

UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA MULTIDISCIPLINAR DE PÓS-GRADUAÇÃO
EM MODELAGEM COIMPUTACIONAL DE CONHECIMENTO

JAILTON CARDOSO DA CRUZ

MODELO DE APOIO AO ESTUDO DE PACIENTES EM ONCOLOGIA
PEDIÁTRICA UTILIZANDO
RACIOCÍNIO BASEADO EM CASOS E MINERAÇÃO DE DADOS

UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA MULTIDISCIPLINAR DE PÓS-GRADUAÇÃO
EM MODELAGEM COIMPUTACIONAL DE CONHECIMENTO

JAILTON CARDOSO DA CRUZ

MODELO DE APOIO AO ESTUDO DE PACIENTES EM
ONCOLOGIA PEDIÁTRICA UTILIZANDO
RACIOCÍNIO BASEADO EM CASOS E MINERAÇÃO DE DADOS

Projeto de Pesquisa apresentado como requisito para obtenção do título de Mestre em Modelagem Computacional de Conhecimento.

Orientador: Prof. Dr. Fábio Paraguaçu Duarte da Costa.

Área de Concentração: Modelagem Computacional de Conhecimento

Maceió – AL

2014

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecário: Valter dos Santos Andrade

C957m Cruz, Jailton Cardoso da.
Modelo de apoio ao estudo de pacientes em oncologia pediátrica
Utilizando raciocínio baseado em casos e mineração de dados / Jailton
Cardoso da Cruz. – 2014.
116 f. : il.

Orientador: Fábio Paraguaçu Duarte da Costa.
Dissertação (Mestrado em Modelagem Computacional de
Conhecimento) – Universidade Federal de Alagoas. Programa de Pós
Graduação em Modelagem Computacional de Conhecimento. Maceió,
2014.

Bibliografia: f. 95-104.
Apêndices: f. 105-113.
Anexos: f. 114-118.

1. Mineração de dados (Computação). 2. Documento eletrônico.
3. Prontuários médico-hospitalares. 4. Oncologia pediátrica - Pacientes.
I. Título.

CDU: 004.5

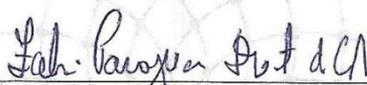


UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
Programa de Pós-Graduação em Modelagem Computacional de Conhecimento
Avenida Lourival Melo Mota, Km 14, Bloco 09, Cidade Universitária
CEP 57.072-900 – Maceió – AL – Brasil
Telefone: (082) 3214-1364



Membros da Comissão Julgadora da Dissertação de Mestrado de Jailton Cardoso da Cruz, intitulada: “Modelo de apoio ao estudo de pacientes em oncologia pediátrica utilizando raciocínio baseado em casos e mineração de dados”, apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas em 14 de abril de 2014, às 17h00min, na sala de aula do Mestrado em Modelagem Computacional de Conhecimento da UFAL.

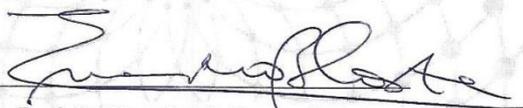
COMISSÃO JULGADORA



Prof. Dr. Fábio Paraguaçu Duarte da Costa

UFAL – Instituto de Computação

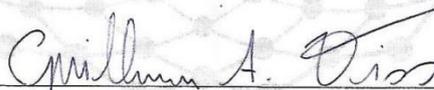
Orientador



Prof. Dr. Evandro de Barros Costa

UFAL – Instituto de Computação

Examinador



Prof. Dr. Guilherme Ataíde Dias

UFPB – Departamento de Ciência da Informação

Examinador

Maceió, abril de 2014.

DEDICATÓRIA

Dedico esta dissertação aos meus pais José Cardoso e Ana Claudino (in memoriam) à minha tia e madrinha Adalgisa Cardoso, à minha esposa Silvana, à minha filha Raíssa e aos meus irmãos José Cardoso, Ana Lúcia, Mércia Maria, Josemar e Jamerson.

AGRADECIMENTOS

A Deus por me dar serenidade e força em todos os momentos, felizes ou não, e por me proporcionar este momento de realização.

Ao meu orientador e amigo Professor Dr. Fábio Paraguaçu pela motivação e pelo apoio durante todo o decorrer do curso e pela confiança em mim depositada.

Ao meu amigo e professor Dr. Evandro Barros pelo incentivo e confiança.

Ao meu grande incentivador que me conduziu nos primeiros passos na carreira de analista de sistemas, Ednamai Nóbrega.

Aos meus amigos que me apoiaram, direta ou indiretamente, ao longo da minha trajetória como aluno deste programa de mestrado em especial a Luiz Frederico que me ajudou nos estudos de mineração de dados, João Henrique, Leonardo Medeiros, Breno Jacinto, Stênio Fernandes, Edel Alexandre Pontes e Eunice Palmeira pela força dada em todos os momentos.

Aos amigos e irmãos de coração Marconi e Wamberto Chaves, André Novaes, Carlos Magno, João Batista, Cezário Lima e Ricardo Amaral, que estão sempre próximos mesmo que o tempo e a distância digam não.

Aos companheiros da Santa Casa de Misericórdia de Maceió e da Macrocoop em especial ao Gerente de Tecnologia da Informação, Marcus Aurélio.

À minha família, em especial aos meus irmãos, José, Lúcia, Mércia, Josemar e Jamerson que, espelhados na figura de nossos pais, José e Ana, estão sempre a postos para apoiar e acolher a todos.

Aos meus cunhados, José Cabral e George Buriti, que, no momento da partida de meus pais, acolheram-me em suas casas como filho e que foram de fundamental importância no meu crescimento como pessoa.

Aos meus sogros, Silas e Genilda, que, como verdadeiros pais, guiam-me longe da minha família, e a toda a família Machado, pela atenção e pelo acolhimento.

À minha cunhada e amiga, Sandra Cristina, e a amiga Luciana Andrade, pelo apoio incondicional.

Por fim, à minha esposa, Silvana, e à minha filha, Raíssa, pela compreensão nos meus momentos de ausência, mas que, com amor e carinho, incentivam-me e fortalecem-me na busca da harmonia familiar. E à nossa cadelinha Juli, pela companhia durante as longas madrugadas de estudos.

SUMÁRIO

Resumo	i
Abstract.....	ii
Lista de Figuras	iii
Lista de Quadros.....	iv
Lista de Tabelas	v
1 Introdução.....	12
1.1 Contextualização	12
1.2 Objetivos	14
1.3 Metodologia	15
1.4 Contribuições	15
1.5 Estrutura do Trabalho	16
2 Informática e a Medicina.....	18
2.1 Introdução.....	18
2.2 A informática e a Medicina	18
2.3 O Prontuário Médico Eletrônico	23
2.4 O sistema MV	26
2.5 O Raciocínio Médico e a Tomada de Decisão	28
2.5.1 Representação do Raciocínio	28
2.6 O Domínio do Câncer Pediátrico	30
2.7 O Câncer Pediátrico no Brasil.....	32
3 O Modelo de Referência CRISP-DM.....	34
3.1 Introdução	34
3.2 O Ciclo de Vida do CRISP-DM	34
3.3 Compreensão do Negócio	35
3.4 Compreensão dos Dados.....	35
3.5 Preparação dos Dados.....	36
3.6 Modelagem	37
3.7 Avaliação	38
3.8 Implementação	38
4 Raciocínio Baseado em Casos.....	39
4.1 Introdução.....	39
4.2 Raciocínio Baseado em Casos	39
4.2.1 Histórico	40

4.2.2	Vantagens do Uso de CBR	42
4.2.3	A Abordagem RBC	43
4.2.4	Sistema Baseado em Casos (<i>Case-based System</i>)	45
4.2.5	Ciclo de Vida de um CBR	46
4.2.6	Aquisição e Representação de Casos	48
4.2.7	Recuperação de Casos.....	51
4.2.8	Reuso.....	53
4.2.9	Retenção.....	54
4.3	Raciocínio Baseado em Casos Aplicado a Saúde.....	55
4.3.1	CBR como Apoio a Tomada de Decisão Médica.....	56
4.3.2	CBR como ferramenta de Aprendizado.....	57
5	Mineração de Dados.....	59
5.1	Introdução.....	59
5.2	Mineração de Dados	59
5.2.1	Métodos	60
5.3	Os Componentes de um Algoritmo de Mineração de Dados.....	62
5.4	Minerando Padrões de Sequências.....	62
5.4.1	Declaração do Problema	62
5.4.2	Algoritmo Apriori.....	63
5.5	Integração de CBR e Mineração de Dados.....	66
6	O Modelo de Apoio a Prescrições Médicas (MAPMED)	69
6.1	Introdução.....	69
6.2	Caracterização do Caso	69
6.3	Modelo Formal.....	71
6.4	Visão Geral da Arquitetura.....	73
6.4.1	Módulo Atendimento.....	74
6.4.2	Modulo KDD	77
6.4.3	Processo de Mineração de Dados	79
7	Implementação do Estudo de Caso	81
7.1	Introdução.....	81
7.2	Entendimento do Negócio	81
7.3	Base de Casos.....	82
7.4	Conhecimento dos Dados	82
7.4.1	Obtenção dos Dados	82

7.4.2	Mascaramento dos códigos dos pacientes	83
7.4.3	Escolha das Palavras Chaves	83
7.5	Tratamento da Base de Dados	87
7.5.1	Criação da Visão Materializada	87
7.5.2	Limpeza da Base de Dados	87
7.5.3	Atualização da base de Casos.....	88
7.5.4	Criação das Visões Utilizadas no Processo de Mineração.....	89
7.6	Modelo de Mineração	89
8	Considerações Finais.....	93
8.1	Conclusão	93
8.2	Trabalhos Futuros.....	94
	Referências	95
	Apêndices	105
	Apêndice A – Descrição das tabelas.....	106
	Apêndice B – Queryes de Extração dos Dados	108
	Extração dos dados dos pacientes	108
	Extração dos dados dos atendimentos	108
	Extração dos dados das prescrições.....	108
	Extração dos dados dos itens da prescrição	109
	Extração dos dados dos tipos de prescrição	110
	Apêndice C – Visão Materializada PRESCRICAO	111
	Apêndice D – Procedure de Carga da Base de Casos.....	112
	Apêndice E – Criação da Visões VW_PRESCRICAO E VW_PESCRICAO_DSC ...	113
	Anexos	114
	Anexo A – Resultados para carcinoma de nasofaringe	115
	Anexo B – Resultados para Osteossarcoma	116
	Anexo C – Resultados para Tumor de Wilms	117
	Anexo D – Resultados para Medicamentos Tumor de Wilms	118

RESUMO

Este trabalho tem como objetivo propor um Modelo de Recomendação de itens de prescrição para Oncologia Pediátrica baseado na extração de dados do Prontuário Eletrônico do Paciente. Esses dados são utilizados como casos indexados, para auxiliar os prestadores de serviço médico baseados na similaridade de prescrições, de acordo com o histórico do paciente. Do ponto de vista do apoio a educação médica, a modelagem objetiva apoiar o estudante ou o profissional de saúde no entendimento do processo de tomada de decisão durante a fase de prescrição de itens de tratamento médico oncológico, como, por exemplo: medicamentos, exames de laboratório ou de imagens, dieta, gases, cuidados, quimioterapia, radioterapia. Para o desenvolvimento do modelo, utilizou-se a abordagem de Raciocínio Baseado em Casos (RBC), através da representação de uma base de casos de prescrição médica, indexada por seus itens de tratamento. Durante a fase de recuperação de casos, utilizou-se a ferramenta de Mineração de Dados aplicando-se o modelo de regra de associação, em conjunto com o algoritmo “apriori” visando a obtenção da similaridade entre casos. Para a atualização da base de casos, foi desenvolvido um procedimento de banco de dados para execução do processo de Extração, Transformação e Carga da base de dados. O modelo desenvolvido foi aplicado na base de dados do Prontuário Eletrônico do Paciente da Santa Casa de Misericórdia de Maceió, baseado no sistema de gestão hospitalar MV Sistemas, implantado na unidade desde 2005. Para a apresentação dos resultados, utilizou-se a ferramenta Oracle Data Miner, que possibilitou o acesso ao banco de dados e a análise dos casos selecionados pela identificação de palavras chaves contidas na evolução do estado clínico do paciente. A aplicação dos experimentos permitiu validar a ocorrência de aplicação conjunta de itens de tratamento de acordo com as palavras chaves, o que pode ser utilizado como elemento para o processo de tomada de decisão médica e tutoria.

Palavras-chave: Mineração de Dados. Raciocínio Baseado em Casos. Oncologia Pediátrica. Prontuário Eletrônico do Paciente. Tutoria., Apoio a Decisão.

ABSTRACT

This work aims to propose a recommendation model of prescription items for Pediatric Oncology based on data extraction from Electronic Patient Record. These data are used as indexed cases to aid providers of medical service based on the similarity of prescriptions, according to the patient's history. From the viewpoint of aid medical education, the modeling objective support the student or health professional in understanding the decision-making process during the prescription items oncological medical treatment, for example drugs, laboratory exams or images exams, diet, gases, care, chemotherapy, radiotherapy. To develop the model, was used the approach of Case Based Reasoning (CBR), through the representation of prescription medical base-case, indexed by their treatment items. During the recovery phase of cases, we used the tool Data Mining by applying the model of association rule, together with the algorithm "apriori" for obtaining the similarity between cases. To update the case base, a procedure database for performing the process of Extraction, Transformation and Load of the database was developed. The developed model was applied in the database of the Electronic Patient Record of the Santa Casa de Misericordia de Maceio, based on Hospital Management Systems "MV Sistemas", deployed in the unit since 2005. For the presentation of results, was used the Oracle Data Miner tool, which allowed access to the database and analysis of selected cases by identifying key words contained in the evolution of the clinical condition of the patient. The application of the experiments validate the occurrence of allowed combined application of items of treatment according to the keywords, which can be used as input in the process of making medical decision and tutoring.

Keywords : Data Mining , Case Based Reasoning , Pediatric Oncology , Health Record, Mentoring, Decision Support

LISTA DE FIGURAS

Figura 1 – Menu Principal dos Módulos do Sistema MV	27
Figura 2 - Modelo Referência CRISP-DM	34
Figura 3 - Ciclo de vida CBR.....	46
Figura 4 - Uma decomposição do método de tarefas de um CBR	47
Figura 5 - Processos envolvidos na recuperação de casos	53
Figura 6 - Passos que Compõem o Processo de KDD	59
Figura 7 - Funcionamento de Algoritmos Apriori com suporte mínimo de 50%	64
Figura 8 - <i>Itens</i> do Pedido.....	65
Figura 9 - Aplicação do Algoritmo Apriori no <i>itens</i> da fig. 8.....	65
Figura 10 - Apresentação da relação <i>If ... Then</i> da regra 8.....	66
Figura 11 - Evolução de paciente com quadro de carcinoma de nasofaringe	70
Figura 12 - Evolução de paciente com quadro de carcinoma de linfoceloma	71
Figura 13 - Arquitetura do Sistema	74
Figura 14 - Caso de Uso Atendimento	75
Figura 15 - Interface Apresentando Dados do Atendimento	75
Figura 16 - Interface de Entrada de Dados dos Itens Prescritos	76
Figura 17 – Modelo de Dados das tabelas envolvidas no processo.....	76
Figura 18 - Processo ETL	77
Figura 19 - Comparação entre itens prescritos em dois atendimentos	80
Figura 20 - Lista de Palavras Medicamentos	86
Figura 21 - Lista de Palavras Tumor	86
Figura 22 - Dados do job de agendamento	88
Figura 23 - Fluxo AssocPorDescr	90
Figura 24 - Fluxo AssocPorCod	90
Figura 25 - Parâmetros de Execução do fluxo	91
Figura 26 - Histograma dos Dados da visão VW_PRESCRICAO	91
Figura 27 - Resultado para Carcinoma de Nasofaringe.....	115
Figura 28 - Resultado para Osteossarcoma.....	116
Figura 29 - Resultado para Tumor de Wilms	117
Figura 30 - Resultado para Medicamentos em Caso de Tumor de Wilms	118

LISTA DE QUADROS

Quadro 1 - Quantitativo de dados gerados.....	82
Quadro 2 - Itens prescritos para a evolução com quadro de OSTEOSSARCOMA.....	84
Quadro 3 - Itens prescritos para a evolução com quadro de OSTEOSSARCOMA.....	84

LISTA DE TABELAS

Tabela 1- Aplicação da CBR na medicina, Holt (2005).....	42
Tabela 2 - Problemas do paradigma baseado em regras, adaptado de Slade (1991).....	44

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

A utilização dos sistemas de informação nas organizações, nos últimos anos, vem experimentando um grande crescimento. Neste intervalo de tempo, tais sistemas passaram do simples papel de suporte para se tornarem uma ferramenta estratégica nas empresas. Nas décadas de 70, 80 os sistemas foram desenvolvidos para dar agilidade aos processos organizacionais tais como contabilidade, folha de pagamento, compras, faturamento dentre outros, reduzindo o tempo de processamento e melhorando o fluxo da informação. Nos anos 90, com o barateamento dos custos de armazenamento, o aumento da velocidade de transmissão de dados, a utilização de sistemas de banco de dados e a internet, o termo processamento de dados deu lugar a algo mais abrangente, e daí os dados foram lapidados e transformados em informação e em uma visão sistêmica em conhecimento. Sendo a organização um sistema aberto, as mudanças ocorridas tanto no ambiente externo quanto no ambiente interno, este banco de conhecimento é incorporado aos ativos organizacionais, tornando-se uma base sólida para o processo de gestão da estratégia de negócio. A grande quantidade de dados gerados pelos sistemas que atendem às camadas operacionais das organizações transformou-se em um verdadeiro tesouro que pode ser explorado de diversas formas.

Com objetivo de alcançar a excelência organizacional bem como proporcionar maior ganho de eficiência nos processos de negócios e fluxo de informação, um dos focos de maior atenção da área de sistema de informação é criar melhores formas de atender às camadas superiores das organizações, tais como gerências e diretorias, através de sistemas de apoio a tomada de decisão a fim de obter maior vantagem competitiva.

Neste contexto, na área de saúde, o desenvolvimento de sistemas informatizados tem sido de fundamental importância para a melhoria e evolução da medicina e, dentre as diversas categorias de sistemas, encontra-se os sistemas de apoio à decisão médica. De acordo com Meunier (2004 apud Bouchon-Meunier, 2000), uma das principais motivações para o uso de sistemas automatizados de apoio à decisão em medicina é a esperança de que eles possam ajudar a superar algumas das dificuldades intrínsecas à própria natureza do conhecimento médico, que é incerto, impreciso e

incompleto. Tais sistemas foram desenvolvidos com o objetivo de auxiliar profissionais de saúde a tomar decisões que influenciem diretamente no diagnóstico clínico ou no tratamento do paciente haja vista que, na atividade clínica, tomar decisão envolve análise profunda dos riscos envolvidos de acordo com a o tratamento escolhido. Nesta situação, uma análise precisa dos resultados e possíveis consequências são fundamentais para atingir a melhor decisão clínica possível. Para obter diagnóstico preciso, profissionais devem estar familiarizados com os diversos padrões que uma determinada patologia apresenta, reconhecê-las e assim diagnosticá-las. Profissionais iniciantes, estudantes, não especialistas na patologia apresentada ou mesmo profissionais experientes também necessitam outras fontes de análise, tais como livros e artigos, a fim de estudar diagnósticos passados com objetivo de obter uma conclusão precisa. Sistemas de apoio à decisão médica apresentam-se como uma ferramenta que pode auxiliar tais profissionais a definir um diagnóstico com rapidez e certa precisão.

No sentido educacional e de treinamento, informações e conhecimento armazenados em bancos de dados tomam importância na medida em que de acordo com (Schmidt 2000, apud Spyropoulos 1999a, 1999b, 2000);

"A otimização dos processos de tomada de decisão em Medicina, exige uma formação contínua. Na formação profissional, contribui para a promoção da pesquisa interdisciplinar, abordando, através da manipulação de casos específicos, de uma forma eficaz, o problema na comunicação intra-especialidade. Por outro lado, o hipertexto e cursos multimídia, estão ganhando importância em Educação Médica e eles podem ser usados a fim de oferecer treinamento orientado à prática clínica. Cursos baseados na *web* e outras alternativas tecnológicas emergentes prometem alcançar vários grupos, oferecendo-lhes serviços de educação continuada. Estes grupos podem incluir também aqueles que já estão engajados no trabalho profissional, como médicos, enfermeiros, engenheiros, físicos, técnicos etc".

Neste contexto de apoio à decisão médica e compartilhamento do conhecimento formal, armazenado em banco de dados, como apoio ao estudo em medicina, este trabalho utiliza a abordagem de Raciocínio Baseado e Casos (RBC), cujo princípio é resolver problemas baseado em casos similares ocorridos no passado onde, um *caso* consiste na descrição de um problema (por exemplo, um sintoma) e uma solução (por exemplo um diagnóstico ou terapia). Casos são armazenados em uma *base de casos*. Para a resolução de um problema atual, a noção de similaridade entre os problemas é usada para construção do raciocínio e *acesso* aos casos similares da base de casos. Em combinação com a abordagem de RBC, utilizamos a tecnologia de mineração de dados cujo objetivo é acessar a base de dados e aplicar o processo de descoberta de conhecimento em base de dados (KDD – *Knowledge Discovery in Database*) para

acessar a base de casos e, a partir da aplicação de medidas de similaridade, retornar os casos similares. Neste sentido, este trabalho propõe a utilização de Raciocínio Baseado em Casos (RBC) em combinação com técnicas de Mineração de Dados no apoio à prescrição médica em casos de Oncologia Pediátrica utilizando uma base de dados de prontuário eletrônico médico armazenado em banco de dados Oracle.

A ideia do desenvolvimento deste trabalho surgiu durante o processo de implantação do projeto de *Business Intelligence* (BI) na Santa Casa de Misericórdia de Maceió, no qual, durante o desenvolvimento do trabalho, identificou-se a possibilidade de combinar Raciocínio Baseado em Casos e Mineração de Dados para, a partir histórico registrado no Prontuário Eletrônico do Paciente (PEP), ajudar na resolução de alguns problemas no contexto de medicina oncológica tais como:

- Apoio ao aprendizado dos prestadores no tratamento de pacientes com quadro clínico semelhantes;
- Apresentar um perfil farmacológico de pacientes com quadro clínico semelhante;
- Apoio à decisão no processo clínico para prescrição de novos tratamentos utilizando casos.

1.2 OBJETIVOS

O objetivo deste trabalho é propor um modelo representação, indexação e recuperação de casos a partir de uma base de dados armazenada em um Prontuário Eletrônico do Paciente – PEP, de pacientes com quadro de Oncologia Pediátrica, utilizando algoritmos de mineração de dados, a fim de dar apoio ao processo de tomada de decisão médica e dar suporte ao ensino na área de medicina.

Os objetivos específicos são:

- Representação de uma base de casos a partir dos dados do Prontuário Eletrônico do Paciente;
- Indexar os dados recuperados;
- Aplicar o algoritmo de mineração de dados para a recuperação dos dados;
- Escolher uma métrica de similaridade;
- Recuperar os dados da base de casos;

- Apresentar os dados para o processo de tomada de decisão utilizando as medidas de suporte e confiança.

1.3 METODOLOGIA

A metodologia utilizada foi um estudo de caso de pacientes com quadro de Oncologia Pediátrica.

A coleta de dados foi feita a partir de consultas SQL à base de dados do PEP da Santa Casa de Maceió, de pacientes com quadro clínico de Oncologia Pediátrica.

Após esta etapa, foi criada uma base de casos contendo visões de banco de dados (*views*) do prontuário eletrônico com os dados necessário para o processo de indexação.

Feita a revisão bibliográfica, estudou-se a metodologia do Raciocínio Baseado em Casos e a técnica de Mineração de Dados, bem como a integração entre a teoria, sua aplicação em Prontuários Médicos Eletrônicos e a utilização na tutoria presencial ou a distância e no processo de tomada de decisão médica.

Por fim, aplicaram-se as medidas de similaridade para a recuperação dos casos indexados, utilizando a ferramenta de mineração de dados Oracle Data Miner.

1.4 CONTRIBUIÇÕES

No contexto do objetivo do trabalho podemos enumerar algumas contribuições, tais como:

- Aquisição do conhecimento – o modelo de extração de casos do Prontuário Eletrônico do Paciente para o processo de coleta de dados;
- Representação e indexação do conhecimento – a partir do conhecimento da base de casos do PEP, que contém dados estruturados e não estruturados, criou-se uma representação contextualizada de casos.
- Tratamento de dados inconsistentes ou “sujos” que criam grandes dificuldades para a atividade de representação e indexação dos casos.
- Recuperação dos casos – utilização de regras de similaridade e mineração de dados para recuperação dos casos.

- Atualização dos casos – criação de agendamento automático e de programa de base de dados (procedimentos e funções) para atualização da base de casos.
- Criação do modelo de recuperação dos casos.

1.5 ESTRUTURA DO TRABALHO

Este trabalho foi desenvolvido em nove capítulos, que permitem o embasamento teórico utilizado para o desenvolvimento da pesquisa, bem como a abordagem para o desenvolvimento do projeto e sua implementação.

No capítulo 1, é feita uma introdução, onde são apresentados os objetivos (geral e específico), a contribuição desejada e a estrutura do trabalho.

O capítulo 2 apresenta a utilização da informática na medicina, aspectos do Prontuário Médico Eletrônico, o processo de tomada de decisão em medicina, que vem servir de base para a construção de sistemas RBC para a área de saúde e, neste cenário, o quadro da Oncologia Pediátrica no Brasil.

O capítulo 3 apresenta um modelo de referência denominado CRISP-DM, que, baseado em um ciclo de vida para o desenvolvimento de sistemas de mineração de dados, apresenta boas práticas para dar mais eficiência ao processo de descoberta de conhecimento.

O capítulo 4 apresenta a abordagem do Raciocínio Baseado em Casos, seu histórico e evolução, vantagens e desvantagens de sua utilização, bem como seu ciclo de vida.

O capítulo 5 faz um estudo da mineração de dados a fim de contextualizar a utilização das técnicas de mineração de dados na recuperação e na indexação de casos e a aplicação dos métodos de mineração para apresentação dos resultados.

No capítulo 6, todo o processo utilizado para a recuperação de casos é apresentado, através de um modelo de apoio a prescrições médicas, utilizando mineração de dados.

O capítulo 7 descreve a implementação do modelo do capítulo 7 através da análise dos dados, desenvolvimento de consultas ao banco de dados e o processo de mineração com a utilização do Oracle Data Miner.

O capítulo 8 mostra as conclusões e trabalhos futuros.

Nos anexos, apresentamos alguns experimentos aplicando o modelo de mineração de dados com alguns tipos de oncologias pediátricas.

2 INFORMÁTICA E A MEDICINA

2.1 INTRODUÇÃO

A informática médica é uma área com importância crescente em todo mundo. O domínio da área da informática médica corresponde a intercessão das áreas das ciências da saúde com as áreas da ciência da computação, sistemas de informação e tecnologia da informação cujo objetivo visa armazenar, recuperar e analisar os dados, informações e conhecimentos biomédicos para o apoio a melhoria dos atendimento ao paciente e apoio ao processo de tomada de decisão médica.

Embora as expressões informática médica e informática em saúde sejam usadas como sinônimo a SBIS, Sociedade Brasileira de Informática em Saúde, adota o termo “em Saúde”, haja vista que, as atividades de relativas à saúde abrangem não só a medicina, mas também a veterinária, odontologia, enfermagem, nutrição, fisioterapia e dentre outras.

A área de informática em saúde engloba diversas especialidades que de acordo com a SBIS (2013) enfatiza as seguintes áreas de atuação: Sistemas de Informação em Saúde, Prontuário Eletrônico do Paciente, Telemedicina, Sistemas de Apoio a Decisão, Processamento de Sinais Biológicos, Processamento de Imagens Médicas, Padronização da Informação em Saúde e Internet em Saúde.

2.2 A INFORMÁTICA E A MEDICINA

Nas últimas décadas, todas as áreas de atividade econômica, sejam elas nas áreas de indústria, comércio, financeiro ou governamental vem experimentando uma crescente utilização das tecnologias de informação e comunicação. Na área de medicina esta tendência também se torna evidente com a utilização dos sistemas informatizados, redes de computadores, internet e sistemas de apoio à decisão, tornando-se um elemento estratégico no armazenamento, distribuição, compartilhamento e análise dos dados dos pacientes não só pelo corpo médico, mas também pela administração hospitalar e parentes, bem como para dar subsídios às políticas públicas na área de saúde, além de abranger o processamento de sinais biológicos e imagens, telemedicina, treinamento médico, acesso a base de dados de saúde dentre outras aplicações. Por outro lado a utilização da tecnologia da informação torna-se importante no que diz respeito à identificação de tratamentos inadequados,

desnecessários ou deficientes, tratamentos incorretos ou ineficientes bem como as questões de custos elevados e desconformidade com as questões éticas e legais.

De acordo com Frank (2006), algumas questões sobre informação clínica devem ser observadas, quais sejam:

- Quanto à manutenção dos registros
 - Quais registros manter?
 - Em que formato?
 - Quais dados informar e quando?
 - Como armazenar os registros e por quanto tempo?
 - Com quem compartilhar os registros?
- Como utilizar as informações contidas nos registros?
 - Para gerenciar os pacientes?
 - Para auditar e melhorar o meu serviço?
 - Para apoiar as minha pesquisa?
 - Para alimentar outro sistema de informação?
- Como se comunicar com os meus colegas e pacientes
 - Face a face?
 - Em *papers*?
 - Usando a internet?
- Fontes de conhecimento clínico
 - Quais as fontes de conhecimento, e como selecioná-las?
 - Como usar essas fontes para responder às minhas próprias perguntas e de minha equipe?
 - Como manter conhecimentos e habilidades?
 - Como usar o conhecimento para melhorar as minhas práticas e de minha equipe?

Estes questionamentos demonstram a importância da utilização racional, planejada e estruturada dos sistemas de informação tanto no que diz respeito ao armazenamento nos prontuários eletrônicos do paciente, mas a distribuição e utilização destes dados com objetivo de possibilitar maior conhecimento, não só das informações clínicas, mas também informações gerenciais.

Segundo Willian (2009, p28), a tecnologia da informação em saúde pode ser caracterizada em quatro categorias:

- Automação - Automação é o uso da TI (Tecnologia da Informação) para executar tarefas que podem ser repetidas com poucas modificações. Os exemplos incluem código barras, administração de medicamentos, geração de resultados laboratoriais e emissão de faturas para pagamento;

- Conectividade - Conectividade começa com a infraestrutura física, garantindo conexões eletrônicas básicas entre várias instalações físicas para que os dados possam ser transmitidos eletronicamente. Os exemplos podem incluir linhas de fibra de alta velocidade e capacidades de roteamento de toda a planta física, redes de área ampla, e a implantação de infraestrutura *wireless*. A conectividade inclui interfaces que mapeiam dados de um sistema para outro.

- Apoio à decisão - Apoio à decisão envolve o uso de aplicativos baseados em TI para fornecer informações em um alto nível conceitual para os médicos com objetivo de facilitar ou melhorar o processo de tomada de decisões. Por exemplo, o apoio à decisão pode incluir simples alertas baseados em regras, como lembretes para os médicos sobre possíveis interações medicamentosas quando os pedidos de medicamentos são introduzidos. Apoio à decisão também pode envolver a apresentação de informações aos prestadores de serviços médicos, de forma a torná-lo mais fácil para eles saber como direcionar a sua atenção através de um *dashboard* ou através de eventos sentinela que mostram alertas em tempo real a ocorrência de situações de exceção. Finalmente, o apoio à decisão também pode se referir ao apoio à decisão estatística e heurística, refletindo uma síntese inteligente de informações sobre o paciente, as informações do ambiente de cuidados, por exemplo, um sistema de apoio à decisão pode recomendar um antibiótico em particular com base na condição do paciente e uma base de dados da recente sensibilidade dos microrganismos a diferentes antibióticos em seu hospital.

- Recursos de mineração de dados - técnicas de mineração de dados usam métodos de descoberta de conhecimento para analisar vários conjuntos de dados semelhantes, ou diferentes, a fim de reconhecer as relações sejam elas conhecidas ou desconhecidas. A técnica de Mineração de Dados, aplicada nos bancos de dados dos Prontuários Eletrônicos dos Pacientes bem como nos sistemas de informação de saúde pública vem sendo utilizada para descobrir relacionamentos entre estes dados para que possa ser feita uma previsão de tendências futuras baseadas no passado. Mineração de dados fornece muitos dos insumos necessários para o apoio à decisão.

Desenvolver, implantar e manter sistemas de informação na área de medicina será um dos grandes desafios para a comunidade de TI nos próximos anos. Segundo Willian (2009), tais mudanças podem ser vistas de forma igualitária a partir de dois aspectos: a mudança evolucionária e a mudança radical. Neste contexto, a mudança evolutiva refere-se à melhora contínua e iterativa dos processos existentes, sustentados por longos períodos de tempo, que não dependem fortemente de novas capacidades tecnológicas. Por outro lado, a mudança radical significa novas maneiras de olhar para os problemas de saúde e novas formas revolucionárias de tratar esses problemas tais como gravação automática de dados, o uso de novos sensores, mineração de dados e técnicas de visualização e o desenvolvimento em genoma. De acordo com os aspectos mencionados alguns princípios de sucesso podem ser citados.

1. De acordo com as mudanças evolucionárias.

Princípio 1 – Concentre-se em melhorias no atendimento.

Tecnologia é secundário

O princípio mais importante para orientar a mudança evolutiva na área da saúde é concentrar esforços em alcançar as melhorias desejadas na área da saúde, em vez de sobre a adoção da Ti em cuidados de saúde por si só.

Princípio 2 – Procure ganho incremental de Esforço Incremental

Programas podem ser estruturados de modo que pequenos investimentos possam render sucesso visível.

Princípio 3 – Registro de dados disponíveis para que possam ser usados para o cuidado, melhoria de processos e pesquisa

Melhoria sistemática dos cuidados de saúde é orientado a dados.

Princípio 4 – Projeto considerando fatores humanos e organizacionais

Projeto centrado no ser humano diminuem as barreiras para sua utilização.

Princípio 5 – Apoiar as funções cognitivas de todos os envolvidos, incluindo profissionais de saúde, pacientes e suas famílias

Investir em cuidados de saúde que pode suportar as funções cognitivas dos indivíduos e organizações como forma iterativa adaptar papéis e processos de trabalho.

2. De acordo com as mudanças radicais.

Princípio 6: Arquiteto de Informação e Sistemas de fluxo de trabalho para acomodar mudanças turbulentas

A infraestrutura de TI deve ser projetada para suportar as necessidades atuais e acomodar as mudanças futuras.

Princípio 7: Arquivo de dados para posterior reinterpretação

Avanços na ciência da computação e disciplinas relacionadas levarão a novas maneiras de extrair conhecimento significativo e útil a partir de repositórios de dados existentes, permitindo reanálise dos dados pré-existentes para revelar medicamente relações significativas e correlações que são presentemente desconhecidos. Esta grande quantidade de dados poderá direcionar, em termos médicos, pesquisas orientadas na aprendizagem de máquina, mineração de dados, apresentação de dados, e outras atividades.

Princípio 8: Pesquisar e desenvolver tecnologias que que identifiquem e eliminem processos de trabalho ineficazes

Automação e melhoria dos processos de trabalho a fim de eliminar os erros e seus efeitos.

Princípio 9: Pesquisar e desenvolver tecnologias que esclarecer o contexto de dados

As organizações devem buscar e desenvolver tecnologias que apresentam novas informações no contexto de outras informações disponíveis sobre o paciente e conhecimento biomédico relevante.

Quanto a utilização dos dados, um problema recorrente é a resistência a utilização das tecnologias por parte dos médicos e que segundo Galvão (2009),

Há uma apreensão por parte de vários gestores e profissionais de saúde em compreender os dados e em utilizar a informação e conhecimento das bases de dados da saúde para promover a gestão da informação e qualidade de cuidados. Isso provavelmente acontece em decorrência de um ritmo alucinante de geração de dados, o que produz uma incapacidade natural no ser humano de explorar, extrair e interpretar estes dados para obter conhecimento dessas bases.

Esta resistência e falta de conhecimento do conteúdo da base de dados tem desafiado os profissionais de TI para a democratização, popularização e utilização dos sistemas tanto em nível operacional, quanto em nível tático e estratégico.

2.3 O PRONTUÁRIO MÉDICO ELETRÔNICO

O prontuário do paciente de acordo com Eduardo (2003) ou, mais frequentemente chamado prontuário médico, é um elemento crucial no atendimento à saúde dos indivíduos, devendo reunir a informação necessária para garantir a continuidade dos tratamentos prestados ao cliente/paciente.

O prontuário do paciente foi desenvolvido por médicos e enfermeiros para garantir que se lembrassem de forma sistemática dos fatos e eventos clínicos sobre cada indivíduo de forma que, todos os demais profissionais envolvidos no processo de atenção de saúde possam também ter as mesmas informações (Slee e Schmidt, 2000). Desta forma, localmente, ou seja, na instituição onde o paciente está recebendo cuidados, o prontuário representa o mais importante veículo de comunicação entre os membros da equipe de saúde responsável pelo atendimento.

O Prontuário Eletrônico do Paciente (PEP) constitui em um sistema informatizado para a guarda e manuseio do prontuário médico. Este sistema tem por objetivo permitir o armazenamento e recuperação dos eventos clínicos dos pacientes tais como, atendimentos, prescrições, dispensação de medicamentos, exames realizados, ou seja, todo o histórico do paciente permitindo assim um melhor acompanhamento ao indivíduo. Outro conceito importante é o Registro Eletrônico de Saúde (RES) que permite o armazenamento e o compartilhamento seguro das informações de um paciente.

As informações registradas no prontuário médico vão subsidiar a continuidade e a verificação do estado evolutivo dos cuidados de saúde, quais procedimentos resultam em melhoria ou não do problema que originou a busca pelo atendimento, a identificação de novos problemas de saúde e as condutas diagnósticas e terapêuticas associadas. Em termos mais gerais, pode-se afirmar que o sistema de saúde de um país é estabelecido graças ao que se tem documentado em um prontuário, uma vez que dele são extraídas as informações sobre a saúde dos indivíduos que formam uma comunidade e uma nação.

A análise conjunta dos dados dos prontuários deveria ser capaz de fornecer, por exemplo, informações desagregadas ou agregadas sobre pessoas atendidas, quais tratamentos foram realizados, quais formas terapêuticas tiveram resultados positivos, como os pacientes responderam e quanto custou cada forma de tratamento por procedimento ou cumulativamente para todos os procedimentos de um paciente, de

grupos de pacientes ou de toda uma população. Estas informações agregadas e sistematizadas são necessárias para caracterizar o nível de saúde populacional e viabilizam a construção de modelos e políticas de atendimento e gestão das organizações de saúde.

O prontuário eletrônico do paciente é um registro eletrônico que reside em um sistema especificamente projetado para apoiar os usuários fornecendo acesso a um completo conjunto de dados corretos, alertas, sistemas de apoio à decisão e outros recursos, como links para bases de conhecimento médico.

O Prontuário Eletrônico do Paciente toma importância no sentido da disponibilização dos dados possibilitando analisar, avaliar as informações e gerar novos conhecimentos de uma forma mais eficiente e confiável além do compartilhamento, facilitando o diagnóstico e tratamento a qualquer hora e por qualquer profissional da área e saúde.

De acordo com CFM (2012), a estrutura de um prontuário, independente de ser eletrônico ou em papel, deve seguir as orientações e determinações da **Resolução CFM N° 1638/2002**, que define prontuário médico e torna obrigatória a criação da Comissão de Revisão de Prontuários nas instituições de saúde.

Apesar da importância do PEP, sua utilização encontra forte resistência junto aos profissionais de saúde dificultando assim sua adoção sob a legação de dificuldade no manuseio do sistema, aumento do tempo de atendimento, dificuldade na utilização da tecnologia além da mudança na forma de relacionamento com o paciente embora o prontuário em papel apresente diversas desvantagens em relação ao eletrônico tais como:

- A informação do prontuário em papel está disponível somente a um profissional ao mesmo tempo;
- Possui baixa mobilidade;
- Está sujeito à ilegibilidade;
- Ambiguidade;
- Perda frequente da informação;
- Multiplicidade de pastas;
- Dificuldade de pesquisa coletiva;
- Falta de padronização, dificuldade de acesso;

- Fragilidade do papel e a sua guarda requerem amplos espaços nos serviços de arquivamento.

Ainda de acordo com Eduardo (2003), para se iniciar a realização do prontuário eletrônico, alguns requisitos devem ser atendidos, tais como:

- Promover mudança comportamental dos profissionais, demonstrando a ineficiência dos registros manuais para lidarem com as grandes quantidades de informações geradas no sistema de saúde.

- Promover mudança nos sistemas e adotar sistemas computacionais abertos, que atendam aos requisitos de interoperabilidade.

- Utilizar tecnologia moderna, evitando a obsolescência.

- Usar normas (padrões) no registro clínico e na transferência das informações em saúde.

- Desenvolver e atualizar a legislação, atendendo às necessidades que foram impostas pelo desenvolvimento científico e tecnológico com criação de leis nacionais para estabelecer requisitos mínimos no uso de padrões e melhoria da qualidade do cuidado (padrões de prática).

Ainda de acordo com CFM (2012), vários estudos ao redor do mundo têm demonstrado o impacto positivo sobre a Saúde que a implantação de um PEP/RES pode trazer tanto para os profissionais de saúde quanto para os pacientes, gestores e toda a equipe envolvida na Atenção à Saúde. Além disso, o PEP é muito mais seguro do que o prontuário em papel além de permitir que as informações sejam compartilhadas automaticamente com outros profissionais e instituições que estão cuidando do paciente, possibilitando dessa forma a continuidade da atenção integral à saúde. Outros benefícios do PEP estão ligados à pesquisa clínica, adesão aos protocolos clínicos e assistenciais, além de usos secundários da informação para fins epidemiológicos e estatísticos.

As normas técnicas para o uso de sistemas de informatizados para guarda e manuseio do prontuário médico, CFM (2012) com relação a utilização de banco de dados regulamenta que: os prontuários eletrônicos deverão ser armazenados em sistema que assegure, pelo menos as seguintes características:

- I. Compartilhamento de dados;

- II. Independência entre dados e programas;

- III. Mecanismos para garantir a integridade, controle de conformidade e validação dos dados;

IV. Controle da estrutura física e lógica;

V. Linguagem para a definição e manipulação de dados SQL (*Structured Query Language*);

VI. Funções de auditoria e recuperação de dados.

Além das resoluções citadas, podemos destacar conforme CFM(2012) os seguintes padrões utilizados como referências na definição do Processo de Certificação SBIS/CFM.

- **Resolução CFM N.º 1638/2002** – que define prontuário médico e atribui as responsabilidades por seu preenchimento, guarda e manuseio.

- **Resolução CFM N.º 1639/2002** – que aprova as "Normas Técnicas para o Uso de Sistemas Informatizados para a Guarda e Manuseio do Prontuário Médico".

- **Resolução CFM N.º 1821/2007** – que aprova as "Normas Técnicas Concernentes à Digitalização e Uso dos Sistemas Informatizados para a Guarda e Manuseio dos Documentos dos Prontuários dos Pacientes, Autorizando a Eliminação do Papel e a Troca de Informação Identificada em Saúde".

- O Padrão TISS (Troca de Informação em Saúde Suplementar) – definido pela Agência Nacional de Saúde Suplementar – ANS (www.ans.gov.br) para registro e intercâmbio de dados entre operadoras de planos privados de assistência à saúde e prestadores de serviços de saúde. O objetivo do padrão TISS é atingir a compatibilidade e interoperabilidade funcional e semântica entre os diversos sistemas independentes para fins de avaliação da assistência à saúde (caráter clínico, epidemiológico ou administrativo) e seus resultados, orientando ao planejamento do setor.

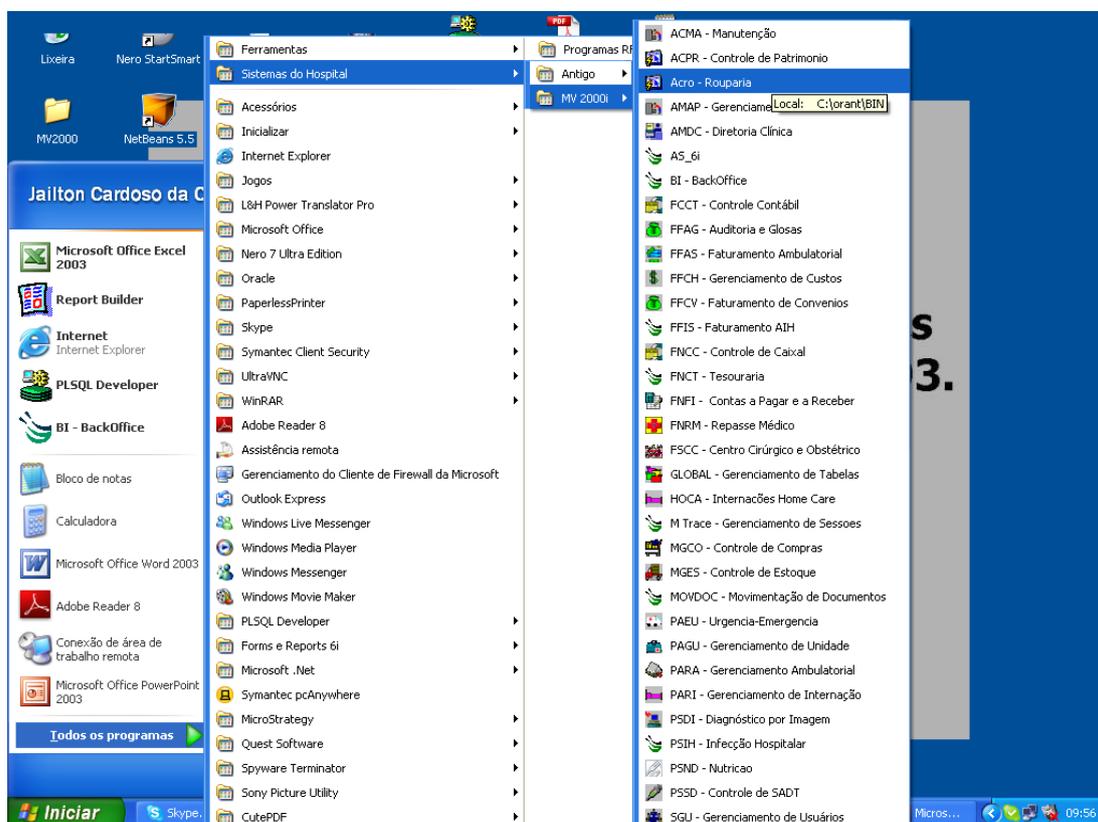
- Normas ISO TC-215 - A norma ISO/TR 20.514:2005 é um documento de referência técnica (TR - *Technical Report*) que estabelece as definições de RES e de Sistemas de RES. Esse relatório descreve as principais categorias de sistemas, define cenários de utilização, e a necessidade de interoperabilidade semântica entre os diferentes sistemas RES.

2.4 O SISTEMA MV

O Sistema de Gestão em Saúde MV é uma solução que integra de todos os setores e processos da instituição, além de disponibilizar informações que permitam a instituição constatar oportunidades e tornar-se mais competitiva, MV

(2013). O sistema é composto de módulos direcionados a cada área de negócio do hospital.

Figura 1 – Menu Principal dos Módulos do Sistema MV



Fonte: o autor

O Sistema MV, adotado pela Santa Casa de Misericórdia de Maceió, possui em torno de trinta módulos integrados que abrange todas as áreas operacionais de um hospital, conforme mostrado na figura 1, que abrange as atividades administrativas, hospitalares e de hotelaria. Dentre os módulos, o Sistema MV possui o Prontuário Eletrônico do Paciente que é alimentado por alguns subsistemas tais como Urgência e Emergência, Gerenciamento de Unidade, Diagnóstico por Imagem, Controle SADT (Serviço Auxiliar Diagnóstico e Terapia), Gerenciamento Ambulatorial, Gerenciamento de Internação.

A adoção do Sistema MV tem contribuído fortemente para tornar a Santa Casa de Misericórdia de Maceió referência em administração hospitalar, possibilitando melhoria na competitividade na área de atendimento hospitalar bem como uma maior integração dos fluxos nos processos de controle interno.

O Sistema MV, além do módulo transaccional também possui módulos direcionados a gestão hospitalar tais como, Soul MV (Sistema de Gestão de Saúde), BI -

Portal de Indicadores, Bússola Gerencial, BSC (*Balance Scorecard*), dentre outros módulos gerenciais.

O sistema MV foi utilizado como base de dados para o desenvolvimento deste trabalho como mostrado nos capítulos 8 e 9.

Toda a base de dados extraída para o desenvolvimento do trabalho bem como o processo de levantamento de dados e de requisitos foi realizado com o acompanhamento do Gerente de Tecnologia da Informação Sr. Marcus Aurélio.

2.5 O RACIOCÍNIO MÉDICO E A TOMADA DE DECISÃO

Profissionais de medicina estão sempre envolvidos no processo de tomada de decisão. Tal decisão deve ser baseada em informações coletadas do paciente juntamente com conjunto de conhecimentos médicos para o diagnóstico, tratamento de doenças ou recomendações para melhoria da qualidade de vida. Uma característica importante no raciocínio médico é que o elemento vida faz com que a precisão da decisão não deixe margens para erro.

2.5.1 REPRESENTAÇÃO DO RACIOCÍNIO

O termo raciocínio refere-se a um processo mental cujo objetivo é, a partir de conhecimentos já adquiridos, alcançar novos conhecimentos, sendo, portanto um processo de pensamento pelo qual podemos derivar, inferir ou justificar conclusões a partir de um conjunto de premissas. De acordo com Vinla (2005) geralmente assumimos dois tipos de raciocínio: o raciocínio **dedutivo** e o **indutivo**, no entanto, o raciocínio no "mundo real" não parece se encaixar perfeitamente para esses tipos básicos. Por esta razão, uma terceira forma de raciocínio tem sido reconhecida na qual dedução e indução são combinados. Este foi denominado raciocínio **abdutivo**. Além destes tipos de raciocínio podemos destacar também os de **senso comum** e **não-monotônico**.

O raciocínio dedutivo é umas das técnicas mais utilizadas do raciocínio humano na resolução de problemas. Este tipo de raciocínio consiste em derivar uma conclusão a partir um conjunto geral de premissas da forma $A \rightarrow B$ (A implica em B). Por exemplo, seja o seguinte pensamento:

Se A é verdadeiro e A implica em B , então B também é verdadeiro.

Premissa 1 – Todo alagoano é brasileiro

Premissa 2. João é alagoano

Conclusão: João é brasileiro

O raciocínio indutivo deriva uma conclusão generalizada a partir de fatos que acontecem corriqueiramente. Segundo Peirce (1955), o raciocínio indutivo, ou sintético, é mais do que a mera aplicação de uma regra geral a um caso particular. Parte de uma premissa menor para uma maior. A indução é a inferência de uma regra a partir do caso e do resultado. Sendo assim, ela ocorre quando generalizamos a partir de certo número de casos em que algo é verdadeiro e inferimos que a mesma regra será verdadeira do total da classe.

Por exemplo: sejam x , y e z membros da família f

X é educado;

Y é educado;

Z é educado;

Então todos os membros da família f são educados.

Para situações do mundo real estes dois tipos de raciocínio parecem não ser suficientes, sendo assim uma terceira forma de raciocínio, na qual dedução e indução são combinadas denominado raciocínio abduativo (Peirce, 1955)

Ainda de acordo com Vinla (2005) toda a teoria do raciocínio médico caracteriza como um processo de abdução, isto é, uma identificação de um conjunto de hipóteses, que são capazes de explicar a caso clínico, com base nos dados disponíveis e testando a explicação (ou seja, avaliação de cada hipótese gerada em função das suas consequências esperadas) para o estado anormal do paciente em questão.

Estes modelos podem ser combinados em um caso mais genérico: o modelo *select and test* no qual o processo de geração de hipóteses e testes pode ser caracterizado em quatro tipos de inferência Peirce (1955) apud Vinla (2005) – abstração, abdução, dedução e indução.

Durante a abstração, os dados são filtrados de acordo com sua relevância para a solução do problema e fragmentada em esquemas representando uma descrição abstrata do problema em questão por exemplo, abstraindo que um homem adulto com concentração de hemoglobina inferior a 14 d/gl é um paciente anêmico. Após isso, hipóteses que poderiam explicar a atual situação estão relacionadas através de um processo de abdução caracteriza-se por um "fluxo para trás" de inferências através de uma cadeia de relações dirigidas, que identificam as condições iniciais, a partir do qual a representação abstrata atual do problema se origina. Isto fornece soluções provisórias para o problema em questão por meio de hipóteses. Por exemplo,

saber que a doença *A* causará sintoma *b*, abdução vai tentar identificar a explicação para *b*, e dedução irá prever que um paciente afetado pela doença *A* irá manifestar sintoma *b*: Ambas as inferências estão usando a mesma relação ao longo de duas direções diferentes. Estes três tipos de raciocínio na medicina são descritos em um artigo escrito por Patel e Ramoni (1997) apud Vimla (2005).

2.6 O DOMÍNIO DO CÂNCER PEDIÁTRICO

O câncer infanto-juvenil, que ocorre em jovens com idade abaixo dos 19 anos, é considerado raro quando comparado com os tumores dos adultos, correspondendo entre 2% e 3% de todos os tumores malignos (INCA, 2008).

O câncer infantil corresponde a um grupo de várias doenças que têm em comum a proliferação descontrolada de células anormais e que pode ocorrer em qualquer local do organismo. Os tumores mais frequentes na infância e na adolescência são as leucemias (que afeta os glóbulos brancos), os do sistema nervoso central e linfomas (que afeta o sistema linfático).

Também acometem crianças e adolescentes o neuroblastoma (tumor de células do sistema nervoso periférico, frequentemente de localização abdominal), tumor de Wilms (tipo de tumor renal), retinoblastoma (afeta a retina, fundo do olho), tumor germinativo (das células que vão dar origem aos ovários ou aos testículos), osteossarcoma (tumor ósseo) e sarcomas (tumores de partes moles) (INCA, 2013).

O câncer infanto-juvenil deve ser estudado separadamente do câncer do adulto por apresentar diferenças nos locais primários, diferentes origens histológicas e diferentes comportamentos clínicos.

De acordo com INCA (2008), o câncer infantil não pode ser considerado uma simples doença, mas sim, como uma gama de diferentes malignidades. Esse tipo de câncer varia de acordo com o tipo histológico, localização primária do tumor, etnia, sexo e idade e agrupadas conforme a Classificação Internacional do Câncer na Infância (CICI).

Grupo I – Leucemia

Grupo de neoplasias corresponde às categorias: I.a. Leucemia linfóide; I.b. Leucemia não linfocítica aguda; I.c. Leucemia mielóide crônica; I.d. Outras leucemias especificadas; I.e. Leucemias não especificadas. Este tipo de câncer é o mais comum em menores de 15 anos na maioria das populações.

Grupo II: Linfomas e neoplasias retículo-endoteliais

Esse grupo de neoplasias corresponde às categorias: II.a. Doença de Hodgkin (DH); II.b. Linfomas não-Hodgkin (LNH); II.c. Linfoma de Burkitt; II.d. Miscelâneas de neoplasias linfo-reticulares e II.e Linfomas não especificados.

Grupo III: Tumores de sistema nervoso central e miscelânea de neoplasias intracranianas e intra-espinais

Esse grupo de neoplasias corresponde às categorias: III.a. Ependimoma; III.b. Astrocitoma; III.c. Tumores neuroectodérmicos primitivos; III.d. Outros gliomas; III.e. Outras neoplasias intracranianas e intra-espinais especificadas e III.f. Neoplasias intracranianas e intra-espinais não especificadas.

Grupo IV: Tumores do sistema nervoso simpático

Esse grupo de neoplasias corresponde às categorias: IV.a. Neuroblastoma e anglioneuroblastoma e IV.b. Outros tumores do sistema nervoso simpático.

Grupo V: Retinoblastoma

O retinoblastoma (RB), tumor intra-ocular maligno.

Grupo VI: Tumores renais

Esse grupo de neoplasias corresponde às categorias: VI.a. Tumor de Wilms, tumor rabdóide e sarcoma de células claras; VI.b. Carcinoma renal e VI.c. Tumores renais malignos não especificados.

Grupo VII: Tumores hepáticos

Esse grupo de neoplasias corresponde às categorias: VII.a. Hepatoblastoma, VII.b. Hepatocarcinoma e VII.c. Tumores hepáticos malignos não especificados.

Grupo VIII: Tumores ósseos malignos

Esse grupo de neoplasias corresponde às categorias: VIII.a. Osteossarcoma; VIII.b. Condrossarcoma; VIII.c. Sarcoma de Ewing; VIII.d. Outros tumores ósseos malignos específicos e VIII.e. Tumores ósseos malignos não especificados.

Grupo IX: Sarcomas de partes moles

Esse grupo de neoplasias corresponde às categorias: IX.a. Rbdomiossarcoma e sarcoma embrionário; IX.b. Fibrossarcoma, neurofibrossarcoma e outras neoplasias fibromatosas; IX.c. Sarcoma de Kaposi; IX.d. Outros sarcomas de partes moles especificados e IX.e. Sarcomas de partes moles não especificados.

Grupo X: Neoplasias de células germinativas, trofoblásticas e outras gonadais

Esse grupo de neoplasias corresponde às categorias: X.a. Tumores de células germinativas intracranianas e intra-espinhais; X.b. Outros tumores de células germinativas não gonadais e tumores de células germinativas não gonadais não especificados; X.c. Tumores de células germinativas gonadais; X.d. Carcinomas gonadais e X.e. Outros tumores gonadais malignos e tumores gonadais não especificados.

Grupo XI: Carcinomas e outras neoplasias malignas epteliais

Esse grupo de neoplasias corresponde às categorias: XI.a. Carcinoma de córtex adrenal; XI.b. Carcinoma de tiróide; XI.c. Carcinoma de nasofaringe; XI.d. Melanoma maligno; XI.e. Carcinoma de pele e XI.f. Outros carcinomas e carcinomas não especificados.

Grupo XII: Outros tumores malignos não especificados

Esse grupo de neoplasias corresponde às categorias: XII.a. Outros tumores malignos especificados e XII.b. Outros tumores malignos não especificados.

Maiores detalhes podem ser obtidos em (INCA, 2008).

2.7 O CÂNCER PEDIÁTRICO NO BRASIL

No Brasil, os dados sobre a ocorrência de câncer são obtidos através dos Registros de Câncer que são centros de coleta, armazenamento, processamento e análise de informações sobre pacientes ou pessoas com diagnóstico confirmado de câncer. Esses centros podem ser de base populacional RCBP (Registros de Câncer de Base Populacional) ou de Base Hospitalar (RHC).

Os registros de câncer, segundo HIJG (2010) fazem parte de um sistema de vigilância estruturado que fornece informações fundamentais para o planejamento e o monitoramento da efetividade de programas de controle de câncer.

Os dados disponíveis em relação à ocorrência da doença, a mortalidade e a sobrevida, permitem analisar a magnitude do problema no Brasil, bem como possibilitam a formulação de hipóteses causais para o desenvolvimento da mesma. As informações obtidas dos registros de câncer auxiliam também na avaliação dos avanços tecnológicos aplicados à prevenção e ao tratamento.

Os registros de câncer de base hospitalar registrados pelo INCA (2008) têm como finalidade contribuir para a melhoria da assistência prestada ao paciente e o planejamento interinstitucional, auxiliar na organização de um Sistema de

Informações em Câncer e promover educação profissional continuada aos profissionais da área de saúde pública. As informações produzidas em um RHC (Registro Hospitalar de Câncer) refletem o desempenho do corpo clínico na assistência prestada ao paciente.

Nos registros de câncer de base populacional o principal objetivo é avaliar o impacto do câncer em uma determinada população. Os RCBP visam conhecer o número de casos novos (incidentes) de câncer, assim como suas distribuições e tendências temporais nas populações pertencentes às áreas geográficas por eles cobertas e, portanto, acompanhar a morbi-mortalidade através dos indicadores de incidência na área onde o mesmo é situado.

Os resultados e comentários sobre a incidência dos RCBPs brasileiros estão detalhados em (INCA 2008).

3 O MODELO DE REFERÊNCIA CRISP-DM

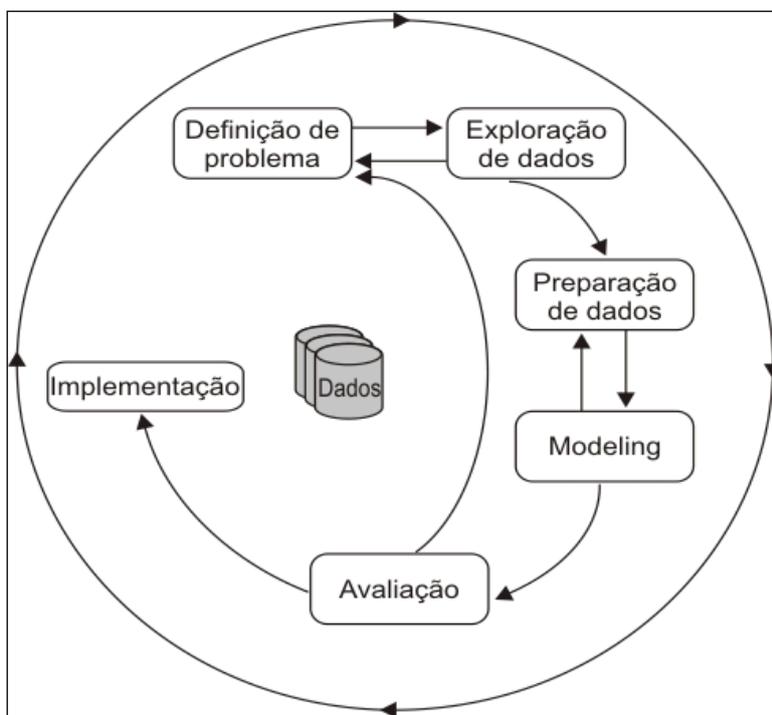
3.1 INTRODUÇÃO

Este capítulo apresenta o ciclo de descoberta do conhecimento. Sistemas RBC são fortemente relacionados ao domínio do problema. Neste trabalho apresentaremos a metodologia CRISP-DM (Chapman, 2000).

3.2 O CICLO DE VIDA DO CRISP-DM

O ciclo de vida CRISP-DM consiste em seis fases, de acordo com a figura 2. Embora a sequência de fases não seja rígida, sua característica cíclica facilita o processo de aquisição de conhecimento. No círculo mais externo, as setas indicam uma dependência mais forte entre as fases. O círculo mais interno indica a natureza cíclica do processo de mineração de dados. As lições aprendidas no processo e na solução implantada podem desencadear novo processo de negócio além de que processos de mineração de dados são beneficiados com as experiências dos processos anteriores.

Figura 2 - Modelo Referência CRISP-DM



Fonte: Adaptado de Chapman (2000)

3.3 COMPREENSÃO DO NEGÓCIO

Esta fase tem como objetivo o entendimento do problema, seus objetivos e transformá-lo em uma definição do problema de mineração de dados (Chapman, 2000).

Identificar as necessidades do usuário sob uma perspectiva de negócio a fim de definir as questões relacionadas ao negócio que irá direcionar o desenvolvimento do projeto. Algumas tarefas desta fase são:

- Definição do problema a ser resolvido;
- Especificar os benefícios a serem obtidos;
- Especificar os critérios de sucesso;
- Especificar as premissas e restrições (p.ex. legais, tempo e recursos e custos) do projeto;
- Identificar riscos;
- Custos e benefícios.

Uma tarefa importante desta fase é determinar os objetivos da mineração de dados. Por exemplo, “Prever o comportamento de clientes a partir de suas compras nos últimos três anos distribuídos por faixa etária, salário, cidade, etc”.

3.4 COMPREENSÃO DOS DADOS

A fase de compreensão dos dados envolve os processo de coleta, descrição, exploração e validação dos dados.

Coleta dos Dados

Obtenção ou acesso aos dados disponibilizados para o processo de mineração. Esta fase inicial inclui a carga dos dados e, se necessário, a análise dos dados. Os dados podem estar disponíveis em uma única fonte ou em fontes diversas.

Algumas atividades desta fase são:

- Especificar os critérios de seleção. Quais os atributos são necessários para o processo de mineração?
- Selecionar tabelas/arquivos de interesse;
- Selecionar os dados;

- Definir o tempo de histórico a ser utilizado.

Descrever dos Dados

Descrever os dados a serem minerados em termos de formato, tipo de dado e quantidade (número de registros) e identificar os dados descrevendo seu significado, caso o mesmo não esteja descrito no dicionário de dados. As atividades desta fase são:

- Identificar os dados e o método de captura;
- Acessar as fontes de dados;
- Utilizar análise estatística, se necessário;
- Checar o volume de dados, cardinalidade e complexidade;
- Apresentar as tabelas e suas relações (MER – Modelo Entidade-Relacionamento)

Explorar os Dados

Nesta tarefa, são tratadas as atividades de mineração de dados, tais como consultas, visualização e técnicas de relatório que podem ser apresentados através de gráficos. Esta análise deve ser direcionada aos objetivos da mineração de dados definidos na fase de compreensão dos dados.

Verificar a Qualidade dos Dados

Examinar o conjunto de dados de acordo com os objetivos do processo de mineração com respeito à completude, ocorrência de erros, valores omitidos e o grau de ocorrência. Algumas atividades deste processo:

- Verificar o significado dos atributos;
- Identificar valores omitidos e em nulos;
- Checar atributos com valores diferentes, mas que possuem o mesmo significado.
- Checar atributos com formatação diferentes (p. ex. maiúsculo e minúsculo).

3.5 PREPARAÇÃO DOS DADOS

Esta fase é focada no pré-processamento a fim de produzir os dados a serem minerados coletando os dados apropriados de uma mesma fonte ou de fontes diversas.

Um processo importante nesta fase de preparação de dados é a limpeza dos dados que envolvem: a exclusão de dados inválidos, utilização de valores default ou adotar valores estimados para dados incertos.

Preparação de dados também envolve a construção de atributos derivados que consiste da criação de novos atributos a partir de um ou mais atributos existentes.

Transformações podem ser necessárias a fim de dar um padrão a valores descritivos, tais como ‘N’ para valores como ‘Não’, ‘NÃO’, ‘n’ ou ‘N’ ou definir faixas de valores para dados numéricos tal como “idade” para “faixa etária”.

Nesta fase a integração dos dados, que é a tarefa de combinar, padronizando informações de múltiplas tabelas em tais como, ‘N’ para valores como ‘Não’, ‘NÃO’, ‘n’ ou ‘N’ armazenando em visões materializadas com objetivo de fazer a junção dos dados, bem como utilizar as funções de agregação para criação de dados sumarizados.

3.6 MODELAGEM

Nesta fase ocorre a escolha da técnica de modelagem. Antes de criar o modelo deve-se decidir qual o modelo de teste que será aplicado, qualitativo ou de validação. Após a escolha do modelo, o mesmo deve ser aplicado utilizando-se uma ferramenta de modelagem a partir de um conjunto de parâmetros a serem aplicados no conjunto de dados. Tal conjunto de parâmetros deve ser documentado a fim de justificar a escolha. Após a execução, o modelo deve ser validado para assegurar que os objetivos do processo de mineração de dados foram atingidos de acordo com os critérios de sucesso com base nas atividades a seguir:

- Avaliar os resultados de acordo com os critérios de avaliação;
- Testar o resultado de acordo com a estratégia de teste;
- Comparar resultados e interpretá-los;
- Criar uma escala de resultados de acordo com os critérios de sucesso;
- Selecionar o melhor modelo;
- Interpretar o resultado em termos de negócio;
- Checar a aplicabilidade do modelo;
- Checar a confiabilidade dos dados;

- Avaliar resultados;

3.7 AVALIAÇÃO

Este passo avalia o grau pelo qual o modelo atingiu os objetivos do negócio e verifica se há alguma deficiência em relação aos objetivos. Ela compara resultados com os critérios de avaliação definidos no início do projeto. Uma questão chave nesta fase é identificar se há alguma questão que não foi suficientemente considerada.

3.8 IMPLEMENTAÇÃO

Após a criação do modelo esta fase se preocupa em dar uma interface ao modelo, seja pela utilização de uma página na internet ou um sistema de banco de dados. Dependendo das características do modelo esta fase pode utilizar-se apenas de relatórios.

4 RACIOCÍNIO BASEADO EM CASOS

4.1 INTRODUÇÃO

Imaginemos o diálogo entre pai e filho indo a um estádio de futebol.

- Filho: Pai, estamos atrasados para o jogo.

- Pai: pois é meu filho temos que nos apressar

- Filho: então papai, teremos que ir por um caminho alternativo para chegarmos logo.

- Pai: Da última vez que estávamos atrasados fomos pela rua x.

- Filho: Mas quando fomos pela rua x tivemos dificuldade de estacionar.

- Pai – é mesmo filho então faremos a trajetória pela rua z, pois apesar do trajeto ser um pouco maior, o trânsito é bem mais tranquilo e não teremos dificuldades para estacionar.

Neste diálogo, pai e filho, a partir de experiências anteriores, buscam uma alternativa para chegar ao estádio da melhor forma possível. Raciocínio baseado em casos significa usar experiências anteriores para entender e resolver novos problemas. Neste capítulo iremos dar uma breve introdução aos sistemas de raciocínio baseado em casos ou CBR (utilizando-se a sigla original para *Case-based Reasoning*).

4.2 RACIOCÍNIO BASEADO EM CASOS

Na era da tecnologia, informação e comunicação, onde somos “bombardeados” diariamente por uma quantidade infinita de informações oriundas de diversas mídias, tais como internet, jornal, TV, sistemas de informação, é comum nos depararmos com a pergunta “com tantas informações, como selecionar as que são úteis?” Tal questionamento já foi inclusive tema de redação de vestibular da UPE (Universidade de Pernambuco) em 2011 (NE10, 2013). Mas, será que estamos preparados para absorver e processar estas informações de modo a filtrar informações relevantes para o desenvolvimento de uma atividade?

Raciocínio baseado em casos é uma abordagem para a resolução de problemas que enfatiza o papel da experiência prévia para resolver problemas futuros, ou seja, novos problemas são resolvidos da reutilização e, se necessário, adaptar as

soluções para problemas semelhantes que foram resolvidos no passado. RBC tem tido um sucesso considerável em uma grande variedade de tarefas de resolução de problemas e domínios (Mantaras, 2005).

4.2.1 HISTÓRICO

Raciocínio baseado em casos tem sua origem na Inteligência Artificial e nas Ciências Cognitivas nos anos 70. Com forte influência dos modelos orientados a esquema de memória, estudados na psicologia (Bartlett, 1932). Muitas características da teoria dos esquemas de Bartlett ocorrem na teoria da memória dinâmica de Schank, (1982) e cujo princípio fundamental é lembrar-se de situações anteriores (episódios e casos) e os padrões de situação (*scripts*, MOPs) para a resolução de problemas e aprendizagem.

A teoria da memória dinâmica tem nos “pacotes de organização da memória” (MOPs) como elementos que tiveram grande impacto no desenvolvimento do raciocínio baseado em casos. Schank e Abelson (1977) propuseram que as sequências de eventos comuns são guardadas em *scripts* que as pessoas utilizam no dia a dia. Um *script* é uma estrutura de conhecimento na memória, representando cenários, sequência de eventos, personagens e demais situações relativas a uma determinada experiência.

Podemos dizer que o sistema CYRUS, desenvolvido por Janet Kolodner, Kolodner (1983) foi o primeiro sistema que seria um raciocinador baseado em casos. O sistema CYRUS foi criado utilizando o modelo de memória dinâmica e teoria MOP de resolução de problemas e aprendizagem (Schank, 1982). O CYRUS era basicamente um sistema de perguntas e respostas (Q&A) com o conhecimento das viagens e reuniões do ex-secretário de Estado dos Estados Unidos Cyrus Vance.

O modelo de casos utilizado no sistema CYRUS serviu como base para o desenvolvimento de outros sistemas de raciocínio baseado em casos tais como:

- MEDIATOR (Simpson, 1985) – utilizado como mediador na disputa em um sistema de planejamento propondo possíveis soluções de consenso.
- Persuader (Sycara, 1987) - mediar negociação entre patrões e empregados.
- CHEF (Hammond, 1989) – que desenvolve novas receitas baseada em outras receitas já conhecidas.

- CASEY (Koton, 1988) – permite diagnosticar problemas no coração a partir da adaptação de diagnóstico de outros pacientes.
- JULIA (Hinrichs, 1992) – utilizado para planejamento de refeições.

Na área de diagnóstico médico, um solucionador de problemas é dado a partir de um conjunto de sintomas e perguntas para explicá-los. Um diagnosticador baseado em casos, pode usar casos para sugerir explicações para sintomas e avisar de explicações inapropriadas que foram feitas no passado. Um exemplo deste sistema é o SHRINK (Kolodner e Kolodner, 1987) projetado para diagnóstico na área de psiquiatria.

Segundo Holt (2005) a aplicação de CBR na medicina foca principalmente no diagnóstico, classificação, planejamento e tutoria. A tabela 4.1 apresenta alguns sistemas desenvolvidos utilizando a abordagem CBR classificado pelo tipo de tarefa.

Tabela 1- Aplicação da CBR na medicina

Tipo de Tarefa	Citação	Nome do Sistema	Domínio de Aplicação
Sistemas de Diagnóstico e Sistemas de Apoio a Decisão	(Kolodner & Kolodner, 1987)	SHRINK	Psiquiatria
	(Koton, 1988)	CASEY	Coração
	(Turner, 1988)	MEDIC	Dispneia
	(Lopez & Plaza, 1993)	BOLERO	Pneumonia
	(Bradburn & Zeleznikow, 1993)	FLORENCE	Planejamento de cuidado do coração
	(Bichindaritz, 1995)	MNAOMIA	Psiquiatria
	(Haddad <i>et al.</i> , 1997)	SCINA	Deteção de doenças coronárias
	(Bichindaritz <i>et al.</i> , 1998)	CAREPARTNER	Transplante de células estaminais
	(Marling & Whitehouse, 2001)	AUGUSTE	Diagnóstico de Alzheimer
Sistemas de classificação	(Bareiss & Porter, 1987)	PROTOS	Alterações audiológicas

	(Macura <i>et al.</i> , 1994)	<i>MACRAD</i>	Análise de imagens
	(Grimnes & Aamodt, 1996)	<i>IMAGECREEK</i>	Análise de imagens
	(Bichindaritz & Potter, 1994; Bichindaritz & Potter, 2004)	<i>PHYLSYST</i>	Classificação filogenética
	(Perner, 1999)	<i>CTS</i>	Análise de imagens
Planejamento de sistemas	(Bichindaritz & Seroussi, 1992)	<i>ALEXIA</i>	Hipertensão
	(Berger, 1994)	<i>ROENTGEN</i>	Terapia de radiação
	(Petot <i>et al.</i> , 1999)	<i>CAMP</i>	Planejamento de cardápio diário
	(Montani <i>et al.</i> , 2000)	<i>T-IDDM</i>	Tratamento de diabetes

Fonte: Holt (2005)

4.2.2 VANTAGENS DO USO DE CBR

A utilização de sistemas CBR traz vantagens em vários pontos de vista em relação a outras abordagens, tal como raciocínio baseado em regras e máquina de aprendizado.

- *A facilidade de explicação*: Os resultados de um sistema de CBR podem ser justificados com base na semelhança do problema atual com o(s) caso(s) recuperado(s).
- *Modularidade*: Cada caso é uma unidade discreta de conhecimento independente, que pode ser inserido ou removido da base de casos.
- *Aprendizagem incremental* – um sistema CBR pode ser colocado em operação com um conjunto mínimo de casos resolvidos. A base de casos será alimentada com novos casos com o uso do sistema, aumentando a capacidade de resolução do sistema.
- *Naturalidade de representação* – casos são um método de representação do conhecimento simples e facilmente compreensível pelo usuário, o que reduz o tempo da fase de aquisição do conhecimento.

Por outro lado, sistemas CBR apresentam alguns problemas, tais como:

- *Incapacidade de expressar o conhecimento geral* - Casos, por natureza, expressam conhecimento especializado. Então, eles não podem expressar o conhecimento geral.
- *Tratar grande base de casos* – Isto requer grande quantidade de memória bem como alto consumo de tempo para retornar os casos.
- *Dificuldades de explicar* – apresentar explicações para as conclusões obtidas bem como todos os passos do raciocínio é uma tarefa difícil.

4.2.3 A ABORDAGEM RBC

No final dos anos 70, os programas que modelavam a memória episódica foram tema de várias pesquisas na área de IA. Tais sistemas foram desenvolvidos em uma área de aplicação denominada Sistemas Especialistas baseados em Regras (*Ruler-base Expert Systems*) os quais se propunham simular a capacidade de resolução de problemas de determinada área de habilidade humana, tais como na área de química ou médica. O sucesso destes sistemas estimulou o desenvolvimento de programas para outras áreas de especialização.

De acordo com Slade (1991), a unidade básica de conhecimento nestes sistemas especialistas era a regra. Uma regra composta por um par teste de ação condicional, por exemplo, “se a condição, então a ação”. Várias centenas de regras podem ser necessárias para uma tarefa típica de diagnóstico ou reparo. Com o crescimento dos sistemas especialistas surgiram os problemas do paradigma baseado em regras.

Tabela 2 - Problemas do paradigma baseado em regras

Problema	Descrição	Consequência
Aquisição do conhecimento	Para construir um sistema especialista, um programador de computador (ou engenheiro de conhecimento) teve que se sentar com o especialista humano para determinar quais as regras eram adequadas para um determinado domínio. Este conhecimento foi difícil de descobrir. O especialista humano não pode simplesmente fazer uma lista de centenas de regras utilizadas para resolver problemas.	A dificuldade de aquisição do conhecimento tornou-se um gargalo na construção de sistemas baseados em regras.

Sistemas baseados em regras não têm uma memória	Sistemas baseados em regras não se lembram dos problemas que já tinham resolvidos	Precisão e eficiência são problemas relacionados para um sistema sem uma memória
Sistemas baseados em regras não eram robustos	Se um problema for apresentado ao sistema que não corresponde a nenhuma das regras, o programa não poderia responder.	Fragilidade do sistema

Fonte: adaptado de Slade (1991)

A tecnologia de sistemas baseados em casos aborda diretamente os problemas encontrados em sistemas baseados em regras, adaptado de Slade (1991), quais sejam;

- **Aquisição do conhecimento** - a unidade de conhecimento é o caso e não a regra;
- **Experiência do desempenho** - Um sistema baseado em casos pode se lembrar do seu próprio desempenho e modificar seu comportamento para evitar a repetição de erros anteriores.
- **Soluções adaptativas** - Por analogia com o raciocínio de casos anteriores, um sistema baseado em casos deve ser capaz de construir novas soluções para os problemas.

Ao construir sistemas baseados em casos devemos ter em mente as seguintes perguntas:

- Primeiro: o que se torna um caso? Como podemos representar a memória do caso?
- Segundo: a memória é organizada? Quais são as regras de indexação?
- Terceiro: como faz alteração de memória? Como a memória do caso e as regras de indexação mudam com o tempo?
- Quarto: como podemos adaptar velhas soluções para novos problemas? Quais são as métricas de similaridade e as regras de modificação?
- Quinto: como podemos aprender com os erros? Quais são as regras de reparação?

A tecnologia de sistemas baseados caso é uma instanciação das teorias psicológicas de raciocínio baseado em casos.

4.2.4 SISTEMA BASEADO EM CASOS (*CASE-BASED SYSTEM*)

O que é raciocínio baseado em casos? Basicamente: resolver um novo problema a partir da lembrança de situações anteriores similares através da reutilização da informação e conhecimento da situação. Na terminologia CBR, um caso geralmente denota uma situação-problema. A situação anteriormente experimentada, que foi obtida e aprendida de uma maneira que ela pode ser reutilizada na resolução dos problemas futuros, é referida como um caso passado, caso anterior, caso armazenados, ou caso retido. Correspondentemente, um novo caso ou caso não resolvido é a descrição de um novo problema para ser resolvido. Raciocínio baseado em casos é - de fato - um processo cíclico e integrado de resolução de um problema que aprende com esta experiência na solução de um problema novo (Aamodt, 1994).

Existem dois tipos de Raciocínio Baseado em Casos Kolodner (1992), CBR interpretativa e CBR para Resolução de Problemas. CBR interpretativa utiliza casos anteriores como pontos de referência para classificar ou caracterizar situações novas. CBR para resolução de problemas usa casos anteriores para sugerir soluções que podem ser aplicadas a novas circunstâncias.

Na CBR interpretativa Leake (1996), o objetivo do raciocinador é formar um julgamento sobre uma nova situação, comparando e contrastando-a com os casos que já foram classificados. CBR interpretativa também é importante para tarefas como diagnóstico, no qual um problema pode ser diagnosticado, comparado e contrastado de acordo com sintomas correntes com aqueles ocorridos em casos anteriores para determinar o melhor diagnóstico.

Na sua forma mais simples, CBR interpretativa envolve quatro etapas. Primeiro, o raciocinador deve realizar avaliação da situação Kolodner (1992), para determinar quais as características da situação atual são realmente relevantes. Em segundo lugar, com base nos resultados da avaliação da situação, o raciocinador recupera um caso anterior ou casos anteriores relevantes. Terceiro, o raciocinador então compara esses casos à nova situação, para determinar qual a interpretação se aplica. Finalmente, a situação atual e a interpretação são salvos como um novo caso que irão basear o raciocínio futuro.

O objetivo da CBR baseado em resolução de problemas Leake (1996) é aplicar uma solução anterior a fim de gerar a solução para um problema novo. Por exemplo, caso baseado em projeto, planejamento e sistemas de explicação todos

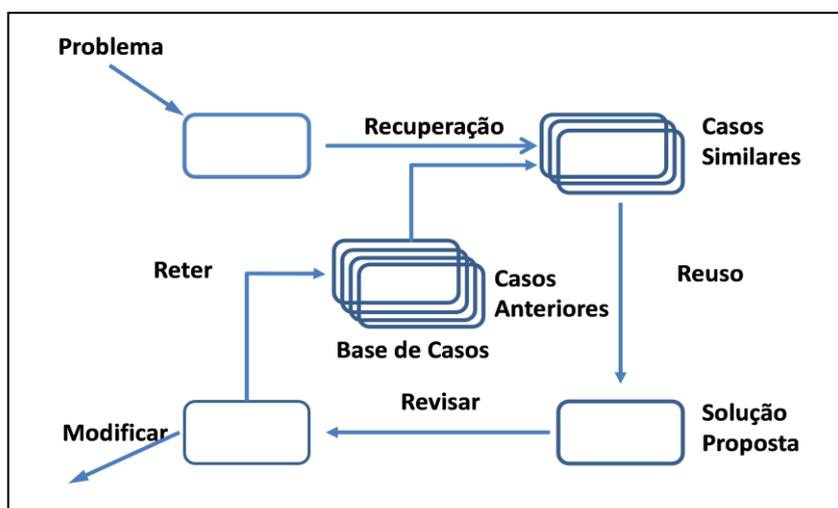
recuperam e adaptam as soluções de problemas similares anteriores. Assim, como CBR interpretativa, CBR baseado em solução de problemas envolve avaliação da situação, a recuperação caso, e avaliação de similaridade. Além disso, por exemplo, um sistema de planejamento baseado em casos gera um novo plano de recuperação de um plano prévio para um objetivo semelhante, determinando as diferenças entre os objetivos antigos e novos, e adaptando o plano para tomar as novas metas em conta.

4.2.5 CICLO DE VIDA DE UM CBR

O processo envolvendo um CBR pode ser representado por um ciclo conforme descrito na figura 3. Este ciclo pode ser descrito tipicamente como um processo representado por quatro RE's (Amoodt & Plaza 1994).

A resolução de um problema CBR, de acordo com Mantaras (2005), envolve a obtenção de uma descrição do problema, a medição da similaridade do problema atual aos problemas anteriores armazenados numa base de caso (ou de memória) com as soluções conhecidas, a *recuperação* de um ou mais casos semelhantes, e a tentativa de *reutilizar* a solução de um dos casos recuperados, para possivelmente depois adaptá-lo para explicar as diferenças nas descrições do problema. A solução proposta pelo sistema é então avaliada (por exemplo, a ser aplicada ao problema inicial ou avaliada por um especialista do domínio). Após *revisão* da solução proposta, se necessário, tendo em conta a sua avaliação, a descrição do problema e a sua solução pode ser mantido como um novo caso, e que o sistema tenha aprendido a resolver um problema novo.

Figura 3 - Ciclo de vida CBR



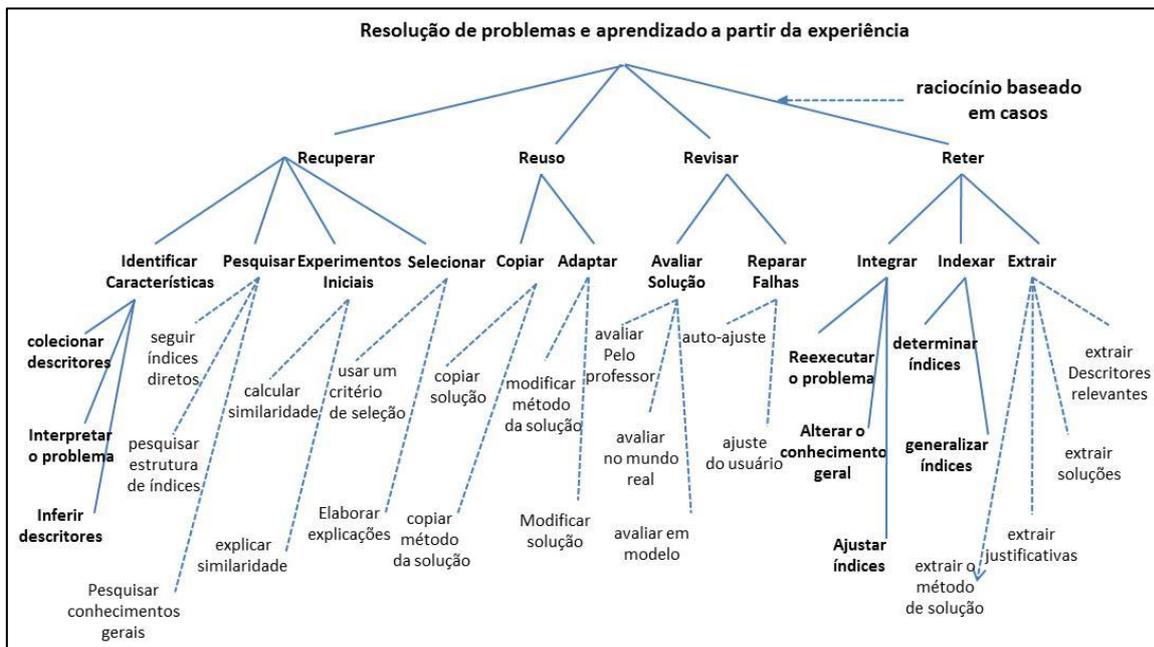
Fonte: adaptado de Amoodt & Plaza (1994)

- *REtrieve* o caso mais similar;
- *REuse* o caso ou casos para resolver o problema;
- *REvise* a solução proposta se necessário, e
- *REtain* a nova solução como parte de um novo caso.

A hierarquia de tarefas em um CBR

A visão mostrada na figura 4 apresenta o problema CBR como uma sequência de etapas. Estas etapas podem ser decompostas em uma visão orientada a tarefas, onde cada etapa ou subprocesso é visto como uma subtarefa que o raciocinador deve executar (Aamodt, 1994). Esta visão permite descrever os mecanismos de forma mais detalhada utilizando-se uma estrutura hierárquica de decomposição onde a descrição do sistema pode ser feita a partir de três perspectivas: tarefas, métodos e modelos de conhecimento de domínio. As tarefas (nomes do nó em negrito) são definidas pelos objetivos do sistema, e uma tarefa é realizada através da aplicação de um ou mais métodos (escritos em itálico). Para um método para ser capaz de realizar uma tarefa, ele precisa de conhecimento sobre o domínio de aplicação geral bem como informações sobre o problema atual e seu contexto. A relação entre as tarefas e métodos (linhas pontilhadas) identificam métodos alternativos aplicáveis para resolver uma tarefa.

Figura 4 - Uma decomposição do método de tarefas de um CBR



Fonte: Aamodt & Plaza 1994

4.2.6 AQUISIÇÃO E REPRESENTAÇÃO DE CASOS

Em Raciocínio Baseado em Casos de acordo com Cordier (2007), a aquisição de conhecimento desempenha um papel importante, pois permite melhorar progressivamente as competências do sistema. Uma das abordagens de aquisição de conhecimento consiste em realizá-lo enquanto o sistema é usado para resolver um problema. Uma vantagem dessa estratégia é que ela não é limitante para o especialista: o sistema explora suas interações para adquirir peças de conhecimentos de que necessita para resolver o problema atual e aproveita a oportunidade para aprender este novo conhecimento para uso futuro. Este processo é apoiado por vários conhecimentos utilizados para raciocinar sobre os casos.

Utilizar CBR como ferramental para aquisição de conhecimento deve ser visto com cuidado haja vista que a aquisição de casos pode ser uma tarefa tão complexa quanto a construção de modelos.

O processo de aquisição e representação de casos deve iniciar com uma análise para determinar o grau de disponibilidade dos casos e esta medida pode ser fundamental para determinar o nível de dificuldade e a estratégia para a construção do sistema. Quanto à disponibilidade os sistemas podem:

Os casos não estão disponíveis em uma fonte externa – neste caso, as descrições dos casos não estão disponíveis em nenhum tipo de registro e apenas o especialista possui a base de casos.

Os casos estão semi-disponíveis em uma fonte externa – neste caso, apesar da existência da base de casos, esta se encontra incompleta, desestruturada ou semiestruturada. Podemos exemplificar a situação em que o especialista tem toda a ficha do paciente, mas não registra toda a evolução da doença. Sendo assim esta base de dados pode ser utilizada como base para a construção do banco de casos e complementada com informações utilizando-se da engenharia do conhecimento.

Os casos estão disponíveis, mas contém erros – os casos são registrados, porém contém informações conflitantes ou erros. Como exemplo podemos citar o registro de evolução do paciente com o mesmo sintoma, mas com diagnósticos diferentes. Assim, como na situação anterior, a construção do banco de casos e complementada com informações utilizando-se da engenharia do conhecimento com apoio do especialista.

Os casos estão disponíveis e corretos – esta é a situação que menos ocorre e geralmente ocorre em domínios cuja solução pode ser descrita através de poucos casos.

Em todas as situações acima citadas, a engenharia do conhecimento, utilizando-se do conhecimento do especialista, deve ser utilizada para definir a melhor forma de representação dos casos, as estratégias de indexação, e a forma de adaptação das soluções recuperadas. (Abel, 1996 apud Strube, 1995).

Após a fase de aquisição do caso deveremos definir a maneira de como os casos serão representados no computador.

De acordo Aamodt (1994) o raciocinador é fortemente dependente da estrutura e conteúdo de sua coleção de casos – muitas vezes referida como memória de casos. O problema da representação de casos no CBR é, principalmente, o problema de decidir o que armazenar em um caso, encontrar uma estrutura adequada para descrevê-lo e decidir como a memória de casos deve ser organizada e indexada para recuperação eficaz e reutilização. Um problema adicional é a forma de integrar a estrutura da memória de casos num modelo de conhecimento geral do domínio, na medida em que tal conhecimento é incorporado.

Dois modelos de memória de casos influenciam fortemente a maneira como os casos são organizados. A forma de armazenamento influi diretamente no planejamento eficiente de um sistema CBR, pois cabe a ela, responder por uma estrutura que suporte eficientes métodos de procura e recuperação na base de casos, esses modelos são: o modelo de memória dinâmica de Schank (1982) apud Aamodt (1994) e o modelo de categoria de exemplares (Porter & Bareiss apud Aamodt 1994).

- **Modelo de Memória Dinâmica**

No modelo de memória dinâmica, o caso é uma estrutura hierárquica denominado de *pacotes de organização de memória episódicos* (E-MOPs). A ideia básica é a de organizar casos específicos que partilham propriedades semelhantes ao de uma estrutura mais geral (um episódio generalizado). O primeiro sistema que pode ser referido como um sistema CBR foi o sistema de Kolodner CYRUS, baseado no modelo de memória dinâmica. De acordo com este modelo, MOPs podem representar o conhecimento de duas maneiras:

- Instâncias – representam casos, eventos ou objetos;
- Abstrações – que representam versões generalizadas de instâncias ou outras abstrações

Este modelo é dito dinâmico porque a inserção de novos casos gera nova E-MOPs a fim de diferenciá-los dos casos gerados anteriormente. O problema é que este processo pode levar a um crescimento absurdo no número de índices com o aumento do número de casos.

- **Modelo de Categoria de Exemplos**

O sistema PROTOS, criado por Ray Bareiss e Bruce Porter, Bareiss (1989), Porter (1990) propôs uma forma alternativa para organização de casos na memória.

Nesta abordagem, casos são referenciados como exemplos de acontecimento, ou seja, este modelo organiza casos baseados na visão do mundo real. Neste modelo, a memória de casos é uma estrutura de rede de categorias, relações semânticas, casos e ponteiros de índice. Cada caso é associado a uma categoria e suas características são importantes para associá-lo a uma categoria. Três tipos de índices são utilizados com a finalidade de apontar para um caso ou categoria:

- *Links de recursos* - que apontam, a partir de descritores de problemas (recursos), para um caso ou categoria;
- *Links de casos* - que apontam, a partir de categorias, para os seus casos associados;
- *Links de diferença* - que apontam, a partir de categorias, para os casos semelhantes que se diferenciam por um pequeno número de características diferentes.

Exemplos de uma categoria são armazenados pelo grau de semelhança com a categoria. Dentro desta organização de memória, as categorias são interligadas dentro de uma rede semântica contendo as características e estados intermediários. Esta rede representa um fundo de conhecimento de domínio geral que permite o suporte explicativo para algumas tarefas RBC. Um novo caso é armazenado a partir da procura de um caso correspondente e pelo estabelecimento de índices característicos relevantes. Se um caso for encontrado com apenas pequenas diferenças para o novo caso, o novo caso não pode ser retido, ou os dois casos podem ser fundidos.

Indexação – indexação de casos refere-se a associar índices a casos para permitir recuperação e comparação de casos relevantes quando na ocorrência de um novo problema. A seguir, apresentaremos os três princípios para indexação de casos (Shi, 2011).

- Índices devem ser relevantes para o domínio concreto. Os índices na base de dados são geralmente suficientes para dividir a base de dados medianamente, assim, a recuperação dos casos pode ser mais rápida. Índices, em um sistema CBR, devem ser predicativos e concretos a fim de simplificar o acesso e recuperação de casos relevantes.

- Os índices devem ser predicativos e abstratos, a fim de permitir o uso futuro da base de casos.

- Os índices devem ser concretos, a fim de serem reconhecidos no futuro.

4.2.7 RECUPERAÇÃO DE CASOS

O processo de recuperação de casos envolve um algoritmo que, a partir de uma descrição do problema e utilizando os índices na memória de casos, deve recuperar o caso mais semelhante para o problema corrente ou situação. Um caso na base de casos contém a descrição do problema e a situação em que ele ocorreu. Quando um novo caso é adicionado à base de casos, índices também são construídos. Quando um novo problema é encontrado, o sistema utiliza o conhecimento de similaridades e índices para recuperar o caso mais semelhante para o problema corrente ou situação. O processo de recuperação influencia a qualidade da solução do problema corrente. O processo de recuperação pode ser dividido em três fases: identificação das características, tentativa do encontro e seleção final.

- **Identificação de características (*feature identification*)** - diz respeito à identificação das características relevantes na análise do problema;

- **Tentativa de busca (*Tentative mathing*)** – significa encontrar um grupo de candidatos na base de casos para o problema corrente;

- **Seleção final (*final selection*)** – é a escolha de um ou mais casos com a maior similaridade com o caso alvo a partir dos casos obtidos na fase de busca.

Recuperação de casos é a maior área de pesquisas em CBR (Sankar, 2004). As técnicas mais utilizadas no processo de recuperação são:

- *Nearest-neighbor retrieval* – nesta técnica, o caso recuperado é escolhido pela soma dos pesos de suas características que estão de acordo com o caso corrente e os casos com maior soma são escolhidos. Em termos gerais, os casos que possuem n características serão recuperados ao invés de um caso com k características,

se $k < n$. Características podem ser ponderadas com pesos e as mais importantes recebem maior peso. O maior problema desta técnica é associar pesos para as características.

- *Inductive approaches* – algoritmos de indução determinam qual a característica do melhor trabalho (*job*) na discriminação dos casos e gera uma árvore de decisão para organizar os casos na memória de maneira eficiente. Esta abordagem é frequentemente utilizada em conjunto com outras técnicas, haja vista o conhecimento explanatório nem sempre está disponível em grandes bases de casos.

- *Template retrieval* – é semelhante a consultas *SQL-like* e retornam todos os casos que possuem certos parâmetros.

Devemos considerar alguns fatores para determinar qual o método a ser adotado para recuperação de casos:

- O número de casos a serem recuperados;
- A quantidade de conhecimento sobre o domínio do problema disponível;
- A facilidade para determinar os pesos das características;
- A maneira como todos os casos devem ser indexados pelas mesmas características ou como os casos podem ter características que variam de importância.

Podemos representar este processo de acordo com a figura 5.

A similaridade é um ponto fundamental no processo RBC, pois a viabilidade do processo ocorre a partir desta etapa na qual verifica-se a similaridade do caso (problema de entrada) com os novos casos. A questão da similaridade entre um caso e outro está nas características que realmente representam o conteúdo e o contexto da experiência.

Segundo Lee (1996) apud Delpizzo (1997), a “métrica de similaridade” é uma função que mede a similaridade entre dois casos e que é utilizada para guiar a busca pelos casos mais similares. De acordo com a função do cálculo do vizinho mais próximo obtêm-se um valor numérico de similaridade onde:

$$\text{Similaridade}(T, S) = \sum_{i=1}^n f(T_i, S_i) \times W_i, \text{ onde}$$

T é o caso de entrada

S é o caso da base

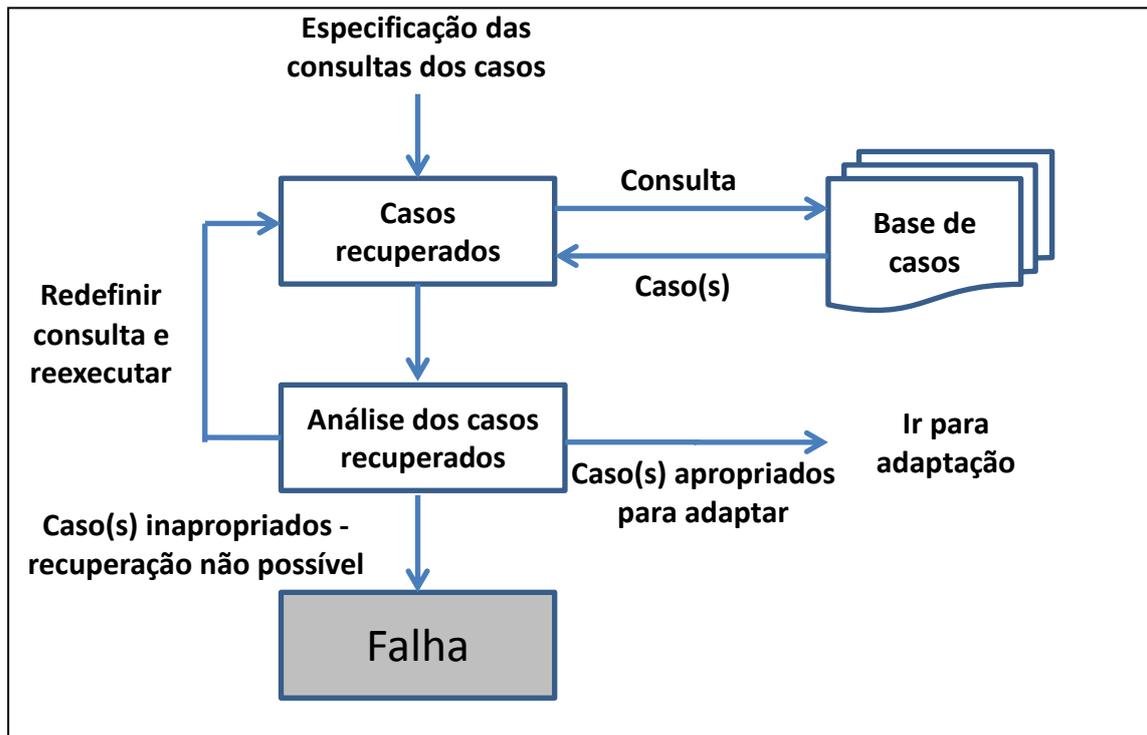
N é o número de atributos de cada caso

i é um atributo individual

f é a função de similaridade para o atributo i nos casos T e S

W é o peso dado ao atributo i

Figura 5 - Processos envolvidos na recuperação de casos



Fonte: adaptado de Sankar, 2004

4.2.8 REUSO

O reuso da solução do caso recuperado no contexto do novo caso está focado em dois aspectos (Aamodt, 1994):

- A quantidade de diferenças entre o caso passado e o caso corrente;
- Que parte do caso recuperado pode ser transferida para o novo caso.

Cópia

Nesta tarefa de classificação simples, as diferenças são abstraídas (elas são consideradas não relevantes enquanto similaridades são relevantes) e a solução do

caso recuperado é transferida para o novo caso como uma classe de solução. Este é um tipo simples de reuso.

Adaptação

Adaptação é o processo de transformação da solução recuperada em uma solução apropriada para o problema corrente. O processo de adaptação percebe diferenças entre o caso recuperado e o caso corrente e então aplica formulas ou regras que identifica estas diferenças e sugere uma solução.

4.2.9 RETENÇÃO

Este é o processo de incorporação que é utilizado para reter a partir de um novo problema solucionado para um conhecimento existente. O aprendizado adquirido com os sucessos ou falhas da solução proposta é alvo de avaliação ou possíveis reparos. Isto envolve selecionar a partir do caso qual informação será retida, que forma retê-lo, como indexar o caso para futuras recuperações a partir de problemas similares e como integrar o novo caso a memória de casos.

Extração

Em CBR, a base de casos é modificada, independentemente de como o caso foi resolvido. Se ele foi resolvido a partir de um caso anterior, um novo caso pode ser criado ou um caso anterior pode ser usado para englobar o novo caso.

As falhas, isto é, informações da tarefa de revisão podem ser extraídas ou retidas separadamente em casos falhos ou em uma base de casos falhos, assim quando uma falha é encontrada, o sistema pode lembrar-se de uma falha semelhante anterior e utilizar para melhorara sua compreensão.

Indexação (*Index*)

O problema da indexação é um dos principais focos dos sistemas CBR. Isto equivale a decidir que tipo de índice em futuras recuperações e como o espaço de busca de índices. Indexar diz respeito a associar índices aos casos a fim de facilitar o processo de recuperação dos casos e comparações futuras.

Segundo Watson (1994), várias diretrizes sobre indexação foram propostos por pesquisadores em RBC, tais como:

- Ser preditivo, ou seja, permitir futuras utilizações da informação para solucionar diferentes casos;
- Endereçar para quais finalidades caso possa ser utilizado;
- Ser abstrato o suficiente para ser utilizado em diversas situações.

Alguns raciocinadores usam apenas um caso ao passo que outros usam uma combinação de casos e casos generalizados.

- Ser concreto o suficiente para ser reconhecido em futuras utilizações.

Integração

Este é o passo final da modificação da base de conhecimento com o novo caso conhecido. Se um novo caso e um conjunto de índices não foram construídos, este é o principal passo da retenção. Por modificar a indexação de casos existentes, sistemas CBR aprendem a se tornar melhores avaliadores das similaridades através do ajuste do poder do índice de acordo com o sucesso ou insucesso na recuperação dos casos.

4.3 RACIOCÍNIO BASEADO EM CASOS APLICADO A SAÚDE

Nos últimos anos, o desenvolvimento de aplicações CBR na área de saúde tem crescido significativamente. Devido a sua característica de reuso de problemas para a solução de novos problemas, a metodologia de CBR tem sido adotada como estratégia de resolução dos mais diversos problemas na área de saúde. De acordo com Bichindaritz (2011), a biomedicina tem sido uma das áreas mais frutíferas para aplicação da metodologia CBR, embora também seja uma das mais complexas. Além da biomedicina, Bichindaritz (2011), também destaca a tendência crescente do desenvolvimento de aplicações nas áreas de diagnóstico bem como tratamento e controle de qualidade. Na área de biomedicina, podemos destacar o projeto eXiT*CBR do grupo *Control Engineering and Intelligent Systems* da Universidade de Girona, Girona (2014), que desenvolveu um *framework* para inferir modelos de câncer de mama usando dados do Instituto Catalão de Oncologia, que resultou em uma ferramenta para

facilitar o diagnóstico médico utilizando métodos CBR. Na área de hemodiálise, o sistema RHENE (*Retrieve HEmodialysis NEphrological disorders*) trabalha no domínio de paciente afetados com nefropatologias (Montani, 2004). Para casos de diabetes, Marling (2008) apresenta uma abordagem baseada em casos para apoio a decisão na administração de tratamento com insulina em paciente com diabetes tipo 1.

A integração de CBR com Prontuário Eletrônico do Paciente foi aplicada ao EXCELICARE, Van Den Branden (2011), sistema PEP utilizado na Inglaterra. O EXCELICARECBR é um sistema CBR que foi desenvolvido para complementar o EXCELICARE, cujo objetivo é integrar CBR para apoiar a tomada de decisão clínica através do aproveitamento de registros eletrônicos de pacientes.

De acordo com o apresentado, a abordagem CBR vem se tornando uma eminente área de estudos aplicada a ciências de saúde com o objetivo de apoiar o processo de tomada de decisão médica.

4.3.1 CBR COMO APOIO A TOMADA DE DECISÃO MÉDICA

A tomada de decisão médica é uma tarefa complexa que muitas vezes exigem dos profissionais de saúde tomar decisões baseadas em informações médicas dos pacientes imprecisas ou incompletas. O processo de tomada de decisão médica combina conhecimento tácito, aquele adquirido ao longo do exercício da atividade médica acumulados a partir de atendimentos realizados anteriormente, juntamente com conhecimento explícito, que são conhecimentos catalogados e armazenados em alguma mídia. Estes conhecimentos representam uma valiosa base de conhecimento a partir da qual os profissionais de saúde recorrem para construir o processo de tomada de decisões a fim de definir um melhor tratamento para o paciente.

A estes conhecimentos, tácitos e explícitos, sistemas de apoio à decisão médica podem agregar conhecimento aos especialistas no processo de tomada de decisão para, por exemplo, selecionar um tratamento específico para um paciente. A abordagem CBR tem sido proposto como base para sistemas de apoio a decisão médica por sua capacidade de lidar com informações incompletas e imprecisas, fazer relacionamentos não triviais por meio de medidas de similaridade, bem como permitir a utilização de atributos numéricos e simbólicos.

Em sistemas de apoio à decisão baseado em casos, um caso é descrito como uma situação passada na qual uma decisão foi tomada. No domínio médico, um caso contém uma descrição de sintomas observados durante a os exames do paciente bem como o diagnóstico definido para o caso apresentado.

Althoff (1998) apresenta alguns requisitos para sistemas de apoio à decisão médica, quais sejam:

- Curto tempo de resposta – considerado o requisito mais importante devido a criticidade, o sistema estar habilitado para dar o diagnóstico e apresentá-lo em um curto intervalo de tempo.
- Justificativa dos resultados – o resultado deve ser justificado pelo sistema de uma forma que o profissional possa validar e julgar sua precisão.

Ainda de acordo com Althoff (1998), desenvolver aplicações CBR geralmente requer definir a área de competência, o propósito do sistema (*call center*, suporte, apoio a ambulância, educação dentre outras áreas) e o público alvo (estudantes, profissionais pouco experientes ou profissionais especialistas). O próximo passo é definir a estrutura do sistema CBR levando em consideração os seguintes itens:

- Definir uma representação de casos apropriada, ou seja, selecionar atributos para os casos;
- Definir medidas de similaridades e os respectivos pesos;
- Coletar os casos.

Diferentemente dos sistemas de suporte clínico, de acordo com Van Bemel (1993), como por exemplo sistemas de imagens que já são bem aceitos pela comunidade médica, por oferecer os dados que o profissional precisa para a tomada de decisão, sistemas de apoio à decisão médica não são aceitos ou são apenas parcialmente aceitos pela mesma comunidade. Ainda de acordo com Van Bemel (1993), os sistemas de apoio à decisão médica podem ser utilizados com sucesso em conjunto com sistemas de apoio à decisão clínica para aumentar o confiabilidade dos dados primários, possibilitando uma maior consistência e completude.

4.3.2 CBR COMO FERRAMENTA DE APRENDIZADO

Estratégias de ensino e aprendizagem baseado em casos podem ser utilizados para auxiliar o aprendizado médico tanto no que diz respeito ao apoio ao

estudante na obtenção de conhecimentos necessários à atividade médica, mas para melhorar o raciocínio clínico.

De acordo com Thomas et al (2001), a aprendizagem baseada em casos traz alguns benefícios, tais como:

- Melhora a organização da informação de uma forma que pode ser recuperado para uso em raciocínio clínico – o raciocínio clínico é melhorado devido a organização apropriada do conhecimento;
- Gera experiências que os alunos não teriam de outro modo – experiências anteriores influenciam na decisão do que é considerado significativo ou que são considerados como sintomas de uma doença;
- Aumenta a visibilidade do raciocínio clínico nos alunos - O discurso ou fluxo de comunicação em relação a casos reflete processos de raciocínio clínico do aluno;
- Melhora a confiança do aluno – aprendizado baseado em casos proporciona uma experiência que pode contribuir para a confiança do aluno.

A utilização da abordagem CBR no aprendizado em medicina tem sido explorada em Eshach (2003) que examina a relação entre PBL (*Problem-based Learning*) e CBR como metodologias complementares que podem ser utilizadas com sucesso no ensino da medicina. Welter (2011) e colegas, em seu artigo, apresenta a ferramenta para diagnóstico baseada em computador (CAD – *Computer Aided Diagnostic*). O ICBR-RE (*Image-based Case Retrieval for Radiological Education*) combina teorias do aprendizado de adultos, raciocínio baseado em casos, recuperação de conteúdo de imagem (CBIR - *Content-based image retrieval*). O ICBR-RE é um sistema para treinamento de diagnóstico incorporado a um framework baseado na metodologia para PBL Seven Jump a fim de proporcionar um ambiente de aprendizado que seja o mais semelhante possível à prática radiológica.

Neste contexto a combinação de CBR com outras metodologias e ferramentas surge como uma nova ferramenta para o processo de treinamento e aprendizagem em medicina.

5 MINERAÇÃO DE DADOS

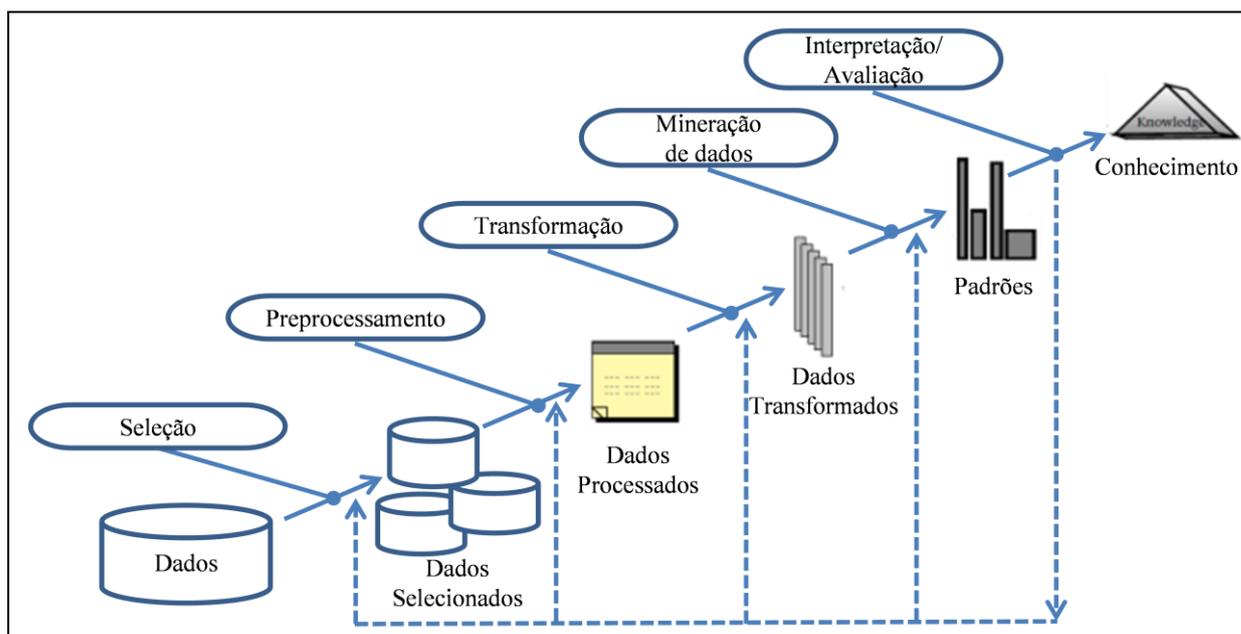
5.1 INTRODUÇÃO

A utilização crescente da tecnologia de banco de dados tem provocado um grande crescimento da disponibilidade de dados tanto nas áreas governamentais, comerciais, indústria, internet e nos mais diversos ramos de atividades. Esta crescente quantidade e disponibilidade de dados têm aumentado a dificuldade de recuperação e interpretação dos mesmos gerando a necessidade de novas ferramentas e técnicas para análise inteligente e em tempo hábil permitindo assim transformar dados estáticos em informações e conhecimento útil para auxiliar no processo de tomada de decisão.

5.2 MINERAÇÃO DE DADOS

Mineração de dados, conforme Ian (2005) é o processo de descobrir padrões em dados. O processo pode ser automático ou semiautomático. Os padrões descobertos devem ser significativos para o processo de tomada de decisão. Esta descoberta de conhecimento tanto pode ser direcionada ao usuário, onde o conhecimento do usuário, juntamente com a análise dos dados são utilizados para produzir algum conhecimento, ou orientados a dados, onde algoritmos são aplicados para produzir conhecimento. Mineração de dados é uma parte do processo de descoberta de conhecimento em uma base de dados.

Figura 6 - Passos que Compõem o Processo de KDD



Fonte: adaptado de Fayyad (1996)

De acordo com Fayyad (1996), KDD (*Knowledge Discovery in Database*) refere-se a todo o processo de descoberta de conhecimento útil a partir de uma base de dados na qual a mineração de dados uma fase deste processo. O processo KDD pode ser dividido em nove passos, conforme a figura 6.

1. Desenvolver um entendimento do domínio da aplicação a partir dos conhecimentos anteriores relevantes para determinar o conjunto de dados a serem explorados e os objetivos do usuário final. Isto pode necessitar a obtenção de direitos do uso dos dados e considerações éticas para exploração dos mesmos;

2. Criar o conjunto dos dados a partir dos dados selecionados para o processo de descoberta;

3. Aplicar o processo de limpeza e pré-processamento através da remoção de ruídos, dados omitidos ou que não sejam importantes para o processo de descoberta;

4. Reduzir quantidade de dados de acordo com os objetivos da tarefa;

5. Escolher o método de mineração de dados;

6. Escolher o algoritmo a ser utilizado na pesquisa de padrões na base de dados;

7. Procurar por padrões de interesse em uma forma representacional particular ou um conjunto de tais representações incluindo regras de associação, árvores de decisão, regressão e *clusterização*.

8. Interpretar os padrões minerados, podendo retornar os passos de 1 a 7 para novas interações.

9. Utilização do conhecimento adquirido incorporando o conhecimento a outros sistemas para novas interações ou simplesmente documentar e reportar as partes interessadas.

Os processos KDD podem envolver interações significantes e pode conter loops entre dois ou mais passos.

5.2.1 MÉTODOS

Baseado nos objetivos do processo de mineração de dados há vários métodos que podemos classificar de acordo com a entrada e a caracterização do algoritmo os quais pode ser de padrões de **predição** e **descrição**. A predição diz respeito à utilização dos dados do banco de dados para prever valores desconhecidos ou

valores futuros das variáveis, enquanto a descrição preocupa-se em encontrar padrões humanos interpretáveis a fim de descrever os dados. Esta classificação não é rígida, podendo os métodos serem utilizados para predição ou descrição de acordo com os objetivos da descoberta.

Os métodos utilizados em mineração de dados são:

- *Classificação* – é o método que mapeia (classifica) dado de entrada em um conjunto finito de categorias (classes) pré-definidas. O objetivo de um algoritmo de classificação é encontrar algum relacionamento entre os atributos e uma classe, de modo que um processo de classificação possa usar esse relacionamento para prever a classe de um exemplo novo e desconhecido. Árvore de decisão são exemplos de algoritmos de classificação.

- *Regressão* – o método de regressão possui as mesmas características que o método de classificação sendo que na regressão o dado de entrada não é um valor discreto e sim contínuo (intervalo). Redes neurais e regressão linear são exemplos de algoritmos de regressão.

- *Agrupamento (cluster)* – é uma típica tarefa descritiva cujo objetivo é encontrar categorias de objetos que são similares. Podemos citar como exemplo o algoritmo minimax.

- *Sumarização* - envolve métodos para encontrar uma descrição compacta para um subconjunto de dados. Técnicas de sumarização são frequentemente aplicados para análise exploratória interativo de dados e geração de relatórios automatizados.

- *Regras de associação* – o propósito da regra de associação é encontrar uma relação escondida dentro de diferentes conjuntos de dados, ou seja, a regra de associação permite verificar quanto a ocorrência de um conjunto de itens em um banco de dados implica a ocorrência de outro conjunto de itens nos mesmos registros. O conceito de regras de associação foi proposto por Agrawal, Imielinsk Swami (Agrawal, 1994). Uma regra de associação é definida como *se X então Y*, ou $X \rightarrow Y$, onde X e Y são conjuntos de itens e $X \cap Y = \emptyset$. Diz-se que X é o antecedente da regra, enquanto Y é o seu conseqüente.

5.3 OS COMPONENTES DE UM ALGORITMO DE MINERAÇÃO DE DADOS

De um modo geral, algoritmos de mineração de dados possuem três componentes básicos: (1) representação do modelo, (2) avaliação do modelo e (3) pesquisa.

Representação do modelo – diz respeito à linguagem utilizada para descrever os padrões de descoberta.

Avaliação do modelo – a avaliação do modelo baseia-se em métricas que definem o quanto um determinado padrão atende aos objetivos da descoberta de conhecimento tais como grau de novidade, utilidade, validade, precisão e compreensão.

Método de pesquisa – consiste de dois componentes: (1) pesquisa baseada em parâmetros (*parameter search*) e pesquisa baseada em modelo (*model search*). O primeiro procura por parâmetros a partir de dados de um modelo fixo, o segundo avalia diferentes possibilidades de modelos de acordo com a base de dados.

5.4 MINERANDO PADRÕES DE SEQUÊNCIAS

O avanço da tecnologia de banco de dados tem permitido o armazenamento de grandes bases de dados nos SGBD's. Uma dos desafios da mineração de dados é identificar padrões de comportamento baseado nos registros do banco de dados para apoiar o processo de tomada de decisão. O exemplo clássico para este tipo de problema é determinar que quando um produto x , ou conjunto de produtos x , é comprado em determinada compra, o produto y , ou conjunto de produtos y , também é.

5.4.1 DECLARAÇÃO DO PROBLEMA

Dado um uma base de dados \mathcal{D} , com transações geradas pelo sistema, por exemplo, transações de compras de clientes, o problema da mineração de dados utilizando regras de associação é gerar todas as regras de associação que tem suporte e confiança maior que um suporte e confiança mínimos definido pelo usuário, denominados de *minsup* e *minconf* respectivamente. Tais medidas de interesse são calculadas de acordo com as regras abaixo:

Confiança:

Seja uma regra $A \rightarrow B$ (A implica em B)

A confiança da regra é dada por:

$$\text{Confiança } (A \rightarrow B) = \frac{\# \text{ tuplas contendo tanto } A \text{ quanto } B}{\# \text{ tuplas contendo } A}$$

Ou seja, uma confiança de 85% (0,85) da regra *compra (homem, cerveja) → compra (homem, amendoim)*, significa que 85% dos homens que compram cerveja também compram amendoim.

Suporte

No suporte a regra é dado por:

$$\text{Suporte } (A \rightarrow B) = \frac{\# \text{ tuplas contendo tanto } A \text{ quanto } B}{\# \text{ total de tuplas}}$$

Ou seja, um suporte de 5% significa que, de todas as transações realizadas, 5% são efetuadas por homens que, comprando cerveja também compraram amendoim.

5.4.2 ALGORITMO APRIORI

Um *itemset* de acordo com Agrawal (1994) é um conjunto de itens não vazio. Uma sequência (*sequence*) é uma lista ordenada de *itemsets*. Nós denotamos um *itemset* i por $(i_1, i_2, i_3, \dots, i_n)$, onde i_j é um item e uma sequência s por $(s_1, s_2, s_3, \dots, s_n)$, onde s_j é um *itemset*.

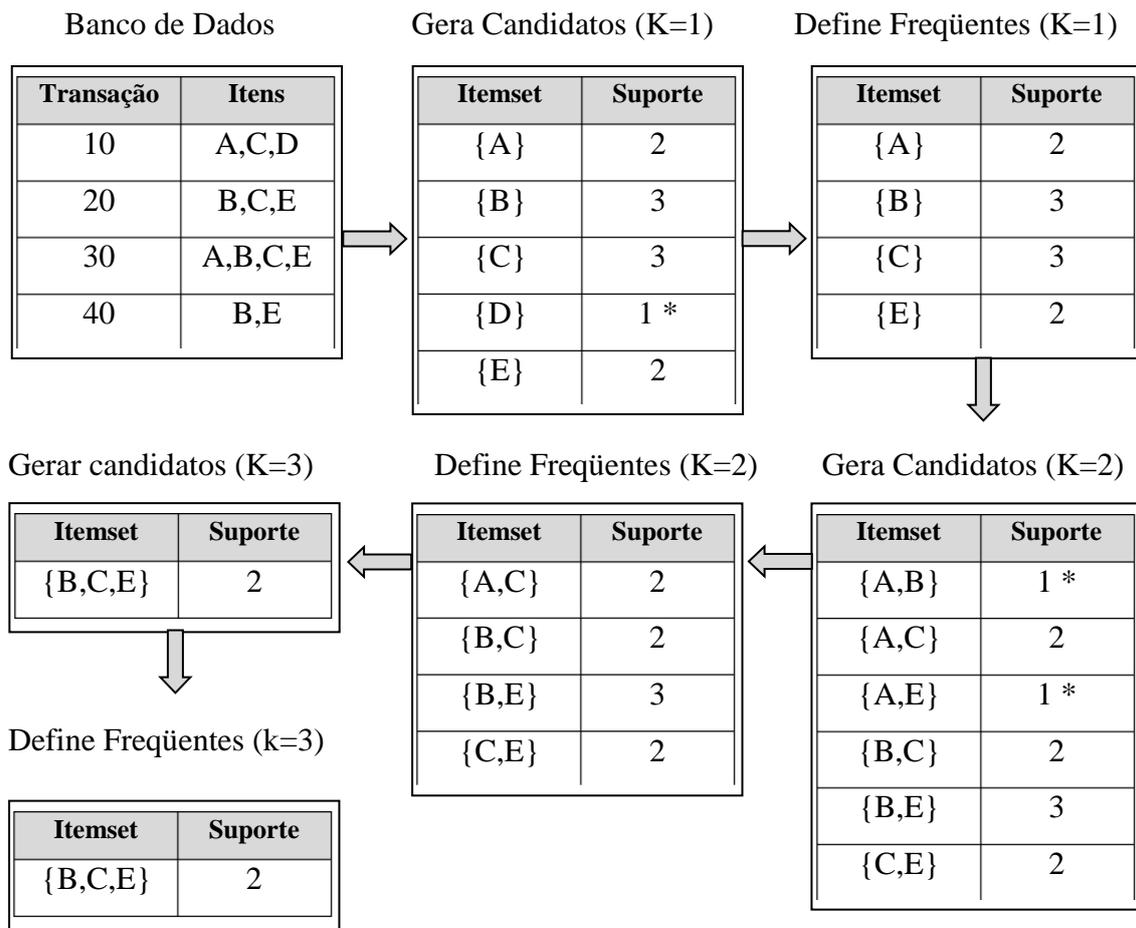
O algoritmo Apriori é o método mais utilizado na mineração de padrões frequentes em base de dados. O algoritmo realiza um processo iterativo no qual, em cada passo, utiliza-se a sequência do passo anterior para gerar as sequência candidatas e, em seguida, medir seu suporte a partir de uma varredura sobre a base de dados. No passo final, o suporte dos candidatos é utilizado para determinar a sequência. No primeiro passo, a saída do *itemset* é utilizado para o conjunto do *1-sequence*, assim os *itemsets* de tamanho k são calculados a partir dos *itemsets* frequentes de tamanho $k-1$.

O ciclo Apriori consiste em:

- Calcular o suporte de todos os *itemsets* possíveis de tamanho l e, em seguida, eliminar os que não atendem ao suporte mínimo;

- A partir dos *itemsets* considerados frequentes na etapa anterior, criar *itemsets* candidatos de tamanho 2 e, novamente, eliminar os que não atendem ao suporte mínimo.
- Repetir o processo até que nenhum *itemset* seja considerado frequente na etapa $k-1$, não gerando, portanto candidatos na etapa k .

Figura 7 - Funcionamento de Algoritmos Apriori com suporte mínimo de 50%



Fonte: o autor

A figura 7 exemplifica a estratégia do algoritmo Apriori para um suporte de 50%. Cada iteração k gera k -*itemsets* como resultado, ou seja, conjuntos de itens frequentes de tamanho k . A solução para o exemplo mostrado na Figura é: 1-*itemset*: {A}, {B}, {C}, {D}, {E}; 2-*itemset*: {A,C}, {B,C}, {B,E}, {C,E}; 3-*itemset*: {B,C,E}.

Isto significa, por exemplo, que em um banco de dados de transação de compras, 50% dos clientes que compraram B também compraram C e E.

No exemplo da figura 8 utilizamos um *itenset* de compras de produtos através de pedidos. Neste *itenset*, temos um total de 13 itens comprados em quatro compras definidas pelo número do pedido (10, 20, 30, 40). Estes produtos fazem parte de um conjunto de cinco produtos diferentes: cerveja, fralda, biscoito, laranja e sabonete.

Figura 8 - Itenset Itens do Pedido

```

select i.num_pedido numero_pedido,
       p.nome_produto nome_produto
from produto p,
     item_pedido i
where p.cod_produto = i.cod_produto
  
```

	NUMERO_PEDIDO	NOME_PRODUTO
▶ 1	10	Cerveja
2	10	Fralda
3	10	Biscoito
4	20	Fralda
5	20	Sabonete
6	20	Laranja
7	30	Cerveja
8	30	Fralda
9	30	Sabonete
10	30	Laranja
11	40	Cerveja
12	40	Fralda
13	40	Laranja

Fonte: o autor

Aplicando-se o algoritmo Apriori com parâmetros de suporte de 50% e confiança mínima de 70% obtemos nove regras de acordo com a figura 9.

Figura 9 - Aplicação do Algoritmo Apriori no *itenset* da fig. 8

ID	Antecedente	Consequente	Levantamento	Confiança(%)	Suporte(%)	Contagem de Itens	Suporte de Antecedente(%)	Suporte Consequente(%)
3	Laranja	Fralda	1,0000	100	75	1	100	75
4	Fralda	Laranja	1,0000	75	75	1	75	100
2	Fralda	Cerveja	1,0000	75	75	1	75	100
1	Cerveja	Fralda	1,0000	100	75	1	100	75
8	Laranja AND Sabonete	Fralda	1,0000	100	50	2	100	50
7	Cerveja AND Laranja	Fralda	1,0000	100	50	2	100	50
5	Sabonete	Fralda	1,0000	100	50	1	100	50
9	Fralda AND Sabonete	Laranja	1,3333	100	50	2	75	50
6	Sabonete	Laranja	1,3333	100	50	1	75	50

Fonte: o autor

A partir das regras apresentadas, podemos fazer algumas análises tais como:

- De acordo com a regra 3, no total de 75% das compras (suporte), 100% (confiança) dos clientes que compraram laranja também compraram fraldas (antecedente laranja e consequente fralda), mas o contrário não ocorre, pois apenas 75% (confiança) dos clientes que compraram fraldas também compraram laranja (regra 4).
- O mesmo ocorre com os produtos cerveja e fralda.
- Com medida de confiança de 100%, clientes que compraram laranja e sabonete também compraram fraldas, sendo que a mesma relação ocorreu em 50% (suporte) das compras conforme apresentado na fig. 10 da relação *If...Then* da relação 8.

Figura 10 - Apresentação da relação *If ... Then* da regra 8

ID	Antecedente	Consequente	Levantamento	Confiança(%)	Suporte(%)	Contagem de Itens	Suporte de Antecedente(%)	Suporte Consequente(%)
2	Fralda	Cerveja	1,0000	75	75	1	75	100
1	Cerveja	Fralda	1,0000	100	75	1	100	75
8	Laranja AND Sabonete	Fralda	1,0000	100	50	2	100	50
7	Cerveja AND Laranja	Fralda	1,0000	100	50	2	100	50
5	Sabonete	Fralda	1,0000	100	50	1	100	50

Detalhes da ...	
IF	
Laranja AND Sabonete	
THEN	
Fralda	
Levantamento	1
Confiança (%)	100
Suporte (%)	50
Suporte de Antecedente (%)	100

Fonte: o autor

5.5 INTEGRAÇÃO DE CBR E MINERAÇÃO DE DADOS

CBR e mineração de dados Mythili (2011) são métodos usados para o apoio à decisão, para organizar e processar informações a fim de disponibilizá-lo para melhorar a qualidade das decisões a serem tomadas por seres humanos dentro de uma organização ou por um sistema de computador.

CBR depende muito da qualidade dos coletados dados, a quantidade de dados, a quantidade de conhecimento de fundo e de uma maneira de comparar os casos para decidir qual é a mais similar. O método é o mais adequado para domínios que mudam, e onde temos pouco conhecimento de processos subjacentes que regem o domínio.

Data Mining é uma forma de extração de informações de bancos de dados e pode assim ser usado para extrair informações que sejam relevantes para uma situação-problema. Também poderia ser usado para encontrar situações problemáticas não vividas a partir de um banco de dados e representá-lo como um caso, possivelmente através da interação com um utilizador. Mineração de dados pode-se inferir as regras, as classificações e gráficos a partir dos dados que podem ser usados como o conhecimento de fundo no sistema de RBC, e também para calcular a semelhança entre os casos.

De acordo com Jansen (2012) a descoberta de conhecimento em Prontuário Eletrônico do Paciente oferece uma abordagem para gerar conhecimento específico e clinicamente útil para apoiar o tratamento de paciente. Tomada de decisão clínica é uma tarefa complicada, em que o prestador deve decidir entre as informações que possui em mãos relativas a um determinado caso e o que é necessário decidir sobre o melhor tratamento.

A aplicação de métodos de descoberta de conhecimento a partir dos dados armazenados no PEP envolve um conjunto de métodos estatísticos, de aprendizado de máquina e computacionais a fim de dar suporte ao processo de tomada de decisão.

Alguns algoritmos de mineração de dados exigem conhecimento de fundo, que pode ser tomada a partir de um sistema CBR. Ao integrar os métodos, esperamos fazer melhor uso da informação, e que isso pode levar a um crescimento em sistemas de informática também.

Combinar CBR e DM, de acordo com Aamodt (1998) requer a adoção de algumas atividades tais como:

- Modelar o domínio do problema;
- Estabelecer parâmetros e regras para os casos de indexação para recuperação eficiente de casos relevantes;
- Seleção de métodos para avaliar a semelhança entre um novo problema e os casos passados;
- Testar e melhorar a medida de similaridade por meio de um processo de aprendizagem;
- Reter e generalizar um novo caso, bem como eles são resolvidos;

A ideia principal é utilizar as duas tecnologias que a partir de dados armazenados em um banco de dados, reutilizar experiências de casos anteriores em

situações problemas para extrair informações relevantes para o aprendizado e tomada de decisão.

Integração CBR e mineração de dados pode ser encontrado em recentes publicações tais como Branden (2011) que mostra o EXCELCARECBR, um sistema desenvolvido para complementar o EXCELCARE. Zhuang (2009) que apresenta uma nova metodologia para integrar mineração de dados e CBR, Jensen (2012) no qual pesquisa correlações entre doenças.

6 O MODELO DE APOIO A PRESCRIÇÕES MÉDICAS (MAPMED)

6.1 INTRODUÇÃO

O Modelo de Apoio a Prescrições Médicas baseado em Mineração de Dados, é um modelo de um ambiente apoiado por computador desenvolvido com o objetivo de auxiliar prestadores (médicos, enfermeiros) e estudantes da área de medicina na prescrição de itens de tratamento a partir de dados registrados no Prontuário Eletrônico do Paciente – PEP, além de auxiliar administradores hospitalares no processo de tomada de decisão.

O modelo MAPMED foi desenvolvido utilizando-se a abordagem de Raciocínio Baseado em Casos (CBR) em combinação com a tecnologia de Mineração de Dados (DM), haja vista que:

- CBR, de acordo com Aamodt (1998) possibilita a recuperação de dados relevantes, comparando situação problema atual do usuário com situações anteriores (casos). Os casos são de conhecimento específico da situação, armazenados em uma base de experiência, juntamente com o conhecimento geral necessário (hierarquias conceituais, relacionamentos, regras de decisão, associações etc.).
- DM é o conhecimento "criador" que, por meio do acesso ao banco de dados, através da aplicação dos métodos de mineração, retorna e informações relevantes a partir de diversas bases de dados.

Desta forma, a combinação, CBR e DM, preenchem a lacuna para a descoberta de dados ocultos e dificilmente compreensíveis em bases de dados existentes para atender as necessidades dos usuários. As seleções de dados, agregações, e abstrações resultantes do processo de mineração de dados são colocadas em uma forma adequada para raciocínio baseado em casos.

6.2 CARACTERIZAÇÃO DO CASO

Um novo caso inicia-se quando da ocorrência de um atendimento, no qual, o paciente relata seu quadro (sintomas), e o prestador registra em um prontuário eletrônico, gerando o que denominamos de “evolução médica”. A partir do quadro relatado e dos dados obtidos nos procedimentos básicos de atendimento, tais como

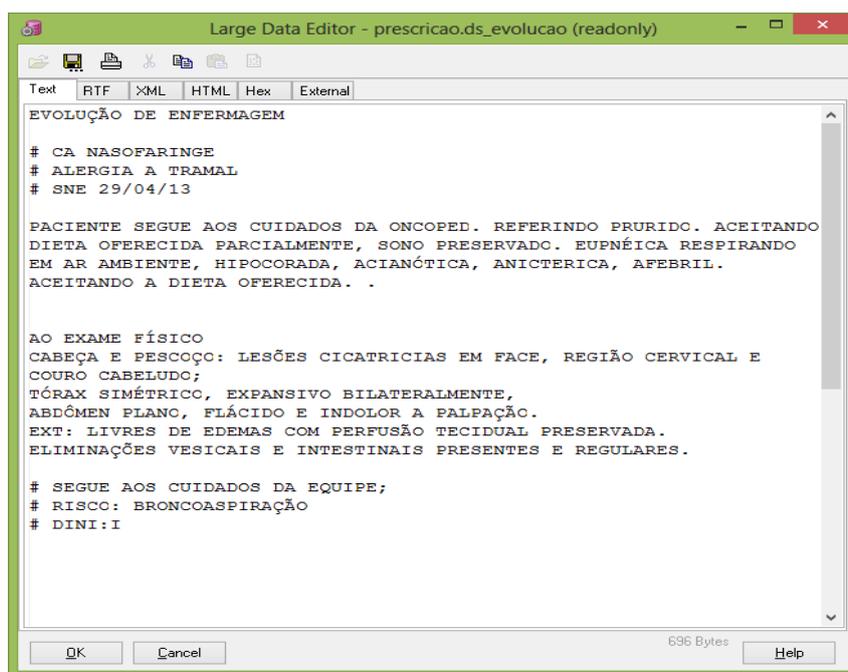
coleta de sinais vitais, exame físico, histórico recente do paciente, histórico familiar e outras intercorrências, como mostrado nas figuras 11 e 12, que apresentam a evolução médica de pacientes com quadro de *carcinoma de nasofaringe e linfoepitelioma*. Estas informações são devidamente registradas na evolução do paciente e, de acordo com os dados coletados, o prestador deverá prescrever itens para tratamento conforme o quadro apresentado. Tais itens podem ser: medicamentos, exames de laboratório ou de imagens, dietas, materiais, órtese/prótese, cuidados, fisioterapia, gases, tratamento radioterápico ou quimioterápico bem como outros procedimentos médicos relacionados ao tratamento.

Logo, um caso caracteriza-se pelo conjunto EVOLUÇÃO (quadro clínico) e itens prescritos.

Caso = {evolução, itens prescritos}

Baseado em casos que contêm palavras-chaves ou termos técnicos relacionados à ocorrência de oncologia pediátrica, de acordo com os grupos de câncer pediátrico apresentados no item 3.6, o problema está em descobrir se há alguma similaridade com relação aos itens prescritos entre estes casos e em que grau esta similaridade ocorre em evoluções que contêm estas palavras-chaves.

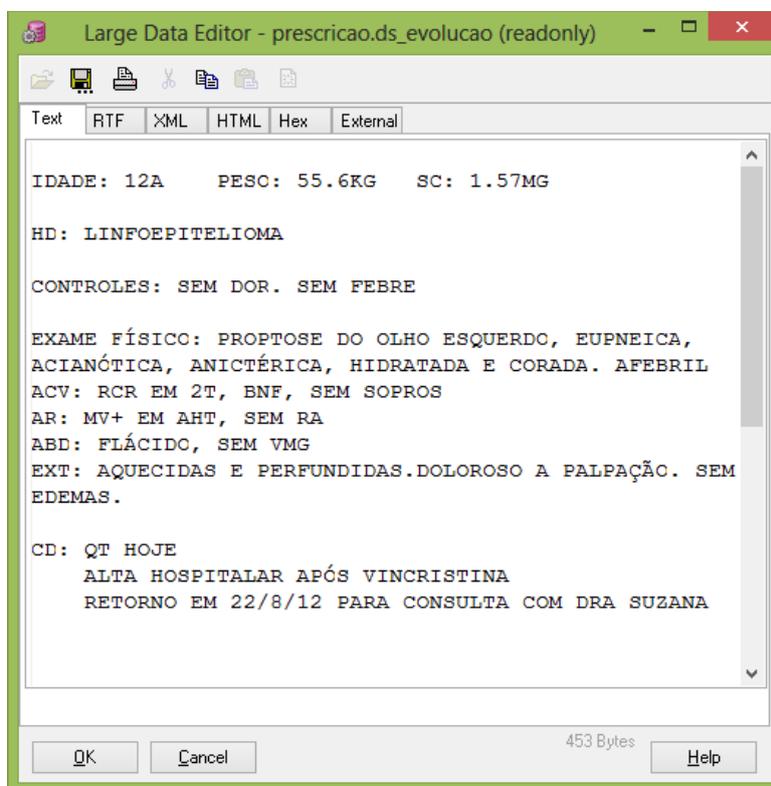
Figura 11 - Evolução de paciente com quadro de carcinoma de nasofaringe



Fonte: o autor

De acordo com os dados da evolução, registrados no prontuário eletrônico do paciente, o prestador prescreve o tratamento adequado àquele quadro.

Figura 12 - Evolução de paciente com quadro de carcinoma de linfoepitelioma



Fonte: o autor

O Modelo de Apoio a Prescrições Médicas baseado em Mineração de Dados foi desenvolvido em módulos permitindo uma visão mais estruturada do ciclo de vida do sistema.

6.3 MODELO FORMAL

Seja $I = I_1, I_2, \dots, I_n$ o conjunto de atributos binários denominados itens. Seja T um banco de dados de transações. Cada transação t é representada como um vetor binário no qual $t[k] = 1$ se o item i_k foi prescrito e $t[k] = 0$ se o item i_k não foi prescrito. Há uma *tupla* registrada no banco de dados para cada transação. Seja X um conjunto de alguns itens em I . Dizemos que uma transação t satisfaz X se, para todos os itens I_k em X , $t[k] = 1$.

De acordo com Agrawal (2003), por uma regra de associação (AR-Association Ruler), queremos dizer uma implicação da forma $X \rightarrow I_j$, onde X é o

conjunto de alguns itens de I , e I_j é um único item que não está presente em X . A regra $X \rightarrow I_j$ é satisfeita no conjunto de transações T com fator de confiança $0 \leq c \leq 1$ se pelo menos $c\%$ das transações em T que satisfazem X também satisfazem I_j . A notação $X \rightarrow I_j \mid c$ será utilizada para especificar que a regra $X \rightarrow I_j$ tem o fator de confiança c . X é o antecedente e Y é o conseqüente da regra $X \rightarrow Y$.

Seja X o item prescrito (antecedente) e Y o item prescrito (conseqüente). Queremos mostrar que em quadros clínicos que contém palavras-chaves semelhantes, quando o item X é prescrito, o item Y também é prescrito.

A medida de suporte $X \rightarrow Y$, $sup(X \rightarrow Y)$, corresponde ao quociente de *tuplas* que contém tanto X quanto Y em relação ao total de *tuplas* do *dataset*, onde:

$$0 \leq sup(X \rightarrow Y) \leq 1$$

Quanto mais próximo de 1, (100%), as medidas de suporte tornam-se mais confiáveis.

$$supp(X \rightarrow Y) = supp(X)$$

A medida de confiança de $X \rightarrow Y$ é o quociente da quantidade de *tuplas* que contém tanto X quanto Y em relação à quantidade de *tuplas* que contém X .

$$conf(X \rightarrow Y) = sup(X \rightarrow Y) / sup(x),$$

$$0 \leq conf(X \rightarrow Y) \leq 1$$

Da mesma forma que a medida de suporte, quanto mais próximo de 1, (100%), melhor a medida de confiança.

Dado o conjunto de transações T estamos interessados em gerar todas as regras que obedecem as duas regras a de seleção dos itens:

Regras de Sintaxe – regras que envolvem restrições em que itens devem aparecer em uma regra, como, por exemplo, regras em que determinado item I_x deve aparecer como conseqüente ou regras em que um item I_y deve aparecer como antecedente. Também regras em que um conjunto de itens X pode aparecer como antecedente ou algum conjunto de itens Y pode aparecer como conseqüente. A esta regra denominaremos “Contagem de Itens”.

Exemplo: Seja um caso de CA de NASOFARINGE

Contagem de itens = 2

Confiança = 89,6552%

Regra:

Antecedente:

Avaliar o acesso venoso durante a infusão

And

Oferecer medidas de conforto durante a infusão

Consequente

Observar hipiremia de trajeto venoso

Analisando esta regra concluímos que em 89,6552% das prescrições em que são prescritos os itens “Avaliar o acesso venoso durante a infusão” e “Oferecer medidas de conforto durante a infusão” também é prescrito o item “Observar hiperemia de trajeto venoso”, Ou seja:

{Avaliar o acesso venoso durante a infusão, Oferecer medidas de conforto durante a infusão} → {Observar hiperemia de trajeto venoso} | 89,6552%

Regra do Suporte – esta regra diz respeito ao número de transações T que suportam uma regra. O suporte de uma regra é definida como a fração de transações em T que satisfazem a união de itens no antecedente e consequente da regra. Se o suporte não é grande o suficiente, isso significa que a regra não é digna de consideração ou que é simplesmente pode ser ignorada.

No caso acima o suporte é de 35,6164%, ou seja, esta regra aparece em apenas em 35,6164% das transações do banco de dados.

6.4 VISÃO GERAL DA ARQUITETURA

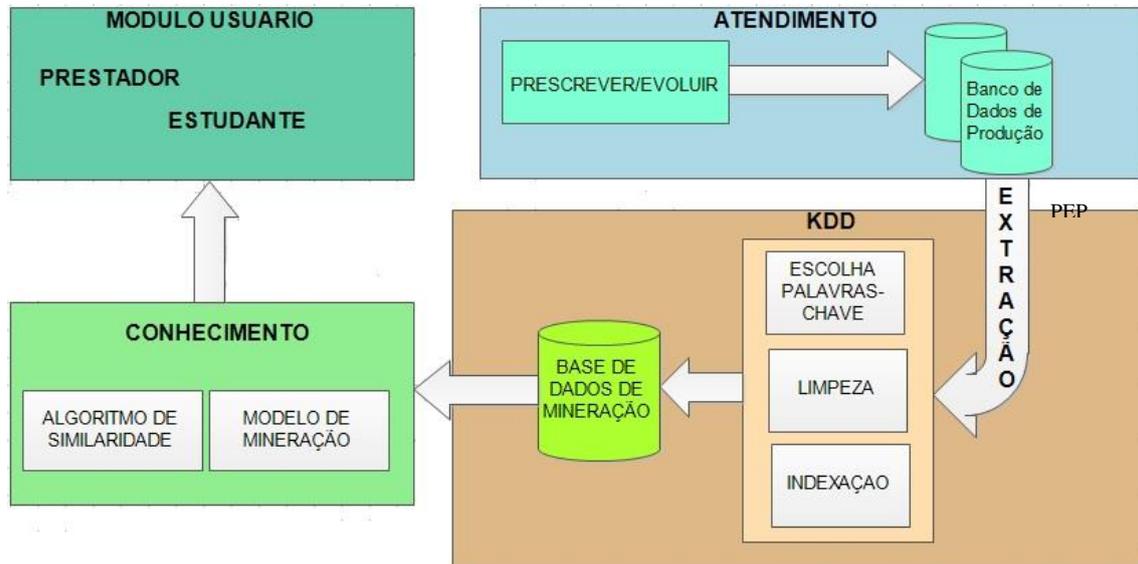
A figura 13 apresenta um esboço geral do fluxo de dados para a geração da base de casos. O modelo é dividido em quatro módulos independentes e logicamente interligados. Um novo caso inicia-se durante o atendimento do paciente com quadro de oncologia pediátrica, momento em que a prescrição/evolução é cadastrada no banco de dados gerando um novo caso. Esta fase é apresentada na arquitetura do sistema pelo módulo ATENDIMENTO e detalhado no caso de uso da figura 6.4.

O módulo KDD é responsável pelo processo de ETL (*Extract, Transform and Load*), Extração, Transformação e Carga, da base de dados de produção para a base de dados de mineração.

O módulo CONHECIMENTO tem o objetivo de aplicar a técnica de mineração de dados adequada, utilizando o algoritmo de mineração apropriado para

gerar as regras de acordo com as medidas de similaridade parametrizadas que serão apresentados pelo módulo do USUÁRIO.

Figura 13 - Arquitetura do Sistema



Fonte: o autor

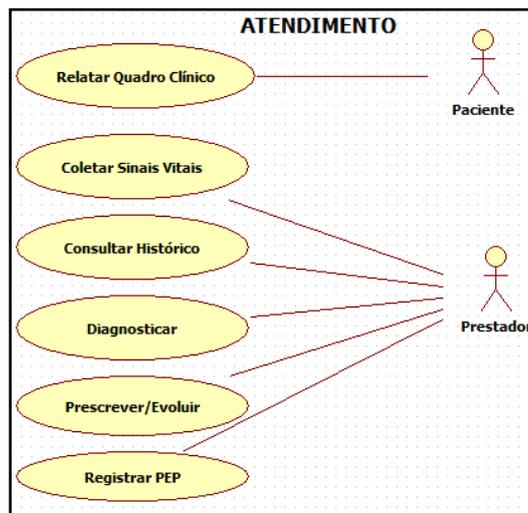
O modelo foi desenvolvido baseado no prontuário de cinco pacientes com quadro de oncologia pediátrica e guiado pelas indagações:

- *Existe similaridade com relação às prescrições médicas no tratamento de pacientes com casos de oncologia pediátrica?*
- *Tal similaridade pode ser utilizada no apoio ao processo de tomada de decisão na ocorrência de novos casos?*
- *Tal similaridade pode ser utilizada no apoio ao aprendizado de estudantes de medicina?*

6.4.1 MÓDULO ATENDIMENTO

Este módulo é alimentado a partir do Sistema de Gestão Hospitalar, MV Sistemas, no qual, durante o atendimento do paciente pelo prestador, os dados do atendimento são cadastrados no Prontuário Eletrônico do Paciente a partir da interface do sistema conforme mostrado nas figuras 14 e 15.

Figura 14 - Caso de Uso Atendimento



Fonte: o autor

O caso de uso da figura 14 apresenta as diversas interações dos atores com o sistema no módulo de atendimento onde o paciente relata o quadro clínico, o prestador aplica os procedimentos de atendimento e registra os dados no Prontuário Eletrônico do Paciente.

Figura 15 - Interface Apresentando Dados do Atendimento

Paciente	11	Maria Filha
Mãe		Maria Mãe
Data de Nascimento	27/07/1939	74 Anos
Data da Internação	20/01/2014	07:01
Médico Assistente	18	Dr. José
Convênio/Plano	X	PLANO EXTRA
Unidade de Internação	26	3º ANDAR
Setor	44	3. ANDAR
Leito		SUITE - (B)
Enfermaria		SUITE
Data Prevista para Alta		21/01/2014
Procedimento	43020208	GASTRECTOMIA TOTAL COM LINFADECTOMIA
CID	C16	NEOPLASIA MALIGNA DO ESTOMAGO

Fonte: o autor

A figura 15 apresenta os dados do atendimento armazenados na tabela ATENDIME e do prontuário do paciente da tabela PACIENTE.

A figura 16 mostra a tela de entrada de dados dos itens prescritos em um determinado atendimento e que serão armazenados nas tabelas PREMEDI, tabela

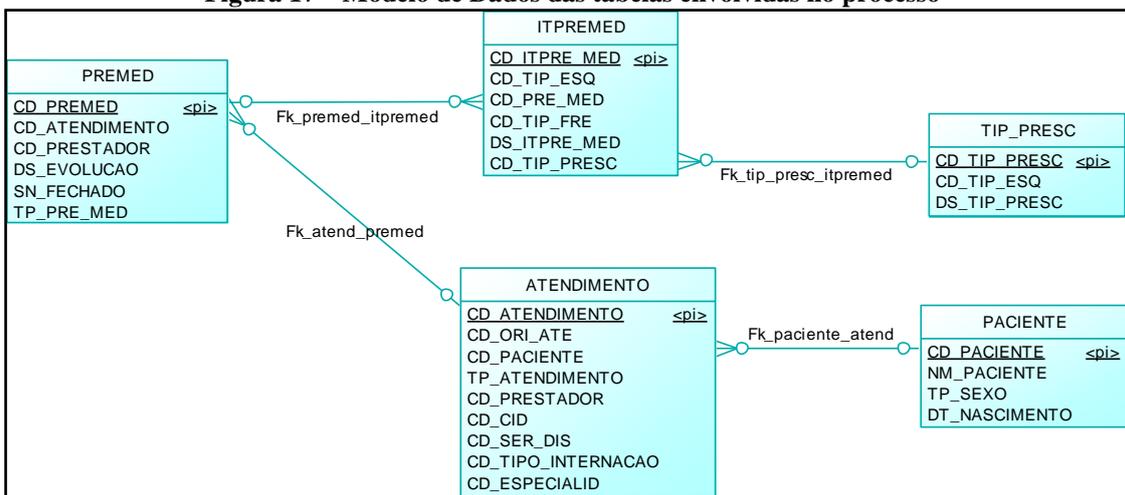
mestre que contém os dados da prescrição médica e na tabela ITPREMED que é a tabela detalhe que contém todos os itens prescritos no atendimento, itens numerados.

Figura 16 - Interface de Entrada de Dados dos Itens Prescritos

Fonte: o autor

Dentre as tabelas do sistema, as envolvidas no processo de extração de dados são mostradas no modelo E-R da figura 17.

Figura 17 – Modelo de Dados das tabelas envolvidas no processo



Fonte: o autor

Este modelo representa parte do banco de dados que será utilizado no processo de extração dos casos. Prontuários dos pacientes são cadastrados na tabela PACIENTE. Pacientes são atendidos e seus atendimentos são registrados na tabela ATENDIMENTO, onde são gravados os dados relativos ao atendimento, tais como data

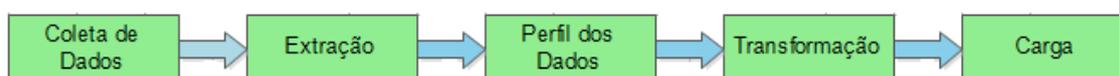
do atendimento, o CID (Classificação Internacional de Doenças) da enfermidade apresentada pelo paciente, o prestador que fez o atendimento, o tipo de atendimento bem como outros dados relativos ao atendimento em si. O quadro clínico do paciente é registrado na tabela PREMEDI na coluna ds_evolucao. De acordo com o quadro clínico do paciente, medicamentos, exames e outros tipos de tratamento prescritos ao paciente são registrados na tabela ITPREMEDI (itens da prescrição).

6.4.2 MODULO KDD

O módulo de KDD consiste na extração de dados do banco de dados de produção para gerar o banco de dados de mineração.

A partir dos dados cadastrados nas tabelas do modelo da figura 18, obtidos no atendimento, iniciaremos o processo de extração de dados a fim de gerar a base de casos. Este processo refere-se a tarefa de exportação da base de dados do sistema de produção para a base de mineração de dados, gerando a base de casos, processo este denominado ETL – Extração, Transformação e Carga que ocorre de acordo com as fases apresentadas na figura 18 (Vasconcelos, 2004).

Figura 18 - Processo ETL



Fonte: adaptado de Vasconcelos (2004)

Coleta de dados – refere-se à seleção ou amostra dos dados com os quais o processo de descoberta será realizado, que consiste da identificação da fonte de dados e análise dos dados a serem extraídos. Os dados serão extraídos utilizando a operação de junção das tabelas PREMEDI e ITPREMEDI nas colunas CD_PREMEDI, DS_EVOLUCAO E CD_TIP_PRESC, descritas no dicionário de dados das tabelas do [Apêndice 01].

Extração – extrair os dados das fontes e colocar em uma área denominada “*Staging Area*” ou “área de preparo”, que normalmente possui a mesma estrutura da fonte de origem.

Perfil dos dados – nesta etapa, é feito um estudo do perfil dos dados, a fim de aplicar as estratégias de limpeza, eliminação de ruídos, unificação e padronização dos dados.

Transformação – na fase de transformação são feitos o mapeamento e desenvolvimento dos procedimentos (*procedures*) de banco de dados para execução da carga.

Carga – refere-se ao processo de carga dos dados no banco de dados de mineração a partir da execução planejada dos procedimentos de banco de dados em momentos específicos segundo agendamento prévio.

O módulo de KDD compreende as seguintes fases:

- **Escolha das palavras-chaves** – a escolha das palavras-chaves será feita a partir dos relatos do quadro clínico registrados na evolução do paciente. Estas palavras foram escolhidas de acordo com os grupos de câncer pediátrico apresentados no item 2.6, tais como: Hepatoblastoma, retinoblastoma, Osteossarcoma, Fibrossarcoma. A partir da escolha da palavra-chave, é aplicado um filtro nas evoluções dos pacientes que contém esta, ou esta(s) palavras(s), gerando-se um novo conjunto de registros que contém a evolução e todos os itens prescritos para cada uma delas.

- **Limpeza** – a limpeza consiste na exclusão dos registros nulos ou que apresente alguma inconsistência nos dados.

- **Indexação** – a indexação possibilita a recuperação de casos mais similares presentes na base de casos com a criação das visões materializadas a serem utilizadas no processo de mineração de dados, gerando a **base de dados de mineração**. Este processo consiste na criação/atualização da base de mineração obtida a partir de um conjunto de itens prováveis, denominados de conjunto de itens candidatos.

6.4.3 PROCESSO DE MINERAÇÃO DE DADOS

Após a obtenção da base de mineração de dados, o processo de mineração se dá em dois passos:

- Escolha do modelo de mineração de dados;
- Aplicação do o algoritmo de similaridade.

Esta fase consiste no processo da análise dos dados sob diferentes perspectivas, a fim de se descobrir informações úteis que normalmente não estão visíveis ou que não são encontradas de maneira trivial.

O processo de mineração de dados envolve na aplicação de análise de dados e algoritmos de descobrimento que produzem uma enumeração de padrões (ou modelos) particular sobre os dados.

Método (ou técnica) de Mineração

Dentre os diversos métodos de mineração o método escolhido, pela própria natureza da pesquisa, foi o método de regra de associação, cuja ideia é buscar padrões encontrados nos dados em um formato que seja interpretável pelo homem.

Conforme o questionamento “*Existe similaridade com relação às prescrições médicas no tratamento de pacientes com casos de oncologia pediátrica?*”, foi feito um estudo com relação aos itens prescritos para o tratamento do paciente em um determinado atendimento. Cada atendimento gera uma prescrição na qual o quadro clínico é relatado, e esta informação armazenada na evolução do paciente, conforme figuras 11 e 12. De acordo com os dados da evolução o prestador prescreve itens para o tratamento conforme o tópico 7.2.

Analisando-se diferentes prescrições na qual existe semelhança entre o quadro, no que diz respeito à ocorrência de palavras-chaves, observamos a repetição entre alguns itens prescritos. Os quadros abaixo mostram a relação de itens prescritos para dois atendimentos com relato de “Neutropenia Febril”.

De acordo com o resultado apresentado, verificamos que, para o atendimento 17xxx01, foram prescritos 11 itens de tratamento, enquanto que, para o atendimento 21xxx09, foram prescritos 16 itens.

Dos itens prescritos nos dois atendimentos, ocorreu a repetição de cinco itens, quais sejam: Dipirona sódica 500mg, Bactrim 400+800mg comp, solução

glicofisiológica, Dieta branda p/ neutropênico, Creatinina. Tal repetição provoca o questionamento feito no início do tópico. A partir do questionamento e dos resultados apresentados, o problema é saber se esta similaridade ocorre e, se ocorre, em que dimensão no banco de dados, se podemos estendê-la aos demais atendimentos e, baseado nos resultados, de que forma ela pode ser útil no aprendizado e no apoio a tomada de decisão.

Baseado no conceito de regras de associação, proposto por Agrawal, Imielinsk Swami (Agrawal, 1993). Uma regra de associação é definida como *se X então Y*, ou $X \rightarrow Y$, onde X e Y são conjuntos de itens e $X \cap Y = \emptyset$. Diz-se que X é o antecedente da regra, enquanto Y é o seu conseqüente. Desta forma, seja X o conjunto antecedente de itens prescritos em um atendimento e Y o conseqüente, logo:

$$X(\text{item_presc}_1, \text{item_presc}_2, \dots, \text{item_presc}_n) \rightarrow Y(\text{item_presc}).$$

Figura 19 - Comparação entre itens prescritos em dois atendimentos

	CD_PRE_MED	DS_TIP_PRESC	
1	17	01	MAXCEF 1 G FR/AMP + BOLSA 100 ML DIL
2	17	01	DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO
3	17	01	SOLUCAO GLICOFISIOLOGICA SIST FECHADO 500 ML BC
4	17	01	REHIDRAT 90 SABOR NATURAL ENV P/ 500 ML
5	17	01	PLASIL 10 MG/2 ML AMP (IV)
6	17	01	DIETA BRANDA P/ NEUTROPENICO
7	17	01	BACTRIN F 400 + 80 MG / 5 ML SUSP PED
8	17	01	AFERIR SINAIS VITAIS
9	17	01	HIDROCORTISONA 100 MG FR/AMP + DIL 2 ML GEN EUR
10	17	01	POLARAMINE 2 MG / 5 ML XAROPE
11	17	01	FUROSEMIDA 20 MG AMP 2 ML GEN HYPOFARMA (IV)

	CD_PRE_MED	DS_TIP_PRESC	
1	21	19	TRAMADOL 50MG/1ML GEN UNIAO (IV/DILUI EM 100 SF
2	21	19	DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO
3	21	19	NISTANTINA 100.000 UI SUSP ORAL CRISTALIA (5 ML)
4	21	19	BACTRIM 400+80 MG COMP
5	21	19	SOLUCAO GLICOFISIOLOGICA SIST FECHADO 500 ML BC
6	21	19	RAIO X TORAX - P.A.
7	21	19	TOMOGRAFIA COMP. DO CRANIO C/ CONTRASTE N/ IOI
8	21	19	DIETA BRANDA P/ NEUTROPENICO
9	21	19	HEMOGRAMA
10	21	19	CONTAGEM DE PLAQUETAS
11	21	19	HEMOCULTURA (POR AMOSTRA)
12	21	19	CEFEPIMA 2 G FR/AMP GEN EUROFARMA IV
13	21	19	PUNCAO DE PORT-A-CATH
14	21	19	AFERIR SINAIS VITAIS
15	21	19	CREATININA
16	21	19	PLAMET 10 MG/2 ML AMP (INTRAVENOSO)

Fonte: o autor

A partir deste conjunto de itens serão aplicadas as medidas de suporte e confiança a fim de determinar as combinações de itens frequentes.

Algoritmo de Mineração

Pelos motivos expostos nos parágrafos anteriores o algoritmo escolhido para o presente trabalho foi o *Apriori*.

7 IMPLEMENTAÇÃO DO ESTUDO DE CASO

7.1 INTRODUÇÃO

Este capítulo apresenta o resultado dos experimentos relacionados ao processo de mineração de dados em prescrições médicas em pacientes com câncer infantil. O trabalho foi motivado pelo seguinte questionamento:

Existe similaridade com relação ao conjunto de itens prescritos em atendimentos médicos para tratamento de pacientes com casos de oncologia pediátrica em evoluções médicas com ocorrência de palavras-chaves semelhantes?

De acordo com o ciclo de vida descrito no capítulo anterior, aplicaremos a metodologia para o desenvolvimento do caso.

7.2 ENTENDIMENTO DO NEGÓCIO

Esta fase envolve os seguintes passos:

Compreensão do problema – a partir do questionamento que dá origem ao problema, devemos colher as informações necessárias à sua compreensão, a fim de que possamos discutir, juntamente com o especialista do negócio, de que maneira como o mesmo lida com o problema. Além do processo de entrevistas, tal coleta de informações pode ser feita através de livros, artigos, pesquisa na internet, manuais de procedimentos e demais fontes de informações.

Um problema nessa fase é criar uma estratégia que facilite o diálogo, a fim de que as diferenças de linguagem entre o técnico e o especialista de negócio sejam minimizadas. A utilização de modelos é necessária para o estabelecimento de uma aproximação, pois permite que, através de uma ferramenta gráfica, ocorra a interação entre o desenvolvedor e o especialista no negócio e este processo não deve se esgotar em apenas uma interação. Todas as informações deverão ser validadas antes de iniciar o processo de implementação.

7.3 BASE DE CASOS

A base de casos é formada por prontuários eletrônicos de 9 (nove) pacientes com quadro de oncologia pediátrica. Apesar de aparentemente pequeno, este conjunto de prontuários gerou uma grande quantidade de atendimentos e consequentemente prescrições médicas, de acordo com o quadro 1.

Quadro 1 - Quantitativo de dados gerados

Quantidade de pacientes (tabela PACIENTE)	9
Atendimentos no período de 2005 a 2013 (tabela ATENDIMENTO)	2.041
Prescrições médicas geradas nos atendimentos (tabela PREMEDI)	3.048
Itens Prescritos (tabela ITPRE_MED)	21.133
Quantidade média de itens da prescritos por prescrição	6,9
Menor quantidade de itens de prescritos	1
Maior quantidade de itens de prescritos	39

Fonte: o autor

A partir das tabelas ATENDIME, PACIENTE, PRE_MED, ITPRE_MED, TIP_PRESC do sistema MV, foram criadas as visões materializadas, apresentadas no modelo entidade-relacionamento da figura 17 e descritas no apêndice [Apêndice 01].

7.4 CONHECIMENTO DOS DADOS

Esta fase, de acordo com Chapman (2000) inicia-se com a carga inicial dos dados e sua exploração, a fim de conhecê-lo.

7.4.1 OBTENÇÃO DOS DADOS

A partir do questionamento escolheu-se os dados de 9 (nove) pacientes com quadro de oncologia pediátrica. Analisando as tabelas do sistema MV, foram identificadas as tabelas “ATENDIME”, “PACIENTE”, “PRE_MED”, “ITPREMED” e “TIP_PRESC”. A extração dos dados foi feita a partir da base de dados do Sistema de Grastão Hospitalar MV Sistemas para planilhas MsExcel, de acordo com as *queries* do [Apêndice 02]. Após a extração dos dados, foi criado um esquema no banco de dados denominado MMCC, para manipulação dos mesmos.

7.4.2 MASCARAMENTO DOS CÓDIGOS DOS PACIENTES

Com objetivo de preservar a privacidade dos dados dos pacientes, os dados do atributo relativo ao código do paciente da tabela PACIENTE, que se refere ao seu número de prontuário, foram modificados para 1, 2, 3, 4, 5, 6, 7, 8, 9.

A fim de manter a integridade referencial, os mesmos também foram modificados nas visões materializadas atendimento, prescrição e itens da prescrição.

A mesma estratégia foi aplicada ao código do atendimento e ao código dos itens da prescrição utilizando-se um procedimento desenvolvido em linguagem PL/SQL para “criptografar” tais códigos.

7.4.3 ESCOLHA DAS PALAVRAS CHAVES

A partir do diagnóstico registrado no atributo DS_EVOLUCAO da tabela PRE_MED, [Apêndice 01], que contém o quadro apresentado pelo paciente naquele atendimento, e, conforme os agrupamentos descritos na sessão 2.6, identificaram-se algumas palavras chaves e aplicamos consultas SQL com objetivo de fazer um levantamento referente a ocorrência das mesmas.

Por exemplo, analisando-se duas prescrições de um paciente com quadro clínico, nas quais ocorre a palavra-chave OSTEOSSARCOMA, no atributo DS_EVOLUCAO da tabela PRE_MED, observamos que existe um grau de similaridade entre os itens de tratamento prescritos para o quadro apresentado, no atributo CD_TIP_PRESC da tabela ITPRE_MED.

Prescrição 1

Quadro clínico do paciente

Premed.ds_evolucao

HD: OSTEOSSARCOMA

EVOLUINDO SEM INTERCORR~ENCIAS MAIORES.

EX FÍSICO: MANTIDO

CD: VPM + AGUARDO EXS DE HOJE

Itens de tratamento prescritos

Itpre_med

Quadro 2 - Itens prescritos para a evolução com quadro de OSTEOSSARCOMA

Cd_tip_esq	Cd_tip_presc	Cd_tip_fre	Ds_itpre_med
DIE	26854		
DEE	39922	22	
MED	43161	9	
MED	20931	8	SE NÁUSEA OU VÔMITO
MED	21995	9	SE DOR
MED	17063	14	SE DOR
MED	42468	9	
MED	40310	27	
MED	17003	14	
MED	17254	6	SE TIVER FEBRE, MAIOR OU IGUAL a 38
MED	42072	6	1,5 ML + 8 ML AD EV SE FEBRE OU DOR.
MED	14187	21	Administrar em bolus de 3-4min ou infusão contínua de 30-40min.
PRO	14509		
PRF	39305	27	CIRURGICO
PRF	30201	27	PAM
PRF	14487	8	
PRF	14490	27	
ANO	9302	6	
FIS	2667	0	

Fonte: o autor

Prescrição 2

Quadro clínico do paciente

Premed.ds_evolucao

HD.: OSTEOSSARCOMA

NÓDULO PULMONAR RESSECADO

CLINICAMENTE ESTÁVEL

EXAME SEGMENTAR INALTERADO

CD.: MANTIDA

Itens de tratamento prescritos

Itpre_med

Quadro 3 - Itens prescritos para a evolução com quadro de OSTEOSSARCOMA

Cd_tip_esq	Cd_tip_presc	Cd_tip_fre	Ds_itpre_med
DIE	26861		LAXANTE
DEE	39922	22	
MED	43161	9	
MED	20931	8	SE NÁUSEA OU VÔMITO
MED	21995	9	

MED	17286	8	
MED	17063	14	
MED	42468	9	
MED	40310	27	
PRF	39305	27	CIRURGICO
PRF	30201	27	PAM
PRF	14487	8	
PRF	14490	27	
ANO	9302	6	
FIS	2667	0	
FIS	2696	0	
GAS	43422	20	

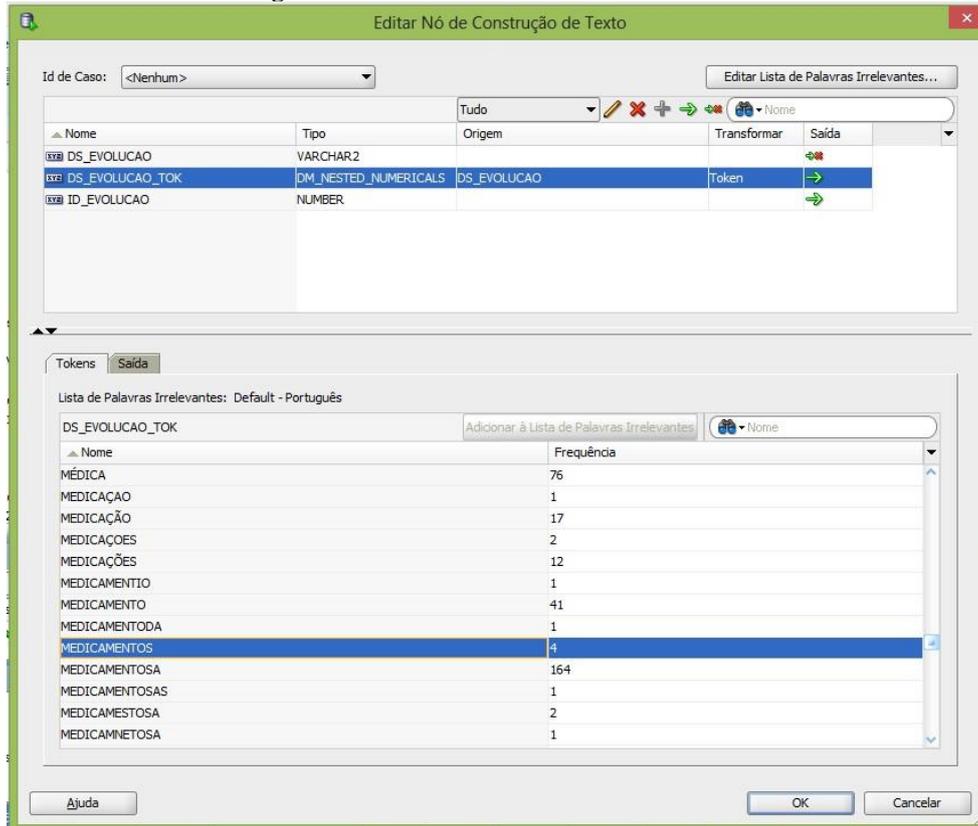
Fonte: o autor

A tarefa de encontrar palavras chaves relevantes em textos, de acordo com Aranha (2006) tem sido um desafio devido a ocorrência de dentre outro problemas quaos sejam:

- Dados ruidosos: nos quais são encontrados erros ortográficos, palavras sem acento, ausência de espaços entre as palavras acarretando o que poderíamos chamar de palavras mutantes;
- Problema da ambiguidade: onde uma palavra pode atender a mais de um significado de acordo com o contexto;
- Super parametrização: Encaixar uma regra para uma dada frase pode ser bastante simples, mas encontrar o número de regras que atendem a todas as possíveis construções da língua pode levar a uma quantidade não tratável de regras.

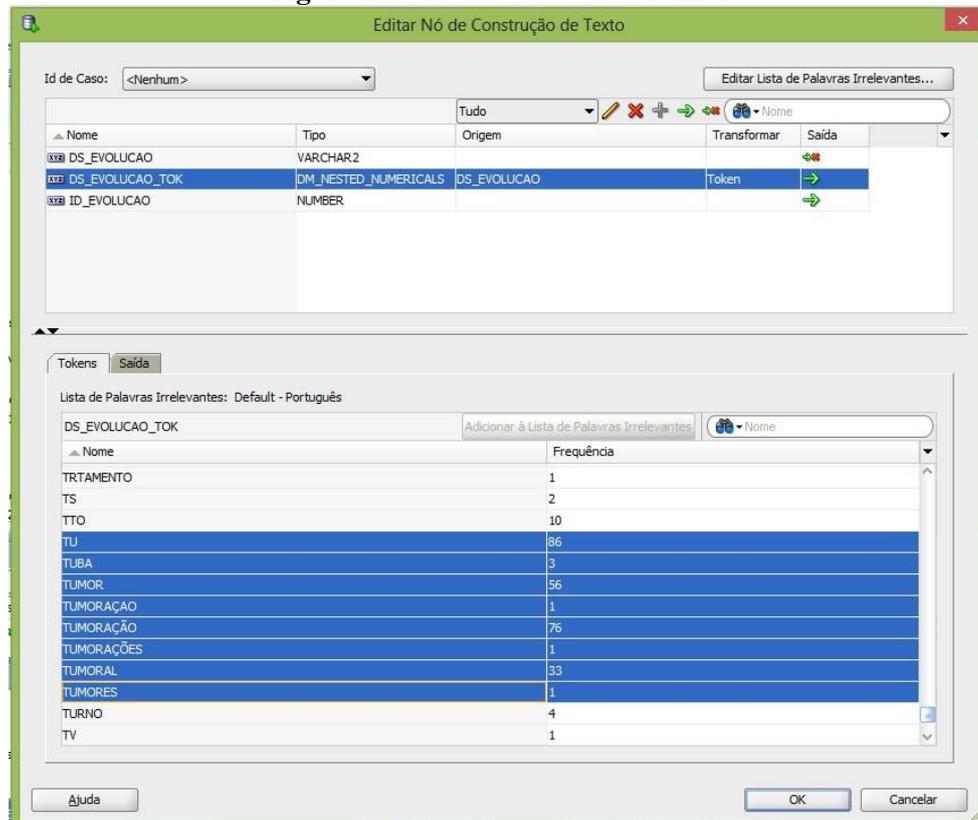
As figuras 20 e 21 mostram a aplicação de mineração de dados, utilizando a ferramenta *Oracle Text Mining* no atributo DS_EVOLUCAO que apresenta a palavra e sua frequência com relação aos registros no banco de dados. Neste caso apresentamos as ocorrências para as palavras MEDICAMENTOSA e TUMOR.

Figura 20 - Lista de Palavras Medicamentos



Fonte: o autor

Figura 21 - Lista de Palavras Tumor



Fonte: o autor

7.5 TRATAMENTO DA BASE DE DADOS

Nesta fase, a fim de criar uma base de dados mais consistente para o processo de mineração, foram efetuados alguns ajustes nos dados, tais como a eliminação de registros no qual o atributo DS_EVOLUCAO continha valor nulo, registros que não puderam ser lidos devido a compatibilidade de migração da base de dados Oracle para MsExcel.

7.5.1 CRIAÇÃO DA VISÃO MATERIALIZADA

Com objetivo de consolidar os dados das prescrições para execução do processo de definição da regra de associação, criou-se a visão materializada [Apêndice 03] denominada PRESCRICAO. Visão materializada, *materialized view*, é um objeto de banco de dados que, ao contrário de uma visão (*view*), que não contém dados e apenas refere-se a uma consulta SQL, a visão materializada contém os dados resultantes da consulta SQL.

Esta visão contém os dados obtidos a partir da junção das tabelas prescrição (PREMED) e dos itens da prescrição (ITPRE_MED).

7.5.2 LIMPEZA DA BASE DE DADOS

Partindo-se do princípio que a coluna PREMED.DS_EVOLUCAO contém as palavras chaves a serem pesquisadas, excluiu-se todos os registros na qual esta coluna possui valor nulo, utilizando o comando SQL DELETE. Esta operação excluiu 6.328 registros.

```
DELETE FROM PRESCRICAO  
WHERE DS_EVOLUCAO IS NULL;
```

Esta exclusão influencia fortemente no cálculo da frequência relativa da ocorrência destas palavras-chaves no banco de dados, haja vista que, o cálculo da frequência relativa leva em consideração o número de ocorrências do termo em relação ao total de registros do banco de dados afetando os valores de suporte e confiança.

7.5.3 ATUALIZAÇÃO DA BASE DE CASOS

Com objetivo de acrescentar novos casos à base de casos, foi desenvolvida uma *procedure* de banco de dados [Apêndice 04] em linguagem de programação PL/SQL denominada PRC_LOAD_BASE_CASOS.

Esta *procedure* é executada com periodicidade diária de acordo com as configurações do *job* de agendamento de tarefas criado no banco de dados Oracle conforme figura 22.

Figura 22 - Dados do job de agendamento

Row 1	Fields
▶ JOB	41
LOG_USER	MMCC
PRIV_USER	MMCC
SCHEMA_USER	MMCC
LAST_DATE	24/08/2013 02:28:39
LAST_SEC	02:28:39
THIS_DATE	
THIS_SEC	
NEXT_DATE	25/08/2013 02:28:39
NEXT_SEC	02:28:39
TOTAL_TIME	0
BROKEN	N
INTERVAL	Sysdate+1
FAILURES	0
WHAT	prc_load_base_casos;
NLS_ENV	NLS_LANGUAGE='BRAZILIAN PORTUGUESE' NLS_TERRIT...
MISC_ENV	0102000200000000
INSTANCE	0

Fonte: o autor

De acordo com os dados do *job* criado em 24/08/2013 as 02h:28m:39s, a próxima execução *procedure* PRC_LOAD_BASE_CASOS será no dia 25/08/2013 as 02h:28m:39s e sua periodicidade de execução será diária de acordo com a coluna *INTERVAL* (*Sysdate* +1). A partir da próxima execução do *job* a coluna *LAST_DATE* assume o valor 25/08/2013 02:28:39 e a coluna *NEXT_DATE* o valor 26/08/2013 as 02:28:39 e esta troca será efetuada a cada execução do *job*.

7.5.4 CRIAÇÃO DAS VISÕES UTILIZADAS NO PROCESSO DE MINERAÇÃO

Com base nas palavras chaves identificadas nos agrupamentos descritos no item 2.5 foram criadas duas visões [Apêndice 04] que serão utilizadas no processo de mineração de dados denominadas VW_PRESCRICAO, que retorna apenas os códigos (CD_PRE_MED, CD_ITPRE_MED, CD_TIP_PRESC), e VW_PRESCRICAO_DSC, que retorna os códigos da prescrição e item da prescrição e a descrição do tipo da prescrição (CD_TIP_PRESC). Estas visões serão as fontes de dados do *workflow* da ferramenta *Oracle Data Miner* para o processo de mineração de dados no qual será aplicado o método de mineração por regra de associação, utilizando o algoritmo APRIORI. Esta visão pesquisa os dados da coluna DS_EVOLUCAO, usando o operador *SQL LIKE*, que utiliza o caractere curinga % (porcentagem) ou _ (sublinhado) para procurar uma *string* padrão, onde o sublinhado (_) serve para marcar uma posição específica e o símbolo de porcentagem (%) qualquer caractere ou *string* antes ou depois da posição desejada. Nos exemplos do [Apêndice 05] a visão VW_PRESCRICAO_DSC retorna todos os registros cuja coluna que contém a descrição da evolução (DS_EVOLUCAO) possui os *strings* “NAUSEAS” ou “NEUTROPENIA FEBRIL” e a visão VW_PRESCRICAO retorna os registros que contém a palavra chave “OSTEOSSARCOMA”.

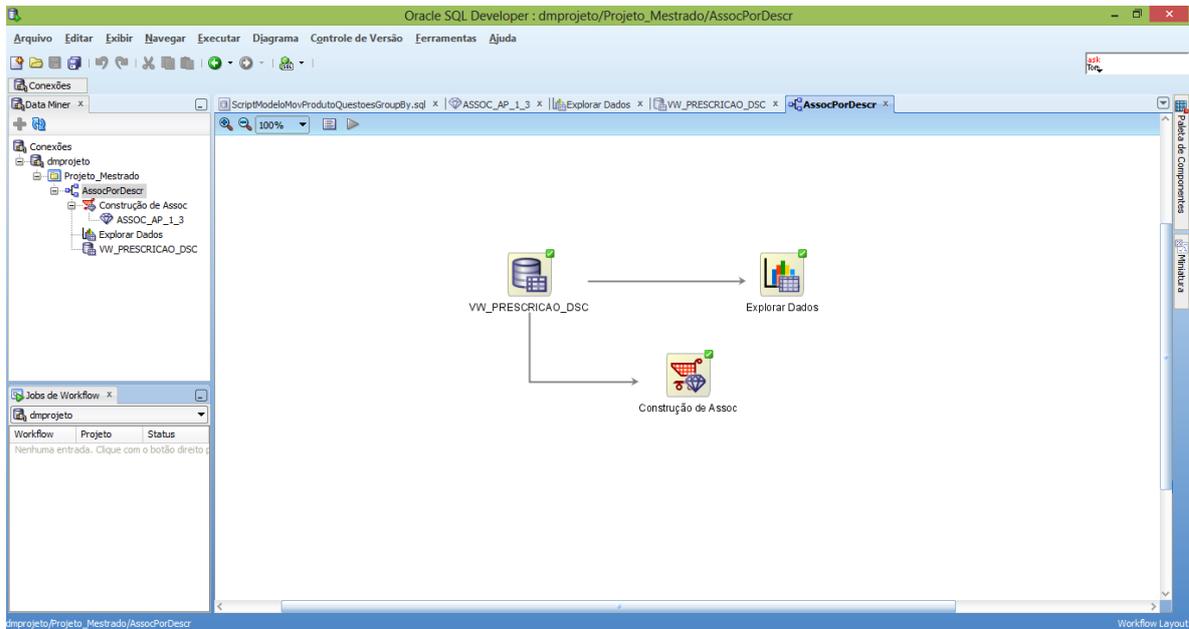
7.6 MODELO DE MINERAÇÃO

A partir das visões VW_PRESCRICAO e VW_PRESCRICAO_DSC, aplicou-se o processo de mineração de dados, utilizando a interface *Oracle Data Miner*.

O *Oracle Data Mining (ODM)* é uma ferramenta gratuita que, de acordo com Oracle (2013) fornece poderosas funcionalidades para mineração de dados. *Oracle Data Mining GUI* é uma extensão do *Oracle Developer SQL* que permite aos analistas de dados trabalharem diretamente com dados dentro do banco de dados, explorar os dados graficamente, construir e avaliar os vários modelos de mineração de dados, além de aplicar modelos de mineração de dados Oracle para novos dados. Fluxos de trabalho do *Oracle Data Miner* permitem capturar e documentar a metodologia analítica do usuário, que podem ser salvas e compartilhadas com outras pessoas para automatizar metodologias analíticas avançadas.

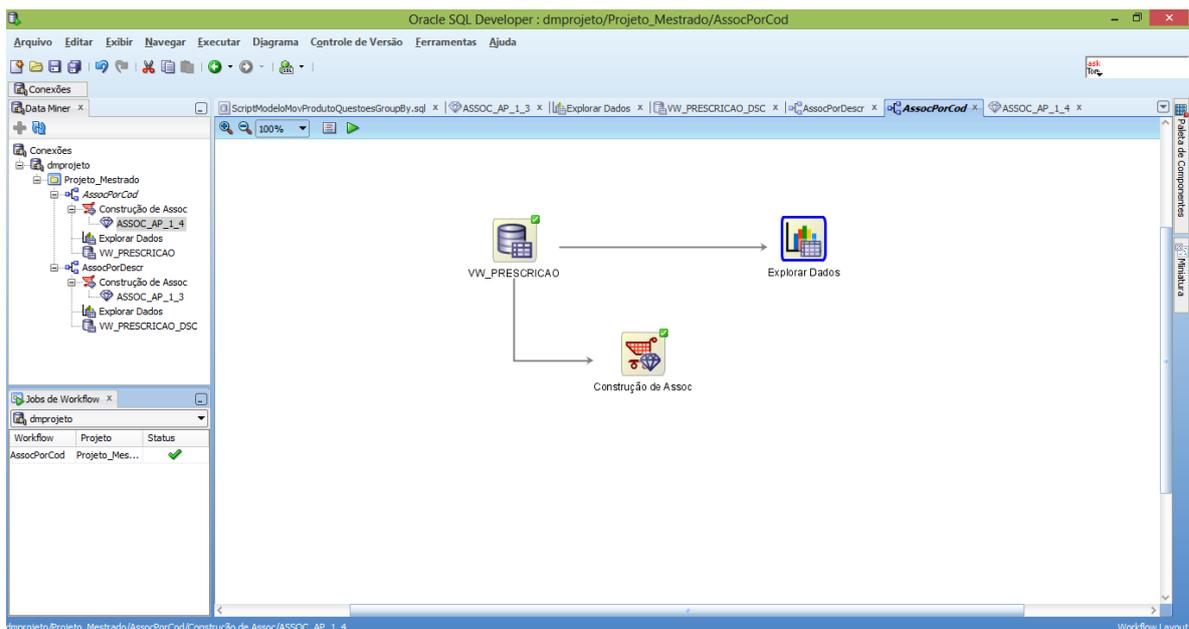
Para visão VW_PRESCRICAO_DSC foi criado um fluxo denominado AssocPorDescri, e para a visão VW_PRESCRICAO, um outro fluxo denominado AssocPorCod, de acordo com as figuras 23 e 24 respectivamente. As análises dos dados estão nos anexos A, B, C e D.

Figura 23 - Fluxo AssocPorDescr



Fonte: o autor

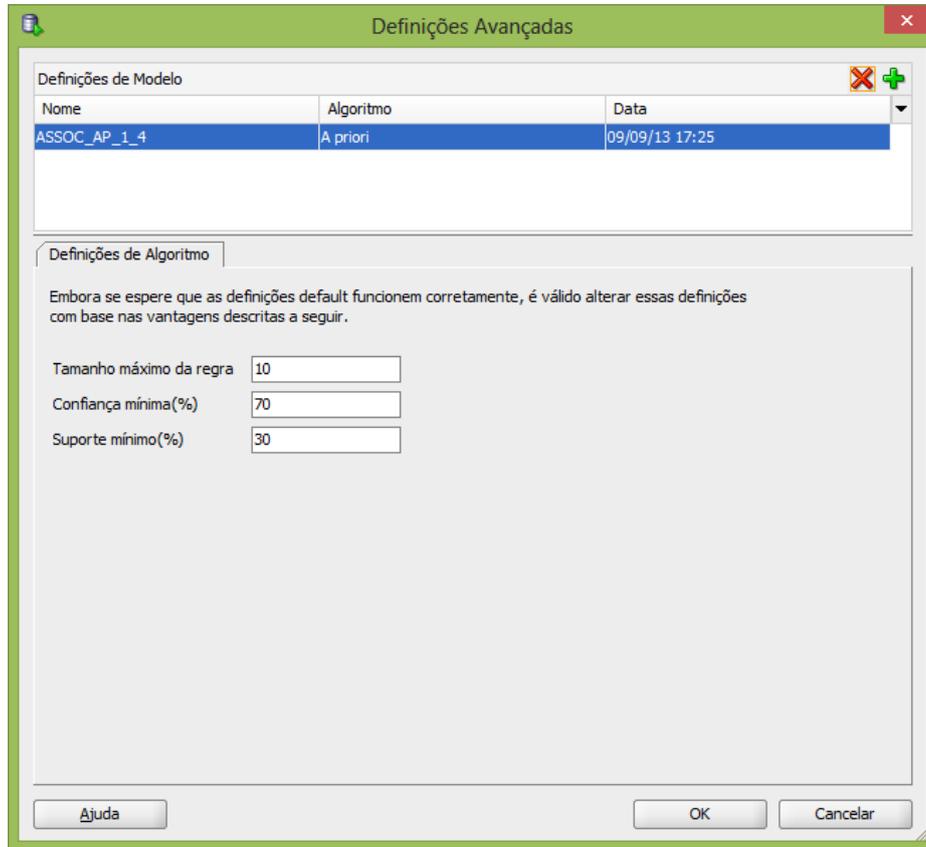
Figura 24 - Fluxo AssocPorCod



Fonte: o autor

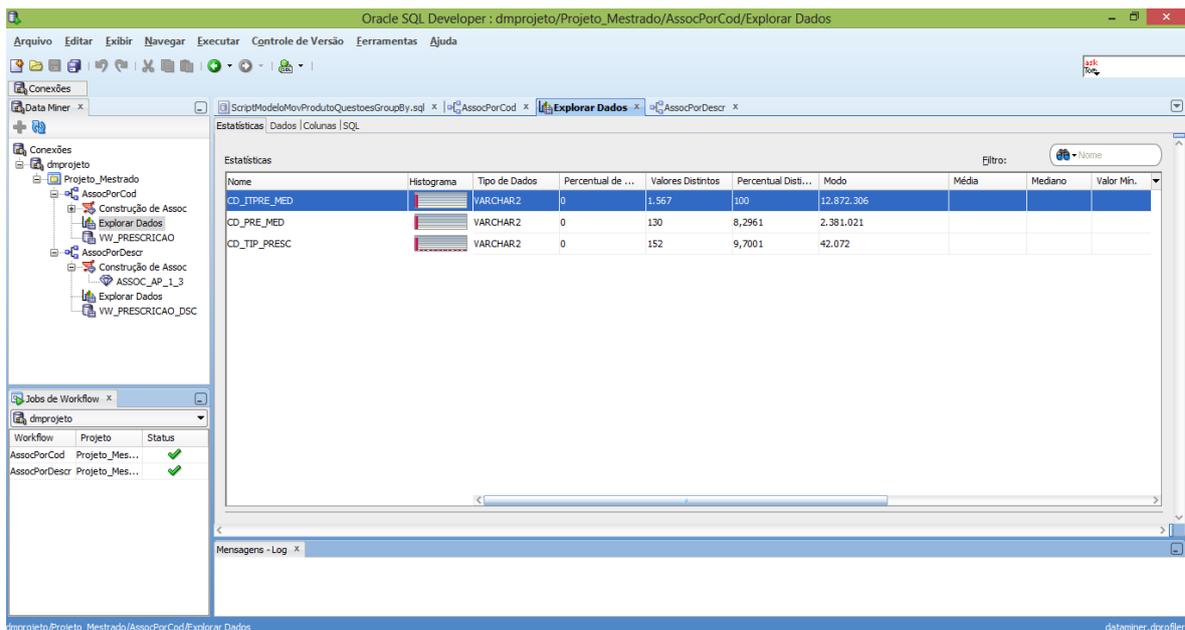
As medidas de suporte e confiança bem como o tamanho da regra podem ser parametrizadas na janela de definições avançadas, mostrada na figura 25.

Figura 25 - Parâmetros de Execução do fluxo



Fonte: o Autor

Figura 26 - Histograma dos Dados da visão VW_PRESCRICAO



Fonte: o autor

A figura 26 mostra um histograma com informações importantes a respeito da base de dados a ser minerada, tais como: Percentual de nulos, valores distintos, média, mediana, maior valor, menor valor, desvio padrão e variação.

8 CONSIDERAÇÕES FINAIS

Neste capítulo apresentaremos as conclusões e sugestões para trabalhos futuros.

8.1 CONCLUSÃO

Conforme discutido no decorrer do trabalho, embora utilização de RBC na medicina, seja uma área ainda em crescimento, vem mostrando resultados bastante positivos no apoio ao processo de tomada de decisão médica bem como mais uma ferramenta para melhoria do ensino de medicina, seja na área de análise de imagens seja na área de diagnósticos médicos.

Os dados armazenados nos prontuários eletrônicos do paciente têm se tornado importante fonte de informações para o aprendizado e a adoção de sistemas RBC, em combinação com ferramentas de BI (*Business Intelligence, Data Warehouse* e mineração de dados), pode ser visto como uma importante ferramenta para as atividades de treinamento, tutoria e apoio à decisão.

A utilização do modelo apresentado facilita a visualização e entendimento do fluxo do processo de descoberta de conhecimento na base de dados e consequente geração e atualização dos casos.

Dentre os problemas encontrados, destacamos que, apesar de aplicação de regras de normalização no desenvolvimento do banco de dados, o que estabelece certa padronização na codificação dos dados, os prontuários eletrônicos apresentam muitos campos (atributos) descritivos, como, por exemplo, o campo no qual o prestador relata a evolução do paciente. Conforme este contexto, a maior dificuldade é a falta de padronização na escrita do texto, informações incompletas ou nulas além de escrita diferentes para o mesmo termo. Esta falta de estruturação e padronização da descrição dos dados implica em uma maior dificuldade na indexação das informações, o que é crucial para estabelecer a similaridade de prescrições e permitir a representação da informação.

O que concluímos é que, devido aos problemas de representação da informação, devemos dar maior atenção ao processo de indexação a fim de melhorar a seleção de bons índices para que as métricas de similaridade retornem valores mais próximos e confiáveis.

Neste contexto, podem-se destacar as seguintes contribuições deste trabalho:

- Permitir maior interação entre prestadores e aprendizes no entendimento da aplicação de itens de prescrição baseado no prontuário médico eletrônico;
- Criar um processo automático de geração da base de casos.
- Facilitar o processo de tomada de decisão médica através da recuperação de casos e avaliação das medidas de similaridade;

8.2 TRABALHOS FUTUROS

A partir deste trabalho podemos destacar alguns trabalhos que podem ser desenvolvidos:

- Desenvolver uma interface interativa para a apresentação dos dados e parametrização das palavras chaves utilizadas no processo de indexação;
- Adicionar o elemento tempo para ter uma informação histórica, a fim de permitir uma análise temporal da evolução do paciente.
- Ampliar a pesquisa para outros elementos, tais como prestador, faixa etária, sexo, tipo de atendimento.
- Ampliar a pesquisas a outras enfermidades tais como cardiologia, pneumologia dentre outros CIDs.
- Ampliar a pesquisas a outros tipos de atendimento tais como Eletivo, Urgência/emergência, Internações, *home care*.
- Validação da aplicação das métricas de similaridade;

REFERÊNCIAS

AAAI AAAI, <http://www.aaai.org/home.html> acessado em 06/10/2013

AAMODT A., PLAZA, E. Case-based reasoning: Foundational issues, methodological variations, and system approaches. **AI communications**, v. 7, n. 1, p. 39-59, 1994.

AAMODT A., Sandtorv H. A., Winnem O. M. Combining Case Based Reasoning and Data Mining - A way of revealing and reusing RAMS experience. *Safety and Reliability; Proceedings of ESREL '98, Trondheim, June 16-19, 1998*.

ABEL, M. Um Estudo sobre Raciocínio Baseado em Casos. Instituto de Informática. Universidade Federal do Rio Grande do Sul. Porto Alegre, 1996

AGRAWAL, R., Srikant R. *Fast algorithms for mining association rules*. IBM Almaden Research Center, Proceedings of the 20th VLDB Conference Santiago, Chile, 1994.

ARANHA C., Passos E. A Tecnologia de Mineração de Textos. RESI-Revista Eletrônica de Sistemas de Informação, Nº2, 2006

BAREISS, E & Porter, B, Protos: An exemplar-based learning apprentice. *In Proceedings of the Fourth International Workshop on Machine Learning*, Morgan Kaufmann, San Mateo, CA, pp. 12-23, 1987.

BARTLETT, F. C. *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press, 1932.

BELLMAN, R. E. *An Introduction to Artificial Intelligence: Can Computers Think?* Boyd & Fraser Publishing Company. 1978.

BERGER, J, Roentgen: Radiation Therapy and Case based Reasoning. *In the Proceedings of the Conference on Artificial Intelligence Applications*, IEEE Computer Society Press, Los Alamitos, CA, pp. 171-177, 1994.

BICHINDARITZ, I. Case-Based Reasoning adaptive to several cognitive tasks. *In the Proceedings of the International Conference on CBR*, Aamodt A. and Veloso M. (eds.), LNAI, No. 1010, Springer, Berlin, pp. 391-400, 1995.

BICHINDARITZ, I, Kansu, E & Sullivan KM, Case-Based Reasoning in CARE-PARTNER: Gathering Evidence for Evidence-Based Medical Practice. *In the Proceedings of the 4th European Workshop on CBR*, Smyth B & Cunningham P, (Eds.), Springer, Berlin, pp. 334-345, 1998.

BICHINDARITZ, I & Potter, S, PhylSyst:un système d'intelligence artificielle pour l'analyse cladistique, *Biosystema*, 12:17-55, 1994.

BICHINDARITZ, I & Potter, S, Knowledge Based Phylogenetic Classification Mining. *In the Proceedings of the 4th Industrial Conference on Data Mining*, Ed. Perner, P, Lecture Notes in Computer Science, Springer-Verlag, Berlin, 3275:163-172, 2004.

BICHINDARITZ, I & Seroussi, B, Contraindre l'analogie par la causalite', *Technique et Sciences Informatiques*, 11(4):69-98, 1992.

BICHINDARITZ, I.; Montani, S. Advances in Case-based Reasoning in the Health Sciences. *Artificial intelligence in medicine*, v. 51, n. 2, p. 75-79, 2011.

BRANDEN, M Et tal. *Integrating Case-based Reasoning With an Electronic Patient Record System* in Artificial Intelligence in Medicine, pp 117-123, 2011

BRADBURN, C & Zeleznikow, J, The Application of Case-based Reasoning to the Tasks of Health Care Planning. *In the Proceedings of the European Workshop on CBR*, Wess, S, et al. (Eds), Springer, Berlin, pp. 365-378, 1993.

CAMARGO A. C. Tumor de Wilms A C Camargo Câncer Center. Disponível em <http://www.accamargo.org.br/tudo-sobre-o-cancer/tumor-de-wilms/39/> acessado em abr/2014.

CFM, SBIS Cartilha Sobre Prontuário Eletrônico, 2012, em http://www.sbis.org.br/certificacao/Cartilha_SBIS_CFM_Prontuario_Eletronico_fev_2012.pdf acessado em 20/04/2013.

CHAPMAN P. et tal. Crisp-DM 1.0. Step by Step Data Mining Guide, Crisp-DM Consortium, 2000

CORBETT, A. T., & Anderson, J. R. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and useradapted interaction*, 4, 253–278, 1995.

CORDIER, A et al. Interactive knowledge acquisition in case based reasoning. In: Workshop on Knowledge Discovery and Similarity, a workshop of the seventh International Conference on Case-Based Reasoning (ICCBR-07), DC Wilson and D. Khemani (volume editors). 2007.

DELPITZO V. L. F. Prescrição de Atividades Físicas através do uso da Inteligência Artificial. Dissertação (Mestrado em Engenharia da Produção) – Programa de Pós-graduação em Engenharia da Produção, UFSC – Florianópolis, 1997.

EDUARDO M., Et tal O prontuário Eletrônico do Paciente na Assistência, Informação e Conhecimento Médico – São Paulo : H. de F. Marin, 2003. P.1-3, 6

ESHACH H, Bitterman H. From Case-based Reasoning to Problem-based Learning, *Academic Medicine*, Vol. 78 Nº 5, May, 2003.

FAYYAD U., Et tal From Data Mining to Knowledge Discovery in Databases – Copyright © 1996, American Association for Artificial Intelligence. P.37-54

FRANK, S., Jeremy C. W. ABC of Health Informatics. Blackwell Publishing Ltd, 2006

GALVÃO N. D. & Marin H. F. Técnica de Mineração de Dados: uma Revisão da Literatura. *Acta Paulista de Enfermagem* vol.22 nº.5 São Paulo set./out. 2009

GIRONA, Control Engineering and Intelligent Systems
<http://exit.udg.edu/applications/exiticbr.aspx> acessado em 04/2014

GRIMNES, M & Aamodt, A, A two layer case-based reasoning architecture for medical image understanding, *Advances in Case-Based Reasoning*, Smith I & Faltings B, (Eds.) Berlin: Springer Verlag, pp. 164-178, 1996.

HADDAD, M, Adlassnig, KP & Porenta, G, Feasibility analysis of a case-based reasoning system for automated detection of coronary heart disease from myocardial scintigrams. *Artificial Intelligence in Medicine*, 9(1):61-78, 1997.

HAMMOND, K. J. Case-Based Planning: Viewing Planning as a Memory Task. Academic Press, Boston, MA, 1989.

HEINZLE, R. Protótipo de uma Ferramenta para Criação de Sistemas Especialistas Baseados em Regras de Produção. Florianópolis: novembro/1995 – Tese de Mestrado.

HIJG - HOSPITAL INFANTIL JOANA DE GUSMÃO Registro Hospitalar de Câncer do Hospital Infantil Joana de Gusmão (Santa Catarina) – 2004 a 2008. / Hospital Infantil Joana de Gusmão. – Florianópolis, HIJG, 2010.

HINRICHS T. R. Problem Solved in Open Worlds: A Case Study in Design. Northvale, NJ: Erbaun, 1992

HOLT, Alec et al. Medical Applications in Case-based Reasoning. *Knowledge Engineering Review*, v. 20, n. 3, p. 289-292, 2005.

IAN H.W. & Eibe F. *Data Mining: Practical Machine Learning Tools and Techniques* – 2nd ed. Morgan Kaufmann series in data management systems, p. 38, 2005

INCA Instituto Nacional de Câncer (Brasil). Coordenação de Prevenção e Vigilância de Câncer. *Câncer da criança e adolescente no Brasil: dados dos registros de base populacional e de mortalidade.* / Instituto Nacional de Câncer. – Rio de Janeiro: INCA, 2008.

INCA Tipos de Câncer Infantil em

<http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/infantil> acessado em 18/08/2013

JANSEN P. B. et al. *Mining Electronic Health Records: Towards Better Research Applications and Clinical Care* in Nature Reviews Vol 13, jun 2012

KOTON, P. Reasoning about Evidence in Causal Explanations. *In the Proceedings of the American Association for Artificial Intelligence.* , Morgan Kaufmann, San Mateo, CA, pp. 256-261, 1988.

KOLODNER, J. L., “Reconstructive memory: a computer model”. *Cognitive Science*, Vol. 7, pp. 281-328, 1983.

_____ Case-Based Reasoning. Morgan Kaufmann Publishers, San Mateo CA ,1993.

_____ An introduction to case-based reasoning. *Artificial Intelligence Review*, v. 6, n. 1, p. 3-34, 1992.

KOLODNER, J. L. & Kolodner, R. M. Using experience in clinical problem solving. *IEEE Transactions on Systems, Man, and Cybernetics* , 1987.

LEAKE, David B. *CBR in Context: The Present and Future em Case-based reasoning: Experiences, lessons and future directions.* MIT press, 1996.

LEE R. W. Pesquisa Jusprudencial Inteligente. Tese (Doutorado em Engenharia da Produção) – Programa de Pós-graduação em Engenharia da Produção, UFSC, Florianópolis, 1998.

LOPEZ, B, & Plaza, E, Case-Base Planning for medical diagnosis. *In the Proceedings of the Methodologies for Intelligent Systems, 7th. International Symposium, ISMIS-93.* Lecture Notes in Artificial Intelligence, Springer-Verlag, 689:96-105, 1993.

LOEB, S.; Terry, D. *Information Filtering*. Communications of ACM, New York, v.35, n.12, p.26, Dec, 1992.

KURZWEIL, R. *The Age of Intelligent Machines*, MIT Press, 1990

MACURA, RT, Macura, KJ, Toro, VE, Binet, EF, Trueblood, JH & Ji, K, 1994. Computerized case-based instructional system for computed tomography and magnetic resonance imaging of brain tumors. *Investigative Radiology*. 29(4):497-506.

MANTARAS, R. L., et al. "Retrieval, reuse, revision and retention in case-based reasoning." *The Knowledge Engineering Review* 20.03 (2005): 215-240.

MARLING C. R. , Shubrook J, & Schwartz F. Case-based Decision Support for Patients with type 1 Diabetes on Insulin Pump Therapy. In: Althoff KD, Bergman R, editors. Proc. European Conference on Case-based Reasoning, LNAI 5239. Berlin: Springer; 2008. p. 325–39.

MONTANI, S, Bellazzi, R, Portinale, L & Stefanelli, M, A Multi-Modal Reasoning Methodology for Managing IDDM Patients, *International Journal of Medical Informatics*, 58-59:243-256, 2000.

_____ RHENE: A Case Retrieval System for Hemodialysis Cases with Dynamically Monitored Parameters, ECCBR 2004, LNAI 3155, pp 659-672, Springer-Verlag Berlin Heidelberg, 2004

MONTANI S. Case-based Reasoning for Managing non-compliance with Clinical Guidelines. In Wilson DC, Khemani D, editors. Proceedings of Case-based Reasoning in the Health Sciences Workshop, International Conference of Case-based Reasoning (ICCBR). Berlin/Heidelberg/New York: Springer Verlag, 2007 p. 325-336

MV, 2013

<http://www.mv.com.br/mv/blogs/variados/2010/04/19/CNT,115,27,418,MV,2060-HOSPITAL.aspx> acessado em 21/04/2013

MYTHILI D. N., Agent-Based CBR for Decision Support System International Journal of Scientific & Engineering Research, Volume 2, Issue 1, 2011

NE10. <http://ne10.uol.com.br/canal/vestibular-2012/noticia/2011/11/20/preconceito-e-avalanche-de-informacoes-sao-temas-da-redacao-do-seriado-da-upe-310807.php>
acessado em 09/03/2013

NORVIG P., Stuart R. Artificial Intelligence *A Modern Approach*, Third Edition Prentice Hall, 2009

PERNER, P, An Architecture for a CBR Image Segmentation System, *Journal on Engineering Application in Artificial Intelligence*, 12(6):749-759, 1999.

PATEL, V. L., & Ramoni, M. F. Cognitive models of directional inference in expert medical reasoning. In P. J. Feltovich, & K. M. Ford (Eds.), *Expertise in Context: Human and Machine* (pp. 67–99): Cambridge, MA: MIT Press, 1997.

PETOT, G. C., Marling, C. R. & Sterling, L. S. An Artificial Intelligence System for Computer-Assisted Menu Planning, *Journal of the American Dietetic Association*, 98(9):1009-1014, 1999

PEIRCE, C. S. Abduction and Induction. In C. S. Peirce, & J. Buchler (Eds.), *Philosophical Writings of Peirce* (pp. 1 50–1 56). New York: Dover, 1955.

PIAGET J. *A Espistemologia Genética*. Petropolis, Vozes, 1972

POOLE D., Milekwordic A. K., & Goebel, P. *Computational intelligence, A logical approach*. Oxford University Press, 1999.

SANKAR K.P. e Simon C. K. S. Foundations of Soft Case-Based Reasoning. : Wiley, . p 41 Jonh Wiley & Sons , Inc, 2004 <http://site.ebrary.com/id/10114063?pg=41>

SBIS, O que é Informática na Saúde <http://www.sbis.org.br/indexframe.html> acessado em 25/11/2013

SCHANK, R.C. *Dynamic Memory: A theory of remaining and learning in computers and people*. Cambridge University Press, 1982.

SCHANK, R.C. and Albelson R. *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Earlbaun Assoc, 1997.

SCHMIDT, R. Gierl L. Case-based Reasoning for Medical Knowledge-based Systems. Proceedings of Medical Infobahn for Europe, 2000.

SHI, Z., *Advanced Artificial Intelligence WSPC*, 2011 Pag. 171-213 Acessado em 18/03/2013 www.ebrary.com.

SIMPSON, R. L. A Computer Model of Case-Based Reasoning in Problem Solving: An Investigation in the Domain of Dispute Mediation. Ph.D. Thesis. Technical Report No. GIT- ICS-85/18. School of Information and Computer Science, Georgia Institute of Technology, Atlanta, GA, 1985.

SLADE, Stephen. Case-based reasoning: A research paradigm. AI magazine, v. 12, n. 1, p. 42, 1991.

SPYROPOULOS (1999a), On-line Educational Means supporting the tutoring of the Physical Principles applied in modern Biomedical Equipment Technology, *Centennial Meeting of the American Physical Society, Atlanta 20-26 March, 1999, Bulletin of the American Physical Society, UB20.04, 44*, (1999), No 1, Part II, p. 1425.

SPYROPOULOS (1999b), I. Marneris, On-line Educational Means on Radiological Protection and Accelerator General Safety Policy in Radiotherapy and Industrial Sterilization Facilities, *Particle Accelerator Conference, New York City, March 29th-April 2nd, 1999*, WEP 124

SPYROPOULOS, Some Aspects of developed On-line Training Means in Biomedical Technology Management, *Chicago 2000, World Congress on Medical Physics and Biomedical Engineering, Chicago, IL, July 23 - 28, 2000*, TU-E206-4.

STRUBE, G., Enzinger, A., Janetzko, D., & Knauff, M. (1995). Knowledge engineering for CBR systems from a cognitive-science perspective. In: Veloso M & Aamodt A (eds). Case-based Reasoning Research and Development; Proceedings of the First International Conference on CBR. Springer

SYCARA, E. P. Resolving Adversarial Conflicts: An Approach to Integrating Case-Based and Analytic Methods. Ph.D. Thesis. Technical Report No. GIT-ICS-87/26, School of Information and Computer Science, Georgia Institute of Technology, Atlanta, GA, 1987.

TANENBAUN A. S. Redes de Computadores”, Tradução da 4ª edição, Rio de Janeiro, Ed. Campus, 2003.

THOMAS M. D. et al Case-based Teaching and Learning Experiences Issues in Mental Health Nursing, 22:517–531, 2001

Van Bommel J., Criteria for the Acceptance of Decision Support Systems by Clinicians, in: S. Andreassen, R. Engelbrecht and J. Wyatt, eds., AI in Medicine – Proc. AIME '93 (IOS Press, Amsterdam, 1993) 7-12.

VAN Den Branden, M. et al. Integrating Case-based Reasoning with an Electronic Patient Record System. *Artificial Intelligence in Medicine*, v. 51, n. 2, p. 117-123, 2011.

WELTER et al. Towards case-based medical learning in radiological decision making using content-based image retrieval. *BMC Medical Informatics and Decision Making* 2011, 11:68

VIMLA, L. P. et tal, The Cambridge Handbook of Thinking and Reasoning – Capítulo 20 Thinking and Reasoning in Medicine, p. 717-750, Cambridge University Press, 2005

WATSON Y., Marir F. Case-based reasoning: A review. *The Knowledge Engineering Review*, 9, pp 327-354. doi:10.1017/S0269888900007098, 1994

WILLIAN, W. S., Herbert S. L. et tal Computational Technology for Effective Health Care The National Academies press, 2009

WINSTON, P. *Artificial intelligence*, 3.ed. Reading, USA: Addison-Wesley, 1992.

ZHUANG, Z. Y. *Combining Data Mining and Case-based Reasoning for Inteligent Decision Support for Pathology Ordering by General Practitioners* in European Journal of Operational Research 195 pp 662–675, 2009

APÊNDICES

APÊNDICE A – DESCRIÇÃO DAS TABELAS

Tabela: ATENDIMENTO – contém os dados dos atendimentos dos pacientes

Atributo	Descrição
Cd_atendimento	Código do atendimento do paciente
Cd_ori_ate	Código do setor que originou o atendimento
Cd_paciente	Código (prontuário) do paciente
Tp_atendimento	Tipo do atendimento (I-internação, U-urgência)
Cd_prestador	Código do prestador (médico) que atendeu o paciente
Cd_cid	Código Internacional da Doença
Cd_ser_dis	Código do serviço
Cd_tip_internação	Código do tipo de internação
Cd_Especialid	Código da especialidade médica

Tabela: PACIENTE – prontuário do paciente, contém os dados básicos do paciente

Atributo	Descrição
Cd_paciente	Código (prontuário) do paciente
Nm_paciente	Nm_paciente
Tpsexo	Tpsexo
Dt_nascimento	Dt_nascimento

Tabela: PRE_MED – tabela que contém os dados das prescrições médicas

Atributo	Descrição
Cd_premed	Código da prescrição médica
Cd_atendimento	Código do atendimento do paciente
Cd_prestador	Código do prestador (médico) que atendeu o paciente
Ds_evolução	Descrição da evolução médica. Neste atributo é registrada a situação do paciente no atendimento, ou seja, o quadro do paciente tais como: pressão, aparências físicas e neurológicas e demais dados clínicos.
Sn_fechado	S-sim, N-Não
Tp_Premed	Tipo da prescrição médica.

Tabela: IT_PREMED – itens da prescrição médica, ou seja, cada prescrição pode gerar 1 (um) ou mais itens de prescrição, como, por exemplo, para uma prescrição (PREMED) pode gerar itens tais como um exame (imagem ou laboratório), dieta, medicamento ou um outro procedimento.

Atributo	Descrição
Cd_itpremed	Código do item da prescrição
Cd_tip_esq	Código do tipo de procedimento atendimento
Cd_premed	Código da prescrição médica
Cd_tip_fre	Código do tipo da frequência a ser aplicado o item
Ds_itpremed	Descrição do item da prescrição
Cd_tip_presc	Código do tipo da prescrição

Tabela: TIP_PRESC – tipo da prescrição

Atributo	Descrição
Cd_tip_presc	Código do tipo da prescrição
Cd_tip_esq	Código do tipo de procedimento atendimento
ds_tip_presc	Descrição do tipo da prescrição

APÊNDICE B – QUERIES DE EXTRAÇÃO DOS DADOS

EXTRAÇÃO DOS DADOS DOS PACIENTES

```
SELECT P.CD_PACIENTE,  
       P.NM_PACIENTE,  
       P.TP_SEXO,  
       P.DT_NASCIMENTO  
FROM DBAMV.PACIENTE P  
WHERE CD_PACIENTE IN (&1, &2, &3, &4, &5, &6, &7, &8, &9)  
ORDER BY 1
```

EXTRAÇÃO DOS DADOS DOS ATENDIMENTOS

```
SELECT A.CD_ATENDIMENTO,  
       A.CD_ORI_ATE,  
       A.CD_PACIENTE,  
       A.TP_ATENDIMENTO,  
       A.CD_PRESTADOR,  
       A.CD_CID,  
       A.CD_SER_DIS,  
       A.CD_TIPO_INTERNACAO,  
       A.CD_ESPECIALID,  
       A.DT_ATENDIMENTO  
FROM DBAMV.ATENDIME A  
WHERE A.CD_PACIENTE IN (&1, &2, &3, &4, &5, &6, &7, &8, &9)  
ORDER BY 1
```

EXTRAÇÃO DOS DADOS DAS PRESCRIÇÕES

```
SELECT ATE.CD_PACIENTE,  
       PM.CD_PRE_MED,  
       PM.CD_ATENDIMENTO,  
       PM.CD_PRESTADOR,  
       PM.DS_EVOLUCAO,
```

```

        PM.SN_FECHADO,
        PM.TP_PRE_MED,
        PM.CD_PRE_PAD
FROM DBAMV.ATENDIME ATE
        , DBAMV.PRE_MED PM
WHERE ATE.CD_PACIENTE IN (&1, &2, &3, &4, &5, &6, &7, &8, &9)
        AND ATE.CD_ATENDIMENTO = PM.CD_ATENDIMENTO
ORDER BY 1,3

```

EXTRAÇÃO DOS DADOS DOS ITENS DA PRESCRIÇÃO

```

SELECT ATE.CD_PACIENTE,
        IPM.CD_ITPRE_MED,
        IPM.CD_TIP_ESQ,
        IPM.CD_TIP_PRESC,
        IPM.CD_PRE_MED,
        IPM.CD_TIP_FRE,
        IPM.CD_FOR_APL,
        IPM.DS_ITPRE_MED,
        IPM.TP_SITUACAO,
        IPM.SN_CANCELADO,
        IPM.CD_UNIDADE,
        IPM.CD_PRODUTO,
        IPM.CD_UNI_PRO,
        IPM.DS_JUSTIFICATIVA,
        IPM.TP_JUSTIFICATIVA,
        IPM.SN_URGENTE
FROM DBAMV.ITPRE_MED IPM,
        DBAMV.PRE_MED PM,
        DBAMV.ATENDIME ATE
WHERE ATE.CD_PACIENTE IN (&1, &2, &3, &4, &5, &6, &7, &8, &9)
        AND PM.CD_PRE_MED = IPM.CD_PRE_MED
        AND ATE.CD_ATENDIMENTO = PM.CD_ATENDIMENTO

```

EXTRAÇÃO DOS DADOS DOS TIPOS DE PRESCRIÇÃO

```
SELECT TP.CD_TIP_PRESC,  
       TP.CD_TIP_ESQ,  
       TP.DS_TIP_PRESC  
FROM DBAMV.TIP_PRESC TP
```

APÊNDICE C – VISÃO MATERIALIZADA PRESCRICAO

```
CREATE MATERIALIZED VIEW PRESCRICAO AS
(SELECT PM.CD_PRE_MED ,
      IPM.CD_ITPRE_MED ,
      PM.TP_PRE_MED ,
      PM.DS_EVOLUCAO ,
      IPM.CD_TIP_ESQ ,
      IPM.DS_ITPRE_MED ,
      IPM.DS_JUSTIFICATIVA ,
      TP.DS_TIP_PRESC
FROM PREMED PM,
      ITPREMED IPM,
      TIP_PRESC TP
WHERE PM.CD_PRE_MED = IPM.CD_PRE_MED
      AND IPM.CD_TIP_PRESC = TP.CD_TIP_PRESC
      AND IPM.CD_TIP_ESQ = TP.CD_TIP_ESQ);
```

APÊNDICE D – PROCEDURE DE CARGA DA BASE DE CASOS

```
CREATE OR REPLACE PROCEDURE PRC_LOAD_BASE_CASOS AS
BEGIN
  FOR PRESC IN
    (SELECT PM.CD_PRE_MED ,
           IPM.CD_ITPRE_MED ,
           PM.TP_PRE_MED ,
           PM.DS_EVOLUCAO ,
           IPM.CD_TIP_ESQ ,
           IPM.CD_TIP_PRESC ,
           IPM.DS_ITPRE_MED ,
           IPM.DS_JUSTIFICATIVA ,
           TP.DS_TIP_PRESC
    FROM ATENDIMENTO ATE ,
         PREMED PM ,
         ITPREMED IPM ,
         TIP_PRESC TP
    WHERE ATE.CD_ATENDIMENTO = PM.CD_ATENDIMENTO
          AND PM.CD_PRE_MED = IPM.CD_PRE_MED
          AND IPM.CD_TIP_PRESC = TP.CD_TIP_PRESC
          AND IPM.CD_TIP_ESQ = TP.CD_TIP_ESQ
          AND ATE.CD_PACIENTE IN (&1, &2, &3, &4, &5, &6, &7, &8, &9)
          AND ATE.DT_ATENDIMENTO = SYSDATE -1)
  LOOP
    INSERT INTO PRESCRICAO VALUES
      (PRESC.CD_PRE_MED,
       PRESC.CD_ITPRE_MED,
       PRESC.TP_PRE_MED,
       PRESC.DS_EVOLUCAO,
       PRESC.CD_TIP_ESQ,
       PRESC.CD_TIP_PRESC,
       PRESC.DS_ITPRE_MED,
       PRESC.DS_JUSTIFICATIVA,
       PRESC.DS_TIP_PRESC);
  END LOOP;
  COMMIT;
END PRC_LOAD_BASE_CASOS;
```

**APÊNDICE E – CRIAÇÃO DA VISÕES VW_PRESCRICAO E
VW_PESCRICAO_DSC**

CREATE OR REPLACE VIEW MMCC.VW_PRESCRICAO

AS

```
(SELECT distinct P1.CD_PRE_MED,  
                P1.CD_ITPRE_MED,  
                P1.CD_TIP_PRESC  
FROM PRESCRICAO P1  
WHERE P1.DS_EVOLUCAO LIKE '%OSTEOSSARCOMA%'  
)
```

CREATE OR REPLACE FORCE VIEW MMCC.VW_PRESCRICAO_DSC

AS

```
(SELECT DISTINCT P1.CD_PRE_MED,  
                P1.CD_ITPRE_MED,  
                P1.DS_TIP_PRESC  
FROM PRESCRICAO P1  
WHERE P1.DS_EVOLUCAO LIKE '%NAUSEAS%'  
      AND P1.DS_EVOLUCAO LIKE '%NEUTROPENIA FEBRIL%'  
)
```

ANEXOS

ANEXO A – RESULTADOS PARA CARCINOMA DE NASOFARINGE

Carcinoma de nasofaringe são **tumores de cabeça e pescoço**. O termo nasofaringe é sinônimo de rinofaringe, cavidade posterior das fossas nasais e cavum. Nas evoluções encontramos termos como “carcinoma de nasofaringe” ou “CA de nasofaringe”.

Quantidade de registros com ocorrência do termo: 612

Número de prescrições médicas: 73

Quantidade de diferentes itens prescritos: 98

Quantidade de regras retornadas após a aplicação do algoritmo apriori: 9

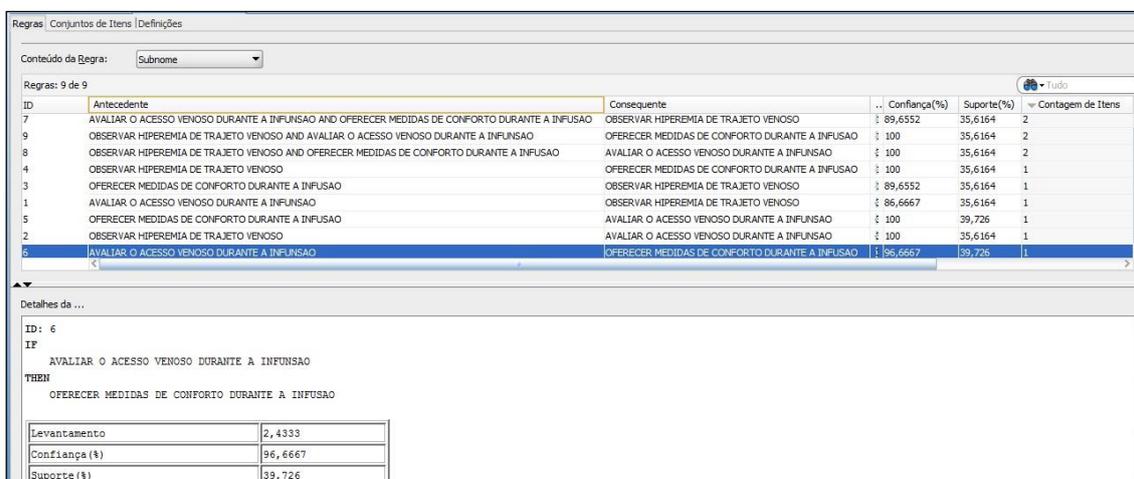
Medidas de Suporte e confiança:

Minsup: 35%

Minconf: 75%

Tamanho máximo da regra (contagem de itens): 6

Figura 27 - Resultado para Carcinoma de Nasofaringe



ID	Antecedente	Consequente	Confiança(%)	Suporte(%)	Contagem de Itens
7	AVALIAR O ACESSO VENOSO DURANTE A INFUSAO AND OFERECER MEDIDAS DE CONFORTO DURANTE A INFUSAO	OBSERVAR HIPEREMIA DE TRAJETO VENOSO	89,6552	35,6164	2
9	OBSERVAR HIPEREMIA DE TRAJETO VENOSO AND AVALIAR O ACESSO VENOSO DURANTE A INFUSAO	OFERECER MEDIDAS DE CONFORTO DURANTE A INFUSAO	100	35,6164	2
8	OBSERVAR HIPEREMIA DE TRAJETO VENOSO AND OFERECER MEDIDAS DE CONFORTO DURANTE A INFUSAO	AVALIAR O ACESSO VENOSO DURANTE A INFUSAO	100	35,6164	2
4	OBSERVAR HIPEREMIA DE TRAJETO VENOSO	OFERECER MEDIDAS DE CONFORTO DURANTE A INFUSAO	100	35,6164	1
3	OFERECER MEDIDAS DE CONFORTO DURANTE A INFUSAO	OBSERVAR HIPEREMIA DE TRAJETO VENOSO	89,6552	35,6164	1
1	AVALIAR O ACESSO VENOSO DURANTE A INFUSAO	OBSERVAR HIPEREMIA DE TRAJETO VENOSO	86,6667	35,6164	1
5	OFERECER MEDIDAS DE CONFORTO DURANTE A INFUSAO	AVALIAR O ACESSO VENOSO DURANTE A INFUSAO	100	39,726	1
2	OBSERVAR HIPEREMIA DE TRAJETO VENOSO	AVALIAR O ACESSO VENOSO DURANTE A INFUSAO	100	35,6164	1
6	AVALIAR O ACESSO VENOSO DURANTE A INFUSAO	OFERECER MEDIDAS DE CONFORTO DURANTE A INFUSAO	86,6667	39,726	1

Detalhes da ...

ID: 6
IF
AVALIAR O ACESSO VENOSO DURANTE A INFUSAO
THEN
OFERECER MEDIDAS DE CONFORTO DURANTE A INFUSAO

Levantamento	2,4333
Confiança (%)	86,6667
Suporte (%)	39,726

Fonte: o autor

ANEXO B – RESULTADOS PARA OSTEOSSARCOMA

Osteossarcoma é o **tumor ósseo maligno** mais frequente em pacientes pediátricos. É mais comum a partir dos 10 anos de idade, ocorrendo principalmente na adolescência, fase da vida em que o osso tem um crescimento rápido. Esse tumor atinge principalmente os ossos longos, sendo os mais comuns o fêmur e a tíbia, especialmente na região do joelho (Camargo, 2014).

Quantidade de registros com ocorrência do termo: 1567

Número de prescrições médicas: 130

Quantidade de diferentes itens prescritos: 152

Quantidade de regras retornadas após a aplicação do algoritmo apriori: 27

Medidas de Suporte e confiança:

Minsup: 40%

Minconf: 75%

Tamanho máximo da regra (contagem de itens): 4

Figura 28 - Resultado para Osteossarcoma

ID	Antecedente	Consequente	Confiança(%)	Suporte(%)	Contagem
27	DRAMIN B6 DL AMP 10 ML AND NISTATINA SUSP ORAL GEN EMS (5 ML) AND DIETA BRANDA P/ NEUTROPENICO	DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	100	40	3
24	DIETA BRANDA P/ NEUTROPENICO AND DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO AND NISTATINA SUSP ORAL GEN EMS (5 ML)	DRAMIN B6 DL AMP 10 ML	100	40	3
25	DRAMIN B6 DL AMP 10 ML AND DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO AND NISTATINA SUSP ORAL GEN EMS (5 ML)	DIETA BRANDA P/ NEUTROPENICO	98,1132	40	3
26	DRAMIN B6 DL AMP 10 ML AND DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO AND DIETA BRANDA P/ NEUTROPENICO	NISTATINA SUSP ORAL GEN EMS (5 ML)	98,1132	40	3
14	DRAMIN B6 DL AMP 10 ML AND DIETA BRANDA P/ NEUTROPENICO	NISTATINA SUSP ORAL GEN EMS (5 ML)	98,1132	40	2
21	NISTATINA SUSP ORAL GEN EMS (5 ML) AND DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	DIETA BRANDA P/ NEUTROPENICO	98,1132	40	2
18	NISTATINA SUSP ORAL GEN EMS (5 ML) AND DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	DRAMIN B6 DL AMP 10 ML	100	40,7692	2
22	DIETA BRANDA P/ NEUTROPENICO AND DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	NISTATINA SUSP ORAL GEN EMS (5 ML)	98,1132	40	2
15	DIETA BRANDA P/ NEUTROPENICO AND DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	DRAMIN B6 DL AMP 10 ML	100	40,7692	2

ID	Antecedente	Consequente	Confiança(%)	Suporte(%)	Contagem
12	DIETA BRANDA P/ NEUTROPENICO AND NISTATINA SUSP ORAL GEN EMS (5 ML)	DRAMIN B6 DL AMP 10 ML	98,1132	40	2
23	DIETA BRANDA P/ NEUTROPENICO AND NISTATINA SUSP ORAL GEN EMS (5 ML)	DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	98,1132	40	2
19	DRAMIN B6 DL AMP 10 ML AND DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	NISTATINA SUSP ORAL GEN EMS (5 ML)	98,1481	40,7692	2
16	DRAMIN B6 DL AMP 10 ML AND DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	DIETA BRANDA P/ NEUTROPENICO	98,1481	40,7692	2
20	DRAMIN B6 DL AMP 10 ML AND NISTATINA SUSP ORAL GEN EMS (5 ML)	DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	100	40,7692	2
13	DRAMIN B6 DL AMP 10 ML AND NISTATINA SUSP ORAL GEN EMS (5 ML)	DIETA BRANDA P/ NEUTROPENICO	98,1132	40	2
17	DRAMIN B6 DL AMP 10 ML AND DIETA BRANDA P/ NEUTROPENICO	DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	100	40,7692	2
6	DRAMIN B6 DL AMP 10 ML	DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	100	41,5385	1
4	DRAMIN B6 DL AMP 10 ML	NISTATINA SUSP ORAL GEN EMS (5 ML)	98,1481	40,7692	1

ID	Antecedente	Consequente	Confiança(%)	Suporte(%)	Contagem
2	DRAMIN B6 DL AMP 10 ML	DIETA BRANDA P/ NEUTROPENICO	98,1481	40,7692	1
7	HIGIENE ORAL COM CEPACOL	DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	92,8571	40	1
10	DIETA BRANDA P/ NEUTROPENICO	DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	96,3636	40,7692	1
9	DIETA BRANDA P/ NEUTROPENICO	NISTATINA SUSP ORAL GEN EMS (5 ML)	96,3636	40,7692	1
1	DIETA BRANDA P/ NEUTROPENICO	DRAMIN B6 DL AMP 10 ML	96,3636	40,7692	1
11	NISTATINA SUSP ORAL GEN EMS (5 ML)	DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	96,3636	40,7692	1
8	NISTATINA SUSP ORAL GEN EMS (5 ML)	DIETA BRANDA P/ NEUTROPENICO	96,3636	40,7692	1
3	NISTATINA SUSP ORAL GEN EMS (5 ML)	DRAMIN B6 DL AMP 10 ML	96,3636	40,7692	1
5	DIPIRONA SODICA 500 MG/ML AMP 2 ML GEN TEUTO	DRAMIN B6 DL AMP 10 ML	76,0563	41,5385	1

Fonte: o autor

ANEXO C – RESULTADOS PARA TUMOR DE WILMS

O tumor de Wilms é o **tumor renal maligno** mais frequente em crianças. A faixa etária de aparecimento é entre 2 a 4 anos. Frequentemente o tumor aparece apenas em um rim. Em cerca de 5% dos casos pode haver tumor bilateral, comprometendo os dois rins (Camargo, 2014).

Quantidade de registros com ocorrência do termo: 1489

Número de prescrições médicas: 135

Quantidade de diferentes itens prescritos: 135

Quantidade de regras retornadas após a aplicação do algoritmo apriori: 2

Medidas de Suporte e confiança:

Minsup: 30%

Minconf: 75%

Tamanho máximo da regra (contagem de itens): 4

Figura 29 - Resultado para Tumor de Wilms

ID	Antecedente	Consequente	Levantamento	Confiança(%)	Suporte(%)	Contagem...
1	FISIO. RESPIRAT./PACIENTE CLINICO INTERNADO	POLARAMINE 2 MG / 5 ML XAROPE	2,5267	95,4545	31,1111	1
2	POLARAMINE 2 MG / 5 ML XAROPE	FISIO. RESPIRAT./PACIENTE CLINICO INTERNADO	2,5267	82,3529	31,1111	1

ID: 2	
IF	
POLARAMINE 2 MG / 5 ML XAROPE	
THEN	
FISIO. RESPIRAT./PACIENTE CLINICO INTERNADO	
Levantamento	2,5267
Confiança (%)	82,3529
Suporte (%)	31,1111

Fonte: o autor

Neste caso podemos observar que a aplicação de Fisioterapia (*f*) Respiratória ocorre em conjunto com a aplicação de Plaramine (*p*). Importante verificar que os valores das medidas de suporte e confiança nas duas regras geradas não é comutativo, ou seja:

$$\text{sup}(f) \rightarrow (p) \neq \text{sup}(p) \rightarrow (f),$$

$$\text{conf}(f) \rightarrow (p) \neq \text{conf}(p) \rightarrow (f)$$

ANEXO D – RESULTADOS PARA MEDICAMENTOS TUMOR DE WILMS

Neste caso selecionamos as prescrições com ocorrência de Tumor de Wilms cujos itens prescritos sejam apenas medicamentos.

$$\text{Regra} = [med_1, med_2, \dots, med_n] \rightarrow med_x$$

Quantidade de registros com ocorrência do termo: 471

Número de prescrições médicas: 77

Quantidade de diferentes itens prescritos: 35

Quantidade de regras retornadas após a aplicação do algoritmo apriori: 10

Medidas de Suporte e confiança:

Minsup: 40%

Minconf: 75%

Tamanho máximo da regra (contagem de itens): 2

Figura 30 - Resultado para Medicamentos em Caso de Tumor de Wilms

ID	Antecedente	Consequente	Levantamento	Confiança(%)	Suporte(%)	Contage...	Suporte de Antecede...	Suporte Cor
8	POLARAMINE 2 MG / 5 ML XAROPE AND NI...	REPARIL GEL BISNAGA	2,1389	100	41,5584	2	46,7532	41,5584
3	REPARIL GEL BISNAGA	NISTATINA SUSP ORAL GEN...	2,1389	94,4444	44,1558	1	44,1558	46,7532
2	NISTATINA SUSP ORAL GEN EMS (5 ML)	REPARIL GEL BISNAGA	2,1389	100	44,1558	1	46,7532	44,1558
10	REPARIL GEL BISNAGA AND POLARAMINE ...	NISTATINA SUSP ORAL GEN...	2,1315	94,1176	41,5584	2	44,1558	44,1558
6	SOLUCAO GLICOFISIOLOGICA SIST FECH...	PLAMET 10 MG/2 ML AMP (I...	1,7907	100	48,0519	1	55,8442	48,0519
7	PLAMET 10 MG/2 ML AMP (INTRAVENOSO)	SOLUCAO GLICOFISIOLOGI...	1,7907	86,0465	48,0519	1	48,0519	55,8442
5	HIDROCORTISONA 100 MG FR/AMP + DIL...	POLARAMINE 2 MG / 5 ML X...	1,4641	96,9697	41,5584	1	66,2338	42,8571
1	REPARIL GEL BISNAGA	POLARAMINE 2 MG / 5 ML X...	1,4259	94,4444	44,1558	1	66,2338	46,7532
9	REPARIL GEL BISNAGA AND NISTATINA S...	POLARAMINE 2 MG / 5 ML X...	1,4210	94,1176	41,5584	2	66,2338	44,1558
4	NISTATINA SUSP ORAL GEN EMS (5 ML)	POLARAMINE 2 MG / 5 ML X...	1,4210	94,1176	41,5584	1	66,2338	44,1558

Fonte: o autor

Neste caso é importante observar que, quando focamos apenas nos medicamentos, ocorre uma mudança significativa nas regras, como por exemplo, na regra 2 que mostra;

- Do total de prescrições médicas (77) o medicamento Nistatina Susp. Oral Gen ENS (5ml) é prescrito em 34 prescrições, 44,15% do total de prescrições e em 100% destas ocorrências o medicamento Reparil Gel é prescrito conjuntamente.