

Integração entre Banco de Dados e Redes de Bayes no Suporte à Medicina

Ademar Rosa de Souza, Anderson Francisco Talon

Departamento de Computação
Faculdade de Tecnologia de Bauru (FATEC) – Bauru, SP – Brasil
ademaroza@gmail.com, anderson.talon@fatec.sp.gov.br

Abstract. *Given the large increase in scientific knowledge in medicine, where, it is not always possible for most physicians to assimilate them, the aim of this project is the creation of a system to support medical decision, based on the use of one relational database (MySQL), the use of Bayes networks and artificial intelligence techniques. In order to get a broader and faster dissemination of human knowledge, the system will be available on the Internet. It was noted a good operability of the system with their lists of differential diagnoses that should be remembered and analyzed by the attending physician.*

Keywords: *Decision Support Systems, Bayesian Networks, Medicine.*

Resumo. *Tendo em vista o grande aumento do conhecimento científico na medicina, onde, nem sempre, é possível para a maioria dos médicos assimilá-los, objetivou-se com este projeto a criação de um sistema de apoio à decisão médica, baseado na utilização de um banco de dados relacional (MySQL), no uso das Redes de Bayes e em técnicas de inteligência artificial. Com o intuito de obter uma maior abrangência e rapidez na disseminação do conhecimento humano, o sistema estará disponível na internet. Notou-se uma boa operacionalidade do sistema com suas listas de diagnósticos diferenciais que devem ser lembrados e analisados pelo médico assistente.*

Palavras-chave: *Sistemas de Apoio à Decisão, Redes de Bayes, Medicina.*

1. Introdução

Com o avanço da medicina, o conhecimento científico ganha proporções gigantescas, onde, nem sempre, é possível para a maioria dos médicos assimilá-los.

O conhecimento sobre qualquer patologia pode ser facilmente encontrado na internet, mas dificilmente encontra-se algum mecanismo que faça o raciocínio e a análise entre os dados obtidos de um mesmo paciente e chegue a uma conclusão para o problema.

Em nosso cotidiano, existe uma necessidade crescente de diagnósticos médicos rápidos e precisos, em virtude de uma maior demanda na área da saúde. Nesse contexto, espera-se que a inteligência artificial (apoiada num banco de dados com informações médicas) possa contribuir de forma significativa.

Atualmente, existem várias pesquisas tentando aplicar a inteligência artificial na área médica (através dos sistemas de apoio à decisão médica). Apesar disso, estes sistemas ainda não estão sendo utilizados de forma rotineira. Existem várias justificativas: a) necessidade prévia de um sistema informatizado de registros médicos, b) programas com interface não amigável entre o usuário e o sistema, c) excessiva perda de tempo no fornecimento dos dados iniciais do paciente contrastando com a necessidade de respostas rápidas nos ambientes médicos, etc. (WESTPHASL, 2003).

Em virtude disso, pretende-se desenvolver um sistema de apoio à decisão médica (utilizando técnicas de inteligência artificial e funcionando via internet) tornando o sistema totalmente simples de operar e compreender.

O objetivo deste projeto será unir duas áreas do conhecimento humano (medicina e informática) com o intuito de otimizar e agilizar de forma confiável os diagnósticos médicos. O fato de um dos autores ser médico, com especialização em Cirurgia Cardíaca, tornará a condução do trabalho mais harmoniosa, detectando precocemente os pontos críticos do sistema.

2. Referencial Teórico

Geralmente, as informações clínicas iniciais dos pacientes são frequentemente imperfeitas, subjetivas ou não específicas. As possíveis hipóteses são numerosas para serem examinadas individualmente, cabendo ao médico reduzir ao máximo esta incerteza através de informações complementares (LINARES, 1997).

Neste contexto, a informática pode prestar um grande auxílio na tomada de decisões com melhoria na qualidade dos diagnósticos.

2.1 Inteligência Artificial (IA)

A Inteligência Artificial (IA) nasceu oficialmente em 1956, durante uma conferência de verão em Dartmouth College, nos Estados Unidos. Ela surgiu do desejo de fazer com que os programas de computadores utilizassem metodologias baseadas na forma como os seres humanos pensam e resolvem problemas, assim, utilizam mecanismos que simulam a inteligência humana.

Segundo Haykin (2008), um sistema de IA deve ser capaz de: a) Armazenar conhecimento, b) Aplicar o conhecimento armazenado para resolver problemas, c) Adquirir novo conhecimento através da experiência.

2.2 Sistemas Especialistas (SE)

Dentre as áreas de pesquisa da IA, encontram-se os SE, os quais atuam na forma de sistemas interativos, respondendo questões, solicitando ou fornecendo esclarecimentos e recomendações, podendo também auxiliar o usuário na tomada de decisões. De modo geral, podem simular o raciocínio humano ao fazer inferências, julgamentos ou projetar resultados (LINARES, 1997).

São utilizados quando um problema não pode ser algoritmizado (não possua regras ou processos claramente definidos), ou sua solução conduza a um processamento muito demorado. Dentre seus objetivos, está o de preservar e transmitir o conhecimento de algum especialista em determinada área (ROSSO, 2002).

Os SE possuem o seu mecanismo baseado em regras heurísticas, que são uma implicação lógica da forma:

SE <condição> **ENTÃO** <consequência 1> **SENÃO** <consequência 2>

De maneira geral, um SE é formado por três elementos básicos (Figura 1):

- **Interface de aquisição de conhecimento:** interface amigável para obtenção do conhecimento do especialista;
- **Base de conhecimento / Mecanismo de inferência:** fatos e dados que representam o conhecimento do especialista, heurísticas, esquemas de raciocínio, inferências;
- **Interface com o usuário:** linguagem natural, explicação, esquema de interação com o usuário final.

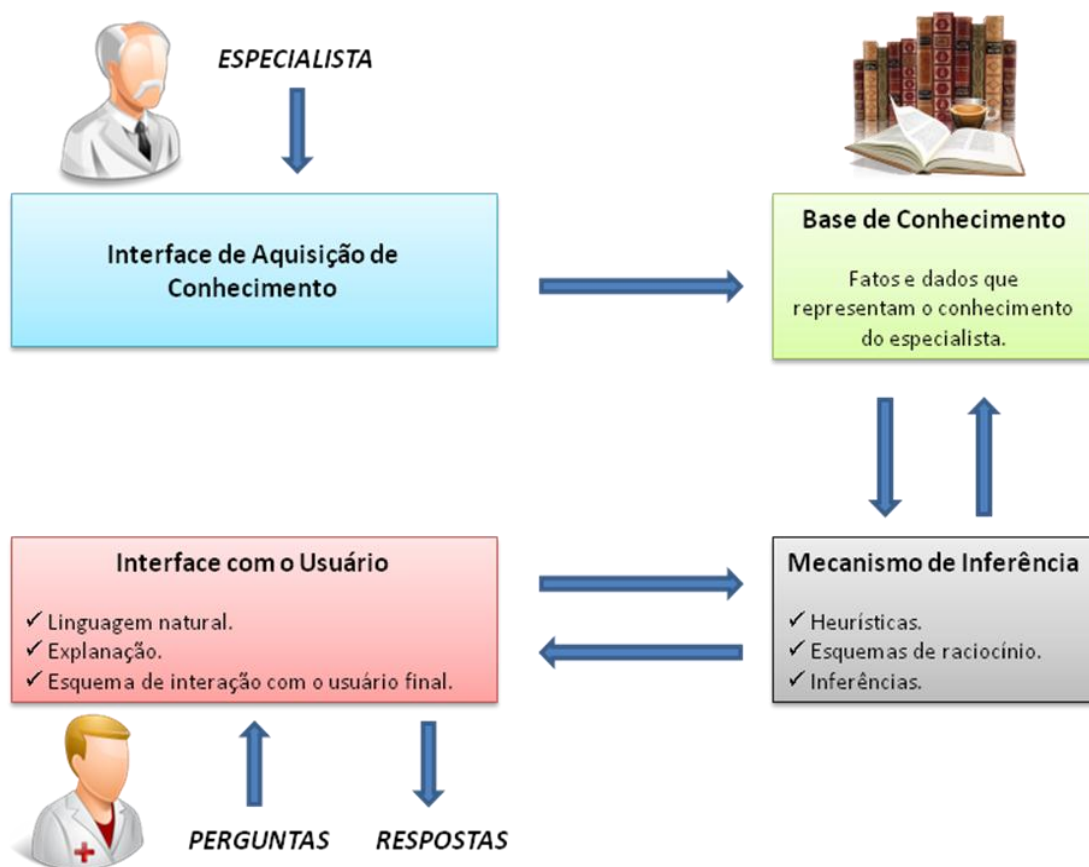


Figura 1. Esquema clássico de um Sistema Especialista, típico de numerosas áreas de aplicação da inteligência artificial. Adaptado de Mendes (1997).

3. Material e Métodos

De origem grega, a palavra diagnóstico significa discernir ou distinguir.

Dentro do contexto médico, é “uma série de procedimentos de ordem intelectual e operacional através dos quais se obtém uma resposta a um problema clínico” conforme definido por Mason (citado por Andrade 1999).

Segundo Linden (2012) os problemas do mundo real podem ser resolvidos através de técnicas de busca, bastando definir onde o problema se inicia e o objetivo aonde se deseja chegar.

Basicamente, tudo que existe na mente humana possui uma relação, e tudo que existe no mundo possui uma relação no tempo e espaço.

“O conhecimento depende da distinção entre os atributos dos objetos do mundo. Sem a distinção, sem o relacionamento, não há conhecimento. Relação define, portanto, conhecimento!” (KREPSKY, 1999)

Dentro deste contexto, optou-se por trabalhar neste projeto com banco de dados relacionais, os quais possuem em sua estrutura a possibilidade de se criar relacionamentos entre os objetos, além de mecanismos de busca muito eficazes e poderosos.

3.1 Banco de Dados

Um banco de dados é um aplicativo que pode armazenar uma grande quantidade de informações e recuperar estes dados de maneira muito rápida.

O modelo mais usado atualmente é o de Banco de Dados Relacional (DIAS NETO, Online), o qual surgiu em meados da década de 1970. Nesse modelo, um banco de dados pode ser composto por várias tabelas, onde os dados de uma tabela podem estar ligados com os dados de outra através de relacionamentos.

3.2 Imprecisão e Incerteza

A modelagem do processo de decisão clínica é uma tarefa complexa, em virtude da imprecisão das informações iniciais e da incerteza das hipóteses diagnósticas.

Hipótese, para a estatística, é o termo utilizado para alguma proposição da qual sua verdade ou falsidade não é conhecida, mas está assegurada com base em alguma evidência.

Tendo em vista que o raciocínio médico é mais baseado em probabilidades (ANDRADE, 1999) e que existem métodos para representar a imprecisão e/ou incerteza, optou-se por adotar o uso das Redes de Bayes, visando uma inferência mais aproximada em relação à realidade.

3.3 Redes de Bayes

Também chamadas de redes de crença, redes probabilísticas ou redes causais, podem ser vistas como um modelo de distribuição de probabilidades para representar uma situação, suas variáveis e estados, e a partir disto realizar inferências. São compostas por várias variáveis de entrada e uma variável de saída (KUBO, 2011).

As Redes Bayesianas (RB) foram desenvolvidas pelo professor Thomas Bayes em 1763 com o intuito de auxiliar o trabalho de predição. São elaboradas a partir de modelos e bases de incertezas e baseiam-se na utilização de fórmulas matemáticas para o cálculo de probabilidades (KUBO, 2011). É uma base atraente para um sistema de raciocínio sob incerteza (LINARES, 1997).

“As RB são grafos acíclicos e direcionados que permitem a representação da distribuição de probabilidades conjunta de um conjunto de variáveis aleatórias.” (KARCHER, 2009) (Figura 2)

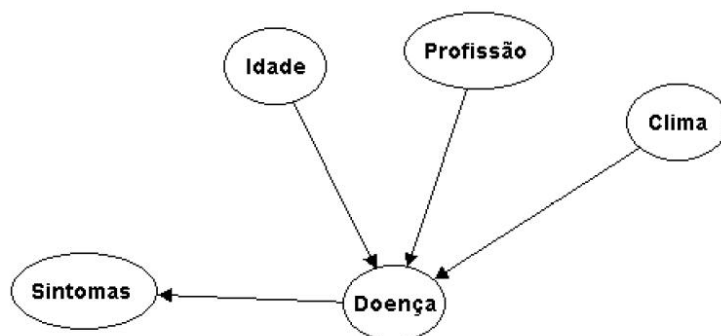


Figura 2. Exemplo de Rede Bayesiana aplicada no diagnóstico médico. (KARCHER, 2009)

Segundo Russell e Norvig (2004), as inferências sobre as redes Bayesianas podem ser realizadas de quatro maneiras distintas: 1) Diagnósticos: partindo dos efeitos para as causas; 2) Causa: partindo das causas para os efeitos; 3) Intercausal: entre causas de um efeito comum; 4) Mistas: combinação de dois ou mais tipos descritos acima.

Quando aplicadas em problemas de classificação de dados passam a ser chamadas de Classificadores Bayesianos, os quais têm como objetivo descrever e distinguir classes: pode ser obtido um sistema de pesos ou escores, sendo atribuídos escores positivos para testes ou achados clínicos em função principalmente da **especificidade** (se presentes ou positivos) e escores negativos em função da **sensibilidade** (se ausentes ou negativos) (ANDRADE, 1999).

Esta proposta de se fazer diagnósticos através de escores é vista com simpatia, pois é muito familiar aos clínicos.

Segundo Russell e Norvig (2004), “A principal vantagem de raciocínio probabilístico sobre raciocínio lógico é fato de que agentes podem tomar decisões racionais mesmo quando não existe informação suficiente para se provar que uma ação funcionará”.

Probabilidade condicional (a probabilidade de ocorrer um evento, na condição de que outro evento já tenha ocorrido): $P(a|b) = x$

Pode ser interpretada como: “Dado o evento **b**, a probabilidade do evento **a** é **x**”.

Regra de Bayes: $P(b|a) = P(a|b) * P(b)/P(a)$

Exemplo de aplicação no diagnóstico médico:

Um médico sabe que a meningite (M) causa torcicolo (T) em 50% dos casos. Porém, o médico também conhece algumas probabilidades incondicionais que dizem que, um caso de meningite atinge 1/50000 pessoas e, a probabilidade de alguém ter torcicolo é de 1/20.

Aplicando a regra de Bayes:

$$P(M | T) = \frac{P(M) \times P(T | M)}{P(T)} = \frac{1/50000 \times 0.5}{1/20} = 0.0002$$

Esta é a probabilidade de que a pessoa tenha meningite, baseado apenas no sintoma torcicolo.

3.4 Classificador *Naive Bayes*

Dentre os classificadores Bayesianos, o classificador *Naive Bayes* é o mais conhecido, em virtude de sua simplicidade e eficiência, destacando-se pelos sucessos obtidos em diversos problemas, mesmo se comparado com outros classificadores mais complexos.

Os Classificadores *Naive Bayes* levam em conta que todos os atributos são condicionalmente independentes, ou seja, uma característica não é relacionada com a outra; embora esta hipótese raramente se verifique no mundo real. De qualquer modo

eles têm apresentado um bom desempenho se as variáveis preditoras não forem fortemente correlacionadas. (KARCHER, 2009)

A probabilidade incondicional da classe pode ser obtida por meio do conhecimento de um especialista ou atribuindo probabilidades iguais para todas as classes.

A função de um Classificador *Naive Bayes* neste projeto será a classificação das hipóteses diagnósticas de maior probabilidade (a partir das hipóteses pré-selecionadas pelo banco de dados relacional), tornando o modelo mais parcimonioso (com menor número de variáveis) e mais poderoso, aumentando o desempenho do sistema e auxiliando no aumento das taxas de acerto das previsões.

3.5 Dicionário de sinônimos

Com o intuito de facilitar a interatividade com o usuário, a estrutura do projeto foi elaborada permitindo a utilização de palavras sinônimas.

Exemplo 1: *Catapora = Varicela*.

Exemplo 2 (regra utilizando sinônimos):

SE (cão/cachorro) **E** (mordeu/mordida/mordedura)

ENTÃO doença provável = **Raiva** (Probabilidade = x%)

3.6 Dicionário de semelhantes

Visando englobar o maior leque possível de doenças a partir dos dados ou sintomas fornecidos pelo usuário, também foi elaborado um dicionário de itens semelhantes.

Exemplo: Se o usuário digitar que o paciente possui o sintoma "febre", o sistema procurará também por doenças que apresentem quaisquer um dos seguintes sintomas: "febre com calafrios", "febre não especificada" ou "febre persistente". Neste caso, os sintomas são semelhantes, porém não iguais (não são sinônimos). Algumas vezes, podem até serem interpretados como graduações de um mesmo item.

3.7 Do Ponto de Vista Ético

É vedado ao médico prescrever tratamentos sem exame direto do paciente. A exceção acontece em casos de urgência e, mesmo assim, a consulta deve acontecer assim que terminado o impedimento. A determinação está no artigo 37 do Código de Ética Médica e foi enfatizada na Resolução 1.974/11, do Conselho Federal de Medicina (CFM), que atualiza regras para divulgação de serviços prestados.

“A relação médico-paciente é insubstituível. Interferências podem induzir a conclusões errôneas” (CENÇO, 2011).

Portanto, neste projeto, o sistema especialista será apenas um Sistema de Apoio à Decisão Médica (SADM) e não um sistema de decisão. Não é destinado a substituir qualquer médico especialista em determinada área, mas sim auxiliar o médico clínico geral de alguma pequena cidade (distante do especialista) a diagnosticar corretamente uma doença.

“Quanto menos avançado é o centro médico, mais se precisa de apoio à decisão”, diz Sigulem (citado por Cardoso, Online).

3.8 Tecnologias utilizadas no projeto

Foi adotada a plataforma .NET, ou seja, as páginas na internet foram elaboradas em ASP.NET em combinação com a linguagem *Visual Basic*.

Foi adotado o banco de dados relacional MySQL, por ser voltado para a internet e com suporte a múltiplas conexões. É um banco livre e gratuito, sendo seu código fonte aberto, além de ser fácil de usar e com melhorias contínuas. Possui uma grande capacidade de armazenamento, além de possibilitar a incorporação de arquivos de áudio e vídeo, dentre outros.

3.9 Resultados

Visando uma maior abrangência e rapidez na disseminação do conhecimento humano, o sistema estará disponível através da internet.

Estarão disponíveis para acesso público algumas páginas, tais como: "*Dicionário*", contendo todas as palavras utilizadas pelo sistema; "*Sintomas*" e todas as "*Doenças*" contidas na Classificação Internacional de Doenças (CID 10), num total de 12.422 itens cadastrados.

Para os usuários médicos cadastrados e devidamente liberados, estará disponível uma página de "*Consulta*", onde podem ser informados os sintomas e/ou sinais clínicos do paciente. Ocorrerá uma seleção das doenças mais prováveis, sendo exibidas em ordem decrescente de probabilidade.

Para os especialistas médicos, estarão disponíveis outras páginas, contendo as "*Variáveis*", onde poderão ser informadas ou cadastradas quaisquer tipos de variáveis a serem utilizadas pelo sistema; uma página de "*Semelhantes*", onde poderão ser informados os dados considerados semelhantes, sejam eles sintomas, doenças, exames ou medicamentos; e outra página contendo as "*Regras*", onde serão informadas ao sistema as mais variadas regras encontradas na medicina.

Nas Figuras 3 e 4 podem ser visualizadas a patologia hipercalcemia com os seus respectivos sintomas.

Nas Figuras 5 e 6 podem ser visualizadas a simulação de uma consulta, onde são informados aleatoriamente alguns sintomas e o sistema sugere as prováveis hipóteses diagnósticas.

Doença e seus sinônimos	Padrão
Hipercalcemia	☒
1	

Figura 3. Formulário contendo um quadro da doença hipercalcemia

Sintomas desta doença	Prob. (%)
Anorexia	
Assistolia	
Astenia	
Coma	
Confusão mental	
Constipação	
Letargia	
Náusea e vômitos	
Poliúria	
Tremores	
1	

Figura 4. Sintomas correspondentes à doença hipercalcemia.

Dados do paciente	Tempo (dias)
Sintoma Cefaléia	 
Sintoma Convulsões	 
Sintoma Febre com calafrios	 
Sintoma Náusea e vômitos	 
1	

Figura 5. Exemplo fictício de um paciente com alguns sintomas.

DOENÇAS PROVÁVEIS		
	Doença	Prob. (%)
1ª	Shigeloses	2
2ª	Babesiose	1,5
2ª	Malária	1,5
2ª	Uremia	1,5
3ª	Tumor cerebral	1,33
1 2 3 4 5 ...		

Figura 6. Lista de prováveis doenças em ordem decrescente de probabilidade (baseado no paciente fictício citado acima).

Nesta fase inicial do projeto, foram disponíveis apenas as doenças com seus respectivos sintomas, preferencialmente da área cardiológica, em virtude da maior familiaridade de um dos autores com o tema.

As informações serão gradativamente inseridas no sistema através de pesquisa em livros médicos ou qualquer outra fonte segura de informação da área médica (artigos científicos, *sites* confiáveis, etc.).

Encontrar fontes médicas contendo as probabilidades dos sintomas ou das doenças foi uma das grandes dificuldades encontradas para se formular sistemas baseados em probabilidades. Neste projeto, quando estes dados não são informados, padronizou-se que o sistema adotará automaticamente o valor de 50%.

4. Conclusão

Mesmo estando o projeto em fase inicial, nota-se um sistema intuitivo e totalmente fácil de operar, gerando pelo menos uma ampla lista de diagnósticos diferenciais que devem ser lembrados ou analisados pelo médico assistente; atendendo ao propósito de ser um sistema de apoio à decisão médica.

À medida em que forem inseridas ou cadastradas mais informações médicas no sistema, isto o tornará cada vez mais confiável.

Pretende-se com isso criar uma grande interação entre o usuário e o computador, onde o sistema poderá sugerir ou interrogar por exames/procedimentos tentando tornar

Como projeto futuro, pretende-se inserir módulos contendo lista de exames médicos, lista de medicamentos disponíveis (com seus respectivos efeitos e interações medicamentosas), protocolos clínicos (padronizados por centros de referência), fluxogramas ou rotinas de condutas médicas, além de fotos e vídeos sobre qualquer tópico da área médica.

Como a estrutura básica do projeto está definida, o acréscimo de outras especialidades médicas dependerá apenas da incorporação em seu banco de dados das respectivas regras, ou seja, não haverá necessidade de reformulação em termos de programação.

Referências

- ANDRADE, P. J. N. Sistemas Especialistas de Apoio ao Diagnóstico em Medicina: relações com o teorema de Bayes e com a lógica do raciocínio diagnóstico, *Arq Bras Cardiol.* v. 73, n. 6, p. 537-544, 1999.
- CARDOSO, F. *A inteligência do chip*: softwares que imitam o raciocínio humano, Disponível em: < <http://super.abril.com.br/tecnologia/inteligencia-chip-softwares-imitam-raciocinio-humano-440972.shtml>>. Acesso em: 08 out. 2011.
- CENÇO, B. A consulta no carrinho virtual, *Revista da APM.* Edição 626, p. 36-37, 2011
- DIAS NETO, A. C. *Banco de Dados Relacionais* - Artigo Revista SQL Magazine 86. Disponível em :<<http://www.devmedia.com.br/bancos-de-dados-relacionais-artigo-revista-sql-magazine-86/20401>>. Acesso em 19 mai. 2013
- HAYKIN, S. Introdução. In _____. *Redes neurais: princípios e práticas*. 2. ed. São Paulo-SP: Artmed Editora. 2008. p. 27-74.
- KARCHER, C. *Redes Bayesianas aplicadas à análise do risco de crédito*, Dissertação (Mestrado em Sistemas Eletrônicos) - Escola Politécnica, Universidade de São Paulo, São Paulo. 2009. Disponível em:

<<http://www.teses.usp.br/teses/disponiveis/3/3142/tde-25052009-162507/>>. Acesso em: 2012-07-18.

KREPSKY, W. *Protótipo de um interpretador para um ambiente de programação lógica*, 70f. Trabalho de conclusão de curso em Ciências da Computação. Universidade Regional de Blumenau, Blumenau. 1999.

KUBO, A. Y.; TENORIO, M. B.; AVARIN, S. Rede Bayesiana Aplicada à Tomada de Decisão na Produção do Leite Bovino, *Sapere*. v. 3, n. 1, 2011. Disponível em: <http://www.revistasapere.inf.br/download/quarta/ADRIANA_MARCELO_SIDNEI.pdf>. Acesso em: 10 jul. 2012.

LINARES, K. S. C. *Sistema Especialista Nebuloso para Diagnóstico Médico*, 116f. Dissertação (Mestrado em Engenharia Elétrica). Universidade Federal de Santa Catarina, Florianópolis-SC. 1997.

LINDEN, R. *Algoritmos Genéticos*. 3. ed. Rio de Janeiro-RJ: Ciência Moderna. 2012. 475p.

MENDES, R. D. Inteligência Artificial: sistemas especialistas no gerenciamento da informação, Brasília: *Ciência da Informação*. v. 26, n. 1, jan./apr. 1997. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651997000100006&lng=en&nrm=iso&tlng=pt>. Acesso em: 23 jun. 2011.

ROSSO, M. *Sistema Especialista de Apoio à Decisão em Ventilação Mecânica*. In *VIII Congresso Brasileiro de Informática em Saúde*, Natal. Anais eletrônicos do VIII Congresso Brasileiro de Informática em Saúde. 2002. Disponível em: <<http://www.avesta.com.br/anais/dados/trabalhos/226.pdf>> Acesso em 23 ago. 2011.

RUSSEL, S.; NORVIG, P. *Inteligência artificial*. 2. ed. Rio de Janeiro-RJ: Elsevier. 2004. 1021p.

WESTPHAL, J. T. *Modelagem difusa de um sistema especialista médico: avaliação dos fatores de internação em crianças queimadas*, 123f. Dissertação (Mestrado em Ciência da Computação). Universidade Federal de Santa Catarina, Florianópolis. 2003.