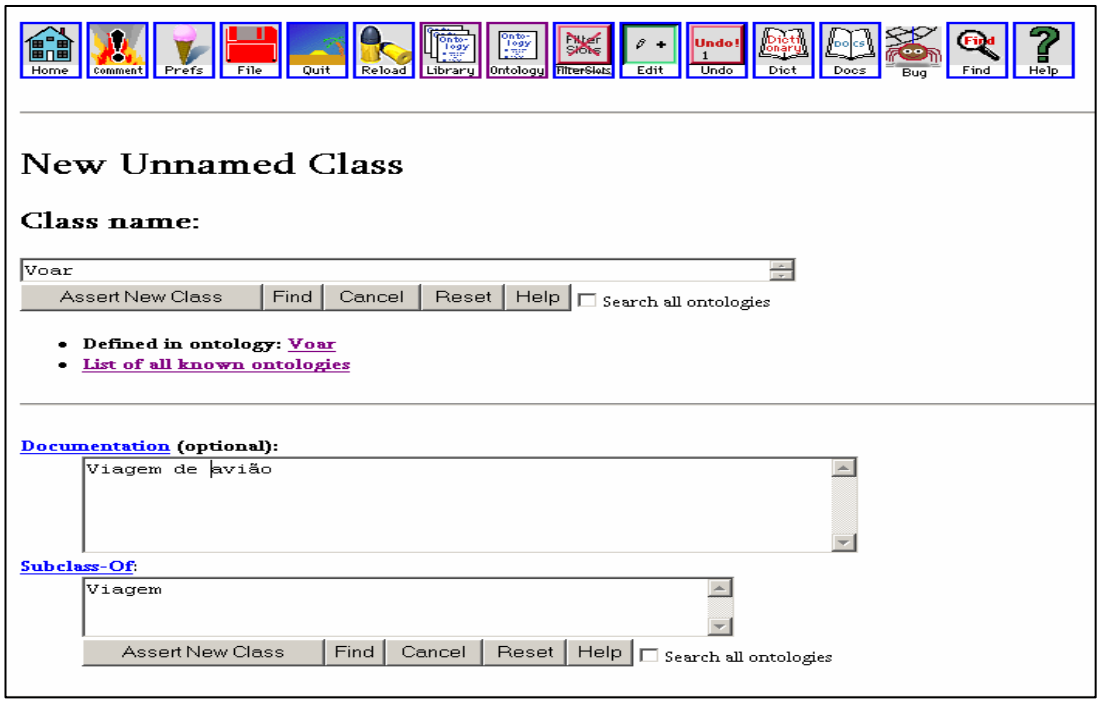


Figura 52. Interface da ferramenta The Ontolingua Server
Fonte: Snowledge Systems Laboratory (1990).

2. OILED

Ferramenta desenvolvida pela Universidade de Manchester, basicamente é um editor simples que permite ao usuário criar e editar ontologias. OilED possui código livre, implementado em Java, tem como objetivo demonstrar o uso, e estimular o interesse no desenvolvimento de uma ontologia. (BECHHOFER et al, 2001, tradução nossa).

OilEd esta disponível para *download* gratuitamente no *site*: http//oiled.mam.ac.uk/.

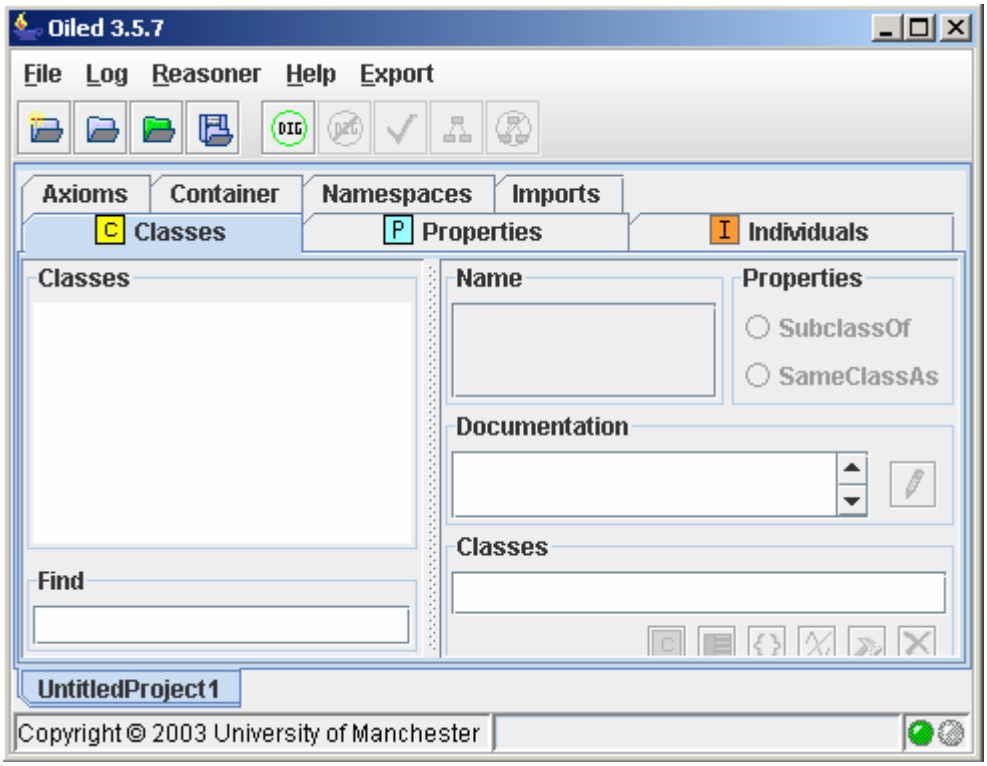


Figura 53. Interface da ferramenta OilEd
Fonte: University of Manchester (2003)

Contundo, conforme observado na Figura 53 sua interface possui muitas funcionalidades que permite a modelagem básica, demonstrar e verificar se as ontologias desenvolvidas estão consistentes.

3. WEBONTO

WebOnto, desenvolvida pelo Knowledge Media Institute (KMI) na Universidade do Reino Unido, foi construída para a criação e edição das ontologias, fornecendo uma relação direta com a manipulação das definições de maneira fácil. A arquitetura da ferramenta é formada por um usuário central e os clientes escritos em Java (DOMINGUE, 1998, tradução nossa).

Consiste em dois módulos principais: o usuário e o editor que estabelece a interação para a construção de ontologias tendo a possibilidade de editar, definir taxonomias e elaborar as relações entre os conceitos, inteiramente de forma gráfica e

representada por meio de árvores (GÓMEZ-PÉEREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).

A Figura 54 mostra a interface desta ferramenta que está disponível no *site*: http://webonto.open.ac.uk/

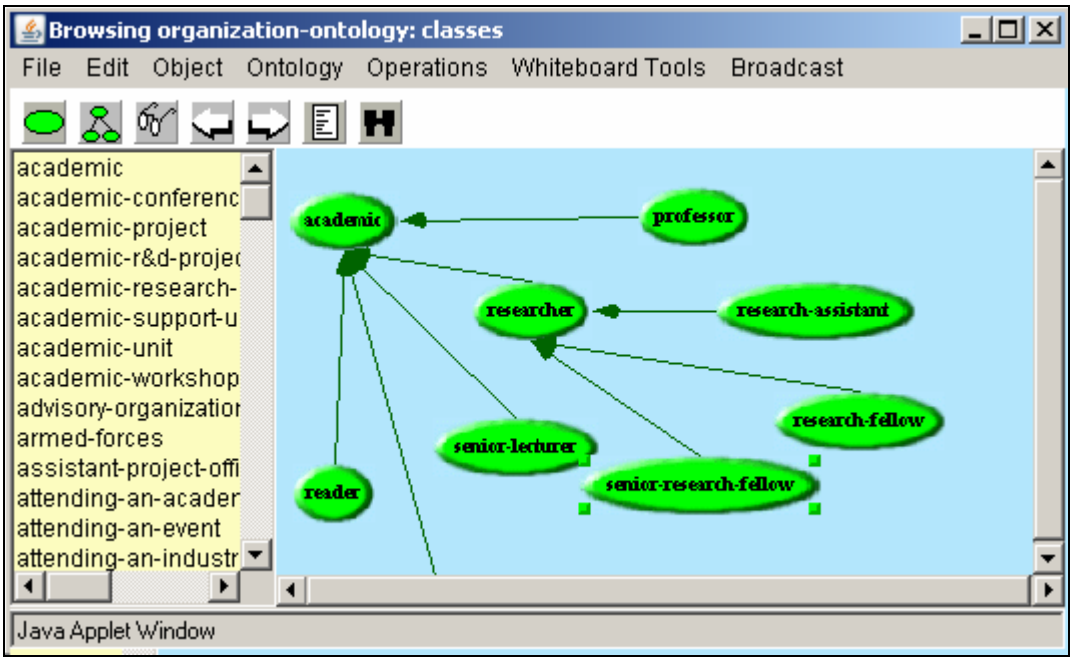


Figura 54. Interface da ferramenta WebOnto
Fonte: Knowledge Media Institute (2008)

4 WEBODE

Desenvolvida no laboratório de Inteligência Artificial da Universidade Técnica de Madri (*Technical University of Madrid)* (UPM), dá suporte a maioria das atividades de desenvolvimento de ontologias (GÓMEZ-PÉEREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).

Ferramenta baseada em tabelas e gráficos que permite personalizar a ontologia construída de acordo com a necessidade, é independente de plataforma, uma vez que é implementada em Java (STAAB; STUDER, 2004, tradução nossa).

O modelo gerado das ontologias é armazenado em um repositório central, sendo utilizado para isto um banco de dados relacional, permite que o mesmo modelo

seja usado para cenários diferentes. WebODE é baseada na metodologia Methontology para o desenvolvimento das ontologias possui todo o suporte necessário (GÓMEZ-PÉEREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).

A WebODE (Figura 55) é uma ferramenta gratuita e está disponível para *download* no *site*: http://webode.dia.fi.upm.es/WebODEWeb/index.html.

Figura 55. Interface do WebODE
Fonte: Ontological Engineering Group (2007)

5. ONTOEDIT

A ferramenta OntoEdit foi desenvolvida por um grupo de pesquisadores do *Institut of Applied Informatics and Formal Description Methods* (AIFB) na Universidade de *Karlsruhe*, e está sendo comercializado pela empresa Ontoprise. A versão atual é a 2.7, de Janeiro de 2007 (GÓMEZ-PÉEREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004, tradução nossa).

OntoEdit possui um ambiente gráfico para edição de ontologias que permite inspeção, navegação, codificação e alteração de ontologias. As ontologias são

armazenadas em bancos relacionais e podem ser implementadas em *XML*, *FLogic*, *RDF(S)* e *DAML+OIL* (SURE et al, 2002, tradução nossa).

A OntoEdit (Figura 56) é baseada na metodologia On-To-Knowledge que está constituída em três grandes processos (STAAB; STUDER, 2004, tradução nossa):

a) especificação: consiste na descrição do domínio e no objetivo da ontologia;

b) refinamento: descrição formal dos resultados obtidos da fase anterior;

c) evolução: verificar se a ontologia desenvolvida está de acordo com os requisitos estabelecidos na especificação.

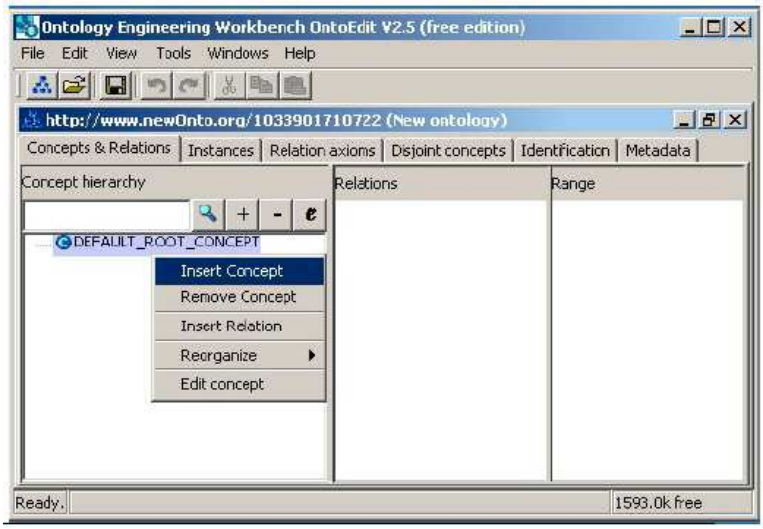


Figura 56. Interface do OntoEdit
Fonte: Ontoprise (2007)

A ferramenta completa profissional não é gratuita, entretanto possui uma versão de demonstração disponível gratuitamente para *download* no *site* http://www.ontoprise.de/customercenter/software_downloads

APÊNDICE C – GLOSSÁRIO DE TERMOS - CONCEITOS

Tabela 24. Descrição dos conceitos utilizados na ontologia DDC

Conceitos	Descrição
Diagnóstico de Doença Coronariana	Primeiro e principal conceito da hierarquia da ontologia DDC.
Conceito de Diagnóstico de Doença Coronariana	Classe contendo a conceitualização do termo DDC
Diagnósticos	Classe contendo alguns diagnósticos de ausência e presença de doença coronariana.
Tipos de Exames	Classe contendo alguns tipos de exames: físicos, laboratoriais e complementares necessários para o DDC

Tabela 25. Descrição dos conceitos utilizados na classe exames físicos

Conceitos	Descrição
Exames Físicos	Classe contendo os exames físicos: dor precordial e pressão arterial para diagnóstico de doença coronariana
Dor Precordial	Classe contendo conceitos e características de dor precordial
Conceitos de Dor Precordial	Definição do conceito de dor precordial
Características de Dor Precordial	Apresenta algumas características de dor precordial
Pressão Arterial	Classe contendo conceitos e estágios de pressão arterial
Conceitos de Pressão Arterial	Definição do conceito de pressão arterial
Estágios de Pressão Arterial	Apresenta alguns estágios de pressão arterial

Tabela 26. Descrição dos conceitos utilizados na classe exames laboratoriais

Conceitos	Descrição
Exames Laboratoriais	Classe contendo exames laboratoriais para diagnóstico de doença coronariana
Colesterol	Classe contendo conceitos e níveis de colesterol
Conceitos de Colesterol	Definição do conceito de colesterol
Níveis de Colesterol	Apresentam algumas informações dos níveis de colesterol
Glicemia	Classe contendo conceitos e informações de níveis de glicemia
Conceitos de Glicemia	Definição do conceito de glicemia
Informações de Níveis de Glicemia	Apresenta algumas informações de níveis de glicemia

Tabela 27. Descrição dos conceitos utilizados na classe exames complementares

Conceitos	Descrição
Exames Complementares	Classe contendo exames complementares para Diagnóstico de Doença Coronariana
Eletrocardiograma	Classe contendo conceitos e informações de eletrocardiograma
Conceitos de Eletrocardiograma	Definição do conceito de eletrocardiograma
Informações de Eletrocardiograma	Apresenta algumas Informações de eletrocardiograma
Teste Ergométrico	Classe contendo conceitos e informações de teste ergométrico
Conceitos de Teste Ergométrico	Definição do conceito de teste ergométrico
Informações de Teste Ergométrico	Apresenta algumas Informações sobre o teste ergométrico
Cateterismo	Classe contendo conceitos e informações sobre cateterismo
Conceitos de Cateterismo	Definição do conceito de Cateterismo
Informações Cateterismo	Apresenta algumas Informações sobre Cateterismo

APÊNDICE D – ÁRVORE DE CLASSIFICAÇÃO DE CONCEITOS

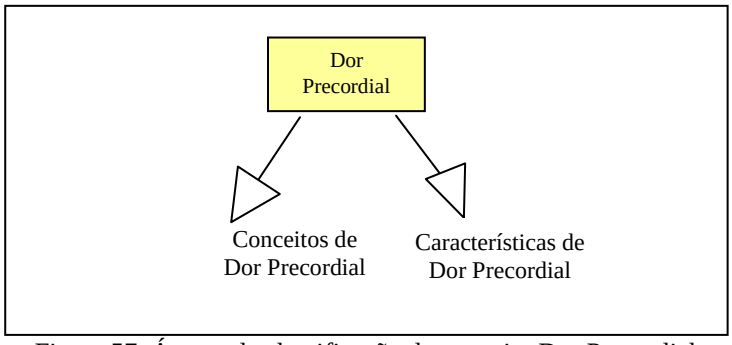


Figura 57. Árvore de classificação do conceito Dor Precordial

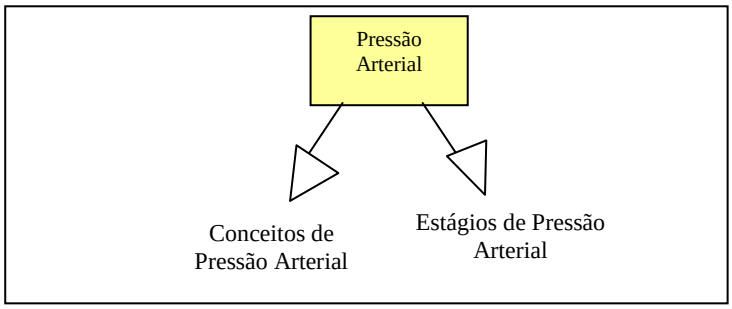


Figura 58. Árvore de classificação do conceito Pressão Arterial

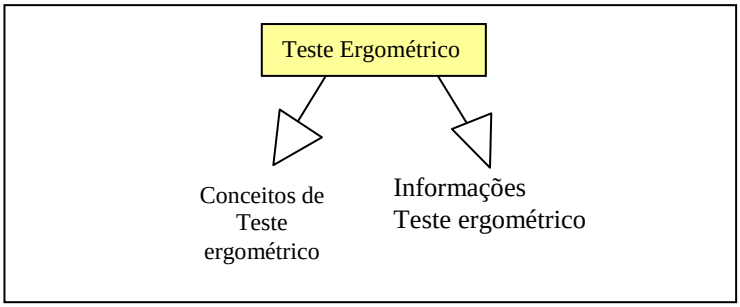
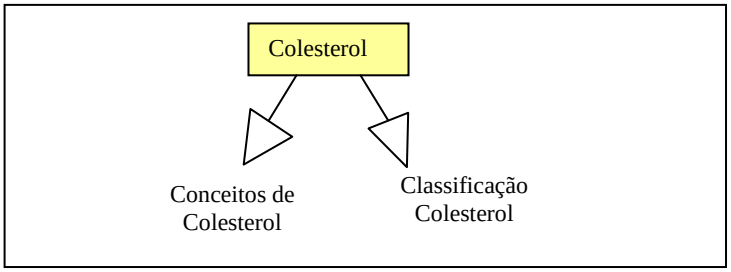


Figura 59. Árvore de classificação do conceito Teste Ergométrico



.Figura 60 Árvore de classificação do conceito Colesterol

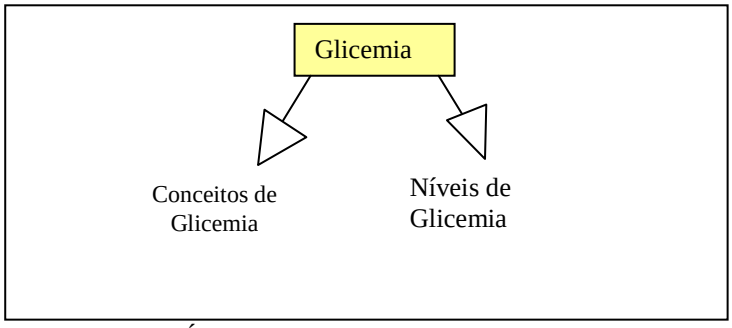


Figura 61. Árvore de classificação do conceito Glicemia

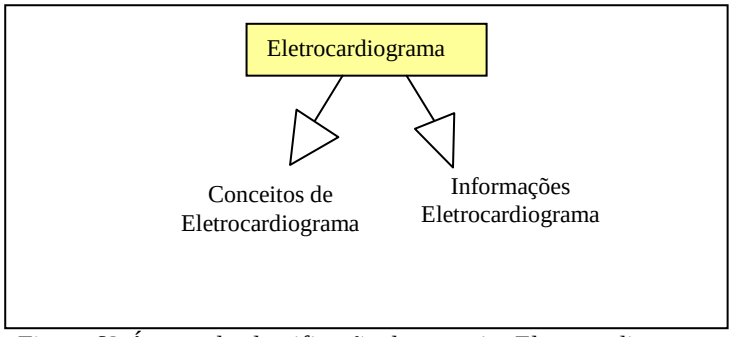


Figura 62.Árvore de classificação do conceito Eletrocardiograma

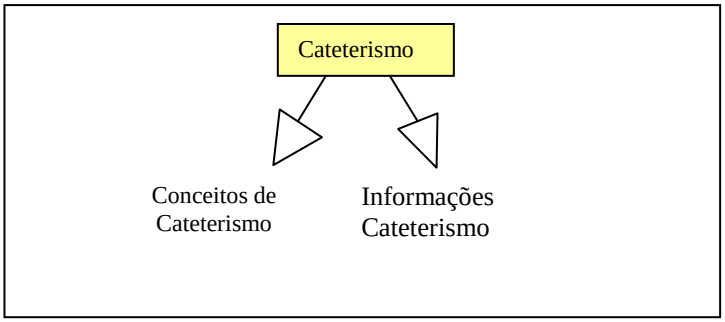


Figura 63. Árvore de classificação do conceito Cateterismo

APÊNDICE E - DICIONÁRIO DE CONCEITOS

Tabela 28. Árvore de classificação de conceitos - subconceitos de DDC

Conceito	Tipo	Hierarquia
Diagnóstico de Doença Coronariana	Abstrato	Conceito Principal
Conceitos Diagnóstico de Doença Coronariana	Concreto	Subconceito de DDC
Diagnósticos	Concreto	Subconceito de DDC
Tipos de Exames	Abstrato	Subconceito de DDC

Tabela 29. Árvore de classificação de conceitos - subconceitos de Tipos de Exames

Conceito	Tipo	Hierarquia
Exames Físicos	Abstrato	Subconceito de Tipos de Exames
Exames Laboratoriais	Abstrato	Subconceito de Tipos de Exames
Exames Complementares	Abstrato	Subconceito de Tipos de Exames

Tabela 30. Árvore de classificação de conceitos - subconceitos de Exames Físicos

Conceito	Tipo	Hierarquia
Dor Precordial	Abstrato	Subconceito de Exames Físicos
Pressão Arterial	Abstrato	Subconceito de Exames Físicos

Tabela 31. Árvore de classificação de conceitos - subconceitos de Exames Laboratorias

Conceito	Tipo	Hierarquia
Colesterol	Abstrato	Subconceito de Exames Laboratoriais
Glicemia	Abstrato	Subconceito de Exames Laboratoriais

Tabela 32. Árvore de classificação de conceitos - subconceitos de Exames Complementares

Conceito	Tipo	Hierarquia
Eletrocardiograma	Abstrato	Subconceito de Exames Complementares
Teste Ergométrico	Abstrato	Subconceito de Exames Complementares
Cateterismo	Abstrato	Subconceito de Exames Complementares

Tabela 33. Árvore de classificação de conceitos - subconceitos de Dor Precordial

Conceito	Tipo	Hierarquia
Conceitos Dor Precordial	Abstrato	Subconceito de Dor Precordial
Características Dor Precordial	Abstrato	Subconceito de Dor Precordial

Tabela 34. Árvore de classificação de conceitos - subconceitos de Pressão Arterial

Conceito	Tipo	Hierarquia
Conceitos Pressão Arterial	Abstrato	Subconceito de Pressão Arterial
Estágios Pressão Arterial	Abstrato	Subconceito de Pressão Arterial

Tabela 35. Árvore de classificação de conceitos - subconceitos de Glicemia

Conceito	Tipo	Hierarquia
Conceitos Glicemia	Abstrato	Subconceito de Glicemia
Nóveis de Glicemia	Abstrato	Subconceito de Glicemia

Tabela 36. Árvore de classificação de conceitos – subcnceitos Colesterol

Conceito	Tipo	Hierarquia
Conceitos Colesterol	Abstrato	Subconceito de Colesterol
Classificação Colesterol	Abstrato	Subconceito de Colesterol

Tabela 37. Árvore de classificação de conceitos - subconceitos de Eletrocardiograma

Conceito	Tipo	Hierarquia
Conceitos Eletrocardiograma	Abstrato	Subconceito de Eletrocardiograma
Informações Eletrocardiograma	Abstrato	Subconceito de Eletrocardiograma

Tabela 38. Árvore de classificação de conceitos - subconceitos de Teste Ergométrico

Conceito	Tipo	Hierarquia
Teste Ergométrico	Abstrato	Subconceito de Teste Ergométrico
Informações Teste Ergométrico	Abstrato	Subconceito de Teste Ergométrico

Tabela 39. Árvore de classificação de conceitos - subconceitos de Cateterismo

Conceito	Tipo	Hierarquia
Conceitos de Cateterismo	Abstrato	Subconceito de Cateterismo
Informações de Cateterismo	Abstrato	Subconceito de Cateterismo

APÊNDICE F – TABELAS DE INSTÂNCIAS

Tabela 40. Instância do conceito de Diagnósticos

Conceito	**Diagnósticos**
Relação	**Instância**
Descrição	Diagnóstico de Ausência de Doença Coronariana
Definição	Em 61,31 % registros foi constatado ausência de doença coronariana.
Faixa etária	Vinte a Setenta anos.
Sexo	Masculino
Fonte	*Heart Disease Database* (http://archive.ics.uci.edu/ml/datasets/Heart+Disease)

Tabela 41. Instância do conceito de Diagnósticos

Conceito	**Diagnósticos**
Relação	**Instância**
Descrição	Diagnóstico de Presença de Doença Coronariana
Definição	Em 38,69 % registros foi constatado ausência de doença coronariana. .
Faixa Etária	Vinte a Setenta anos
Sexo	Feminino
Fonte	*Heart Disease Database* (http://archive.ics.uci.edu/ml/datasets/Heart+Disease)

Tabela 42. Instância do conceito de dor precordial

Conceito	**Dor Precordial**
Relação	**Instância**
Descrição	Dor Precordial
Definição	Dor precordial abrange uma enorme e complexa gama de possibilidades clínicas que devem ser investigadas.
Observação	De acordo com as características podem ser classificadas em angina típica, angina atípica e assintomático.
Fonte	MICHIELIN, Francisco**. Doenças do Coração**. São Paulo: Robe Editorial, 2003.

Tabela 43. Instância do conceito de dor precordial – subconceito características

Conceito	**Características de dor precordial**
Relação	**Instância**
Descrição	Angina Típica
Definição	Pode ser causada por: Desconforto ou dor retroesternal; Desencadeada pelo exercício, ou estresse emocional;
Observação	Pode ser aliviada com repouso
Fonte	NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** Barueri, SP: 2005.

Tabela 44. Instância do conceito de dor precordial

Conceito	Características de dor precordial
Relação	**Instância**
Descrição	Dor Precordial Angina Típica
Definição	A angina manifestar-se por uma sensação de opressão, a dor não é intensa, contudo causa desconforto.
Observação	Além da dor, deve-se observar os sintomas e sinais
Fonte	COWAN, J. Campbell. **Cardiologia.** 6. ed. São Paulo: Santos, 2000.

Tabela 45. Instância do conceito de dor precordial – subconceito características de dor precordial

Conceito	Características de dor precordial
Relação	**Instância**
Descrição	Dor Precordial Angina Atípica
Definição	Possui características da angina mas com variações.
Observação	A localização da dor é um fator importante para o diagnóstico.
Fonte	FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005.

Tabela 46. Instância do conceito de dor precordial

Conceito	Características de dor precordial
Relação	**Instância**
Descrição	Dor de origem não coronariana e/ou assintomático.
Definição	Dor precordial assintomático não possui dores de origens cardíacas.
Observação	A localização da dor deriva das estruturas torácicas, dando a impressão de dor cardíaca.
Fonte	FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005.

Tabela 47. Instância do conceito de pressão arterial

Conceito	Pressão Arterial
Relação	**Instância**
Descrição	Pressão Arterial
Definição	A pressão arterial é aquela existente no interior das artérias e comunicada às suas paredes.
Observação	Pressão sistólica, momento em que o coração bombeia o sangue. Pressão diastólica, momento em que o coração relaxa.
Fonte	NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** 2005.

Tabela 48. Instância do conceito de pressão arterial – subconceito estágios de pressão arterial

Conceito	**Estágios de pressão arterial**
Relação	**Instância**
Descrição	Pressão Arterial Normal
Definição	Os valores para a pressão normal são: Pressão sistólica menor que 130mmHg
Observação	Pressão diastólica menor 85mmHg
Fonte	NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** Barueri, SP: Manole, 2005.

Tabela 49. Instância do conceito de pressão arterial – subconceito estágios de pressão arterial

Conceito	**Estágios de pressão arterial**
Relação	**Instância**
Descrição	Pressão Arterial ótima
Definição	Os valores para a pressão arterial ótima são: Pressão sistólica menor que 120 mmHg
Observação	Pressão diastólica menor que 80 mmHg
Fonte	NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** Barueri, SP: Manole, 2005.

Tabela 50. Instância do conceito de pressão arterial – subconceito estágios de pressão arterial

Conceito	**Estágios de pressão arterial**
Relação	**Instância**
Descrição	Pressão Alta ou Hipertensão
Definição	Os valores para hipertensão são: Pressão sistólica maior e igual a 140mmHg Pressão diastólica maior e igual a 90mmHg
observação	pressão sistólica acima de 159mmHg e pressão diastólica acima de 99mmHg hipertensão grave.
fonte	HESS, Michael L. **Doenças cardíacas: primeiros cuidados.** São Paulo: Monole, 2002.

Tabela 51. Instância do conceito de pressão arterial – subconceito estágios de pressão arterial

Conceito	**Estágios de pressão arterial**
Relação	**Instância**
Descrição	Informação Pressão Arterial Limítrofe
Definição	Os valores para a pressão arterial limítrofe: Pressão sistólica entre 130 mmHg e 139 mmHg.
Observações	Pressão diastólica entre 85 e 89 mmHg.
Fonte	NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** Barueri, SP: Manole, 2005.

Tabela 52. Instância do conceito de colesterol

Conceito	**Colesterol**

Relação	Instância
Descrição	Colesterol
Definição	Níveis de gorduras no sangue composto por lipídios e triglicerídios.
Observações	Os lipídios são formados pelo HDL e LDL.
Fonte	FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005.

Tabela 53. Instância do conceito de colesterol – subconceito classificação de colesterol

Conceito	Classificação de colesterol
Relação	**Instância**
descrição	Classificação Colesterol Total
definição	Classificação de colesterol total Desejável - menor que 200mg/dl Limítrofes - entre 200 a 239mg/dl
Observações	Aumentados - maior ou igual a 240mhg/dl
fonte	FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005. 266 p.

Tabela 54. Instância do conceito de colesterol – subconceito classificação de colesterol

Conceito	Classificação de colesterol
Relação	**Instância**
Descrição	Classificação Colesterol LDL
Definição	Desejável - menor que 100mg/dl Acima do normal - entre 100 a 129mg/dl Limítrofes - entre 130 a 159mg/dl Alto - entre 160 a 189mg/dl
Observações	Muito alto, maior e igual a 190mg/dl
Fonte	FARRET, Jacqueline Faria. **Nutrição e oenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005.

Tabela 55. Instância do conceito de colesterol – subconceito classificação de colesterol

Conceito	Classificação de colesterol
Relação	**Instância**
Descrição	Colesterol HDL
Definição	Colesterol HDL Baixo - menor que 40mg/dl
Observações	Alto - maior que 60mg/dl
Fonte	FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005.

Tabela 56. Instância do conceito de hiperglicemia

Conceito	Glicemia
Relação	**Instância**
Descrição	A Glicemia em Jejum
Definição	Níveis de açúcar no sangue, mas precisamente no plasma.
Observações	A Glicemia em Jejum deve ser maior e igual a 110mg/dl
Fonte	NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** Barueri, SP: Manole, 2005.

Tabela 57. Instância do conceito de eletrocardiograma

Conceito	Eletrocardiograma
Relação	**Instância**
Descrição	Definição Eletrcardiograma ECG
Definição	A análise do ECG é feita por meio de uma minuciosa avaliação de todas as medidas e duração das ondas e intervalos.
Observações	Ondas e segmentos analisados; P, PR,PRS, J, R, Q, S, QRS, ST,QT, T e U
Fonte	MICHIELIN, Francisco**. Doenças do Coração**. São Paulo: Robe Editorial, 2003.

Tabela 58. Instância do conceito de eletrocardiograma – subconceito informações eletrocardiograma

Conceito	Informações eletrocardiograma
Relação	**Instância**
Descrição	Informação ECG em repouso resultado anormal
Definição	O ECG anormal, tendo inversões na onda T e ou onda ST e ainda elevação ou depressão maior que 0,05 mV.
Observações	Outros exames são necessários para o diagnóstico de alguma doença cardíaca.
Fonte	MICHIELIN, Francisco**. Doenças do Coração**. São Paulo: Robe Editorial, 2003.

Tabela 59. Instância do conceito de eletrocardiograma- subconceito informações eletrocardiograma

Conceito	Informações eletrocardiograma
Relação	**Instância**
Descrição	Informação Eletrocardiograma em repouso resultado normal
Definição	Um traçado ECG normal expressa ritmo sinusal, medidas de segmentos e intervalos dentro dos limites da normalidade e duração normal.
Observações	
Fonte	MICHIELIN, Francisco**. Doenças do Coração**. São Paulo: Robe Editorial, 2003.

Tabela 60. Instância do conceito de teste ergométrico

Conceito	Teste ergométrico
Relação	**Instância**
Descrição	Teste Ergométrico
Definição	O teste ergométrico é de grande importância para o diagnóstico de doenças cardiovasculares devido à sua grande disponibilidade e à amplitude das informações fornecidas.
Observações	Teste de esforço físico tipo: esteira e bicicleta .
Fonte	MICHIELIN, Francisco**. Doenças do Coração**. São Paulo: Robe Editorial, 2003.

Tabela 61. Instância do conceito de teste ergométrico

Conceito	Informações Teste ergométrico
Relação	**Instância**
Descrição	Informações Teste Ergométrico
Definição	No teste ergométrico o princípio básico é tentar levar o paciente a atingir sua Frequência cardíaca máxima (FCM), que é obtida subtraindo-se 220 da idade.
Observações	A FCM pode variar de acordo com o paciente devido alguns fatores como: sedentarismo e obsidade.
Fonte	MICHIELIN, Francisco**. Doenças do Coração.** São Paulo: Robe Editorial, 2003.

Tabela 62. Instância do conceito de cateterismo

Conceito	Cateterismo
Relação	**Instância**
Descrição	Conceito de Cateterismo
Definição	Exame que permite a avaliação quantitativa da estrutura da função cardíaca e das artérias coronárias.
Observações	Observa-se o grau de estreitamento das artérias.
Fonte	MICHIELIN, Francisco**. Doenças do Coração**. São Paulo: Robe Editorial, 2003.

APÊNDICE G – TABELAS DESENVOLVIDAS PARA O POSTGRESQL

Tabela 63. Campos da tabela dor precordial

Campo	Tipo de campo
Descricao_dor	Character varying (200)
Dor_precordial	Character varying (200)

Tabela 64. Campos da tabela colesterol

Campo	Tipo de campo
Descricao_colesterol	Character varying (200)
Colesterol	Character varying (200)

Tabela 65. Campos da tabela glicemia

Campo	Tipo de campo
Descricao_glicemia	Character varying (200)
Glicemia	Character varying (200)

Tabela 66. Campos da tabela eletrocardiograma

Campo	Tipo de campo
Descricao_ecg	Character varying (200)
Ecg	Character varying (200)

Tabela 67. Campos da tabela cateterismo

Campo	Tipo de campo
Descricao_ca	Character varying (200)
Cateterismo	Character varying (200)

Tabela 68. Campos da tabela teste ergométrico

Campo	Tipo de campo
Descricao_depressao	Character varying (200)
Depressao	Character varying (200)

Tabela 69. Campos da tabela

Campo	Tipo de campo
Descricao_induzido	Character varying (200)
Exercicio	Character varying (200)

Tabela 70. Campos da tabela freqüência cardíaca máxima

Campo	Tipo de campo
Descricao_fcm	Character varying (200)
FCmaxima	Character varying (200)

Tabela 71. Campos da tabela pressão arterial

Campo	Tipo de campo
Descricao_pressao	Character varying (200)
Pressao_repouso	Character varying (200)

Tabela 72. Campos da tabela exercício

Campo	Tipo de campo
SlopeST	Character varying (200)
SlopeST	Character varying (200)

APÊNDICE H– GLOSSÁRIO DE TERMOS – CONCEITOS E RELAÇÕES

Tabela 73. Descrição dos conceitos utilizados na ontologia DDC

Conceitos	Descrição
Diagnóstico de Doença Coronariana	Primeira classe para diagnóstico de doença coronariana
Conceito de Diagnóstico de Doença Coronariana	Classe contendo conceitualização de doença coronariana
Diagnósticos	Classe contendo diagnósticos de doença coronariana

Tabela 74. Descrição dos conceitos utilizados na classe de sintomas

Conceitos	Descrição
Sintomas	Classe contendo sintomas para diagnóstico de doença coronariana
Dor Precordial	Classe contendo conceitos e características de dor precordial
Conceitos de Dor Precordial	Definições do Conceito de dor precordial
Características de Dor Precordial	Apresentam algumas características de dor precordial

Tabela 75. Descrição do conceitos utilizados na classe exames

Conceitos	Descrição
Exames	Classe contendo exames para diagnóstico de doença coronariana
Glicemia	Classe contendo conceitos e informações de glicemia
Conceito de Glicemia	Definição do conceito de glicemia
Níveis de Glicemia	Apresentam algumas Informações dos níveis de glicemia
Colesterol	Classe contendo conceitos e informações de colesterol
Conceito de Colesterol	Definição do conceito de colesterol
Classificação de Colesterol	Apresentam Informações sobre a classificação do colesterol
Pressão Arterial	Classe contendo conceitos e estágios de pressão arterial
Conceito Pressão Arterial	Definição do conceito de pressão arterial
Estágios de Pressão Arterial	Algumas Informações de estágios de pressão arterial
Eletrocardiograma	Classe contendo conceitos e informações de eletrocardiograma
Conceito Eletrocardiograma	Definição do conceito de eletrocardiograma
Informações Eletrocardiograma	Apresentam algumas Informações de eletrocardiograma
Teste Ergométrico	Classe contendo conceitos e informações de teste ergométrico
Conceito de Teste Ergométrico	Definição do conceito de teste ergométrico
Informações Teste Ergométrico	Apresentam algumas Informações de teste ergométrico
Cateterismo	Classe contendo conceitos e informações de cateterismo
Conceito de Cateterismo	Definição do conceito de cateterismo
Informações de Cateterismo	Algumas Informações do exame de cateterismo

Tabela 76. Descrição das relações utilizadas na ontologia DDC

Conceito	Relação
Conceitos Diagnósticos de Doença Coronariana	Descrição Definição Fonte
Diagnósticos	Descrição Definição Sexo Fonte
Sintomas	Descrição Definição Observação Fonte
Dor precordial Pressão Alta	Descrição Definição Observação Tipo Fonte
Eletrocardiograma Teste Ergométrico Cateterismo	Descrição Definição Observação Fonte
Glicemia Colesterol	Descrição Definição Observação Valores Fonte

APÊNDICE I – DICIONÁRIO DE CONCEITOS

Tabela 77. Árvore de classificação de conceitos - subconceitos de Diagnóstico de Doença Coronariana

Conceito	Tipo	Hierarquia
Diagnóstico de Doença Coronariana	Abstrato	Conceito Principal
Conceitos Diagnósticos de Doença Coronariana	Concreto	Subconceito de DDC
Diagnósticos	Concreto	Subconceito de DDC
Sintomas	Abstrato	Subconceito de DDC
Exames	Abstrato	Subconceito de DDC

Tabela 78. Árvore de classificação de conceitos - subconceitos de Sintomas

Conceito	Tipo	Hierarquia
Dor Precordial	Abstrato	Subconceito de Sintomas

Tabela 79. Árvore de classificação de conceitos - subconceitos de Dor Precordial

Conceito	Tipo	Hierarquia
Definição de Dor Precordial	Concreto	Subconceito de Dor Precordial
Características de Dor Precordial	Concreto	Subconceito de Dor Precordial

Tabela 80. Árvore de classificação de conceitos - subconceitos de Diagnóstico de Exames

Conceito	Tipo	Hierarquia
Glicemia	Abstrato	Subconceito de Exames
Colesterol	Abstrato	Subconceito de Exames
Pressão Arterial	Abstrato	Subconceito de Exames
Eletrocardiograma	Abstrato	Subconceito de Exames
Teste Ergométrico	Abstrato	Subconceito de Exames
Cateterismo	Abstrato	Subconceito de Exames

Tabela 81. Conceitos Árvore de classificação de conceitos - subconceitos de Glicemia

Conceito	Tipo	Hierarquia
Conceito de Glicemia	Concreto	Subconceito de Glicemia
Níveis de Glicemia	Concreto	Subconceito de Glicemia

Tabela 82. Conceitos Árvore de classificação de conceitos - subconceitos de Colesterol

Conceito	Tipo	Hierarquia
Conceito de colesterol	Concreto	Subconceito de Colesterol
Classificação colesterol	Concreto	Subconceito de Colesterol

Tabela 83. Conceitos Árvore de classificação de conceitos - subconceitos de Pressão Arterial

Conceito	Tipo	Hierarquia
Conceito de pressão arterial	Concreto	Subconceito de Pressão Arterial
Estágios de pressão arterial	Concreto	Subconceito de Pressão Arterial

Tabela 84. Conceitos Árvore de classificação de conceitos - subconceitos de Eletrocardiograma

Conceito	Tipo	Hierarquia
Conceito de eletrocardiograma	Concreto	Subconceito de Eletrocardiograma
Informações de Eletrocardiograma	Concreto	Subconceito de Eletrocardiograma

Tabela 85. Conceitos Árvore de classificação de conceitos - subconceitos de Teste Ergométrico

Conceito	Tipo	Hierarquia
Conceito de teste ergométrico	Concreto	Subconceito de Teste Ergométrico
Informações de teste ergométrico	Concreto	Subconceito de Teste Ergométrico

Tabela 86. Conceitos Árvore de classificação de conceitos - subconceitos de Cateterismo

Conceito	Tipo	Hierarquia
Conceito de Cateterismo	Concreto	Subconceito de Cateterismo
Informações de cateterismo	Concreto	Subconceito de Cateterismo

APÊNDICE J – TABELAS DE ATRIBUTOS DE INSTÂNCIAS

Tabela 87. Instâncias dos atributos utilizados na ontologia DDC

Conceito	Relação	Função
Conceitos Diagnósticos de Doença Coronariana	Descrição	Caractere
	Definição	Caractere
	Fonte	Caractere
Diagnósticos	Descrição	Caractere
	Definição	Caractere
	Sexo	Símbolos, valores a assumir: feminino, masculino ou ambos
	Faixa etária	Caractere
	Fonte	Caractere
Sintomas e o subconceito Dor Precordial. Pressão Arterial	Descrição	Caractere
	Definição	Caractere
	Observação	Caractere
	Tipo	Caractere
	fonte	Caractere
Exames e os subconceitos Eletrocardiograma, Teste Ergométrico e Cateterismo	Descrição	Caractere
	Definição	Caractere
	Observação	Caractere
	Tipo	Caractere
	Fonte	Caractere
Os conceitos Glicemia e Colesterol	Descrição	Caractere
	Definição	Caractere
	Observação	Caractere
	Valores	Caractere
	Fonte	Caractere

APÊNDICE K – TABELA DE INSTÂNCIAS

Tabela 88. Instância do conceito dor precordial – subconceito características de dor precordial

Conceito	**Características de dor precordial**
Relação	**Instância**
Descrição	Classificação Dor Precordial
Tipo	Assintomático
Definição	Não possui dores de origens cardíacas
Observação	Precisa de uma investigação pra descobrir a origem da dor.
Fonte	FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005.

Tabela 89. Instância do conceito de glicemia – subconceito níveis de glicemia

Conceito	**Níveis glicemia**
Relação	**Instância**
Descrição	Glicemia
Definição	Açúcar no sangue. Níveis normais de Glicemia
Valores	Maior ou igual 110 mg/dL
Observação	
Fonte	FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005. 266 p.

Tabela 90.Instância do conceitos de colesterol – subconceito classificação de colesterol

Conceito	**Classificação de colesterol**
Relação	**Instância**
Descrição	Colesterol LDL
Definição	Os lipídios formados pelo LDL, ou colesterol ruim.
Valores	Desejável menor que 100mg/dl
Observação	.
Fonte	FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005.

Tabela 91. Instância do conceitos de colesterol

Conceito	**Colesterol**
Relação	**Instância**
Descrição	Colesterol Total
Definição	Aumentados
Valores	Maior e igual a 240mg/dl
Observação	Total de HDL e LDL juntos.
Fonte	FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005.

Tabela 92. Instância do conceitos de colesterol – subconceito classificação de colesterol

Conceito	**Classificação de colesterol**
Relação	**Instância**
Descrição	Níveis de Colesterol Total
Definição	Nível Desejável
Valores	Menor que 200mg/dl
Observação	Total de HDL e LDL juntos.
Fonte	FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005.

Tabela 93. Instância do conceito de colesterol – subconceito classificação de colesterol

Conceito	**Classificação de colesterol**
Relação	**Instância**
Descrição	Colesterol Total
Definição	Nível Limítrofe
Valores	Entre 200 e 239mg/dl
Observação	
Fonte	FARRET, Jacqueline Faria. **Nutrição e doenças cardiovasculares:** prevenção primária e secundária. São Paulo: Atheneu, 2005.

Tabela 94. Instância do conceito de pressão arterial – subconceito estágios de pressão arterial

Conceito	**Estágios de pressão arterial**
Relação	**Instância**
Descrição	Estágio de Pressão Arterial
Tipo	Normal
Definição	Pressão arterial sistólica menor que 130 mmHg
Observação	Diastólica menor que 85 mmHg
Fonte	NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** Barueri, SP: Manole, 2005.

Tabela 95. Instância do conceito de pressão arterial – subconceito estágios de pressão arterial

Conceito	**Estágios de pressão Arterial**
Relação	**Instância**
Descrição	Estágio de pressão arterial
Tipo	Ótima
Definição	Pressão sistólica menor que 120 mmHg e
Observação	Diastólica menor que 80 mmHg
Fonte	NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** Barueri, SP: Manole, 2005.

Tabela 96. Instância do conceito de pressão arterial – subconceito estágios de pressão arterial

Conceito	**Estágio de Pressão Arterial**
Relação	**Instância**
Descrição	Estágio de pressão arterial
Tipo	Limítrofe
Definição	Pressão sistólica entre 130 mmHg e 139 mmHg
Observação	Diastólica entre 85 mmHg e 89 mmHg
Fonte	NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** Barueri, SP: Manole, 2005.

Tabela 97. Instância do conceito de pressão arterial – subconceito estágios de pressão arterial

Conceito	**Estágio de Pressão Arterial**
Relação	**Instância**
Descrição	Pressão Arterial
Tipo	Hipertensão
Definição	Pressão sistólica maior que 140 mmHg e
Observação	Diastólica maior que 90 mmHg
Fonte	NOBRE, Fernando; SERRANO JÚNIOR, Carlos V. **Tratado de cardiologia SOCESP.** Barueri, SP: Manole, 2005.

Tabela 98. Instância do conceito de teste ergométrico – subconceito informações teste ergométrico

Conceito	**Informações teste ergométrico**
Relação	**Instância**
Descrição	Teste Ergométrico
Definição	Princípio de tentar levar o paciente a atingir sua frequência cardíaca máxima (FCM)
Observação	O calculo para FCM subtraindo-se 220 da idade
Fonte	MICHIELIN, Francisco. **Doenças do Coração.** São Paulo: Robe Editorial, 2003.

Tabela 99. Instância do conceito de teste ergométrico – subconceito informações teste ergométrico

Conceito	**Informações teste ergométrico**
Relação	**Instância**
Descrição	Informações Teste Ergométrico
Definição	Neste teste são analisados as ondas e segmentos: P, PR, PRS, J, R, Q, S, QRS,ST,QT, T e U
Observação	O segmento ST
Fonte	MICHIELIN, Francisco**. Doenças do Coração**. São Paulo: Robe Editorial, 2003.

Tabela 100. Instância do conceito de cateterismo – subconceito informações de caterterismo

Conceito	**Informações Cateterismo**
Relação	**Instância**
Descrição	Informações cateterismo
Definição	Avalia o grau de estreitamento das artérias, percentual maior que 50 % há comprometimento da artéria.
Observação	Quando o paciente não tem artérias comprometidas diz-se que não possui doença coronariana.
Fonte	MICHIELIN, Francisco. **Doenças do Coração**. São Paulo: Robe Editorial, 2003.