

AQUISIÇÃO DE CONHECIMENTO: O GRANDE DESAFIO NA CONCEPÇÃO DE SISTEMAS ESPECIALISTAS

Welbson Siqueira Costa¹ e Shirly Christiany Macedo Silva²

¹ Aluno do Mestrado em Sistemas e Computação da UFRN. Tecnólogo em Informática pelo CEFET/RN. welbson@ppgsc.ufrn.br / welbsonsiqueira@yahoo.com.br

² Aluna do Mestrado em Sistemas e Computação da UFRN. Graduada em Ciências da Computação pela UERN. shirly@ppgsc.ufrn.br

Artigo Recebido em Agosto de 2004 e Aceito em Setembro de 2005

RESUMO

Durante as últimas décadas o mundo vem se tornando um ambiente de convívio totalmente dependente das informações. Nesse contexto surgem os Sistemas Especialistas. Uma das etapas na construção desses sistemas apresenta bastante dificuldade, a Aquisição de Conhecimento. Nesse trabalho é exposto o conceito de Sistema Especialista mostrando sua arquitetura, as dificuldades em adquirir o conhecimento para esse tipo de sistema e algumas técnicas de aquisição de conhecimento.

Palavras-chave: aquisição de conhecimento, sistema especialista, regras de produção.

KNOWLEDGE ACQUISITION: THE LARGE CHALLENGE ON CONCEPTION OF THE EXPERT SYSTEMS

ABSTRACT

During the last decades the world is turning a atmosphere totally dependent of the information. In this context appear the Expert Systems. One of the stages in the construction of these systems presents plenty difficulty, the knowledge acquisition. In this paper is exposed the concept of Expert Systems showing your architecture, the difficulties in acquiring the knowledge for that system type and some techniques of knowledge acquisition.

Key-words: knowledge acquisition, expert systems, production rules.

AQUISIÇÃO DE CONHECIMENTO: O GRANDE DESAFIO NA CONCEPÇÃO DE SISTEMAS ESPECIALISTAS

1. INTRODUÇÃO

O desenvolvimento mundial ocorre e vem se intensificando em virtude da crescente quantidade de informações compartilhadas entre os povos. Esse desenvolvimento acontece em conjunto com o desenvolvimento tecnológico, o qual abre maiores caminhos para a disseminação de mais informações ou conhecimento, por exemplo, a Internet é uma grande responsável pela divulgação de conhecimento ou informações. Nesse contexto, de mundo das informações, é cada vez mais crescente a necessidade que têm os profissionais e empresas de possuírem conhecimento sobre o seu domínio de atuação para que possam mostrar-se competitivos no mercado em que atuam. O diferencial competitivo de uma empresa situa-se essencialmente na capacidade que ela tem de oferecer soluções para seus clientes. Essa capacidade provém do conhecimento adquirido pelas empresas, ou melhor, por profissionais que trabalham nelas durante vários anos. Por não ser um produto concreto o conhecimento pode simplesmente sumir ou não estar presente quando for necessitado, estando ele armazenado na mente das pessoas que trabalham na empresa pode desaparecer com uma mudança de quadro de pessoal. Estando ele registrado em documentos e relatórios, poderá não ser percebido facilmente quando realmente for necessitado. Assim o uso de sistemas inteligentes vem se tornando popular e essencial em empresas de médio e grande porte. Esses sistemas inteligentes são os Sistemas Especialistas (SEs), sistemas que armazenam conhecimento e possibilitam o resgate dele facilmente e de maneira organizada, Sistemas de Apoio à Tomada de Decisões são exemplos de SEs. Os Sistemas de Apoio à Tomada de Decisões armazenam informações relevantes a um determinado assunto e auxiliam a tomada de decisões sobre aquele assunto, por exemplo, a compra e venda de ações na bolsa de valores que se dá em função da cotação do dólar, do índice de risco de investimento em determinado país, do preço do petróleo, entre outros, pode ser auxiliada por um Sistema de Apoio à Tomada de Decisões.

Outro fato que ocorre com a grande quantidade de informações existente é que não há mais um profissional que conheça tudo em sua área de trabalho, por exemplo, na medicina há várias áreas de especialização: um neurologista estuda e trabalha com problemas no cérebro humano, um cardiologista com problemas no coração, um ginecologista com o aparelho reprodutor feminino. Logo um especialista é aquele profissional que possui um conhecimento profundo e específico em determinado assunto. Mas alguns tipos de especialistas são raros, caros e algumas vezes podem falhar. Dessa maneira os SEs demonstram ser de grande valia na realização de tarefas que requerem especialização (SOLANGE, 2003, p.68). Esses sistemas podem, em algumas tarefas, não substituírem o especialista humano, como em um diagnóstico médico, mas auxiliá-lo, aumentando assim a eficácia no diagnóstico e tratamento de doenças.

Já é um fato o uso de SEs em diversas áreas de trabalho. No entanto, em detrimento do sucesso no uso dos SEs, a construção deles apresenta um grande desafio, a construção de sua Base de Conhecimento (BC), que para ser realizada necessita de uma abordagem de Engenharia de Conhecimento, a qual extrai informações de especialistas, documentos e fatos e os insere na BC que será usada pelo sistema. Essa tarefa é essencial e bastante difícil tendo em vista a natureza do problema (o conhecimento).

2. SISTEMAS ESPECIALISTAS

Sistemas Especialistas, também conhecidos como Sistemas Baseados em Conhecimento, são Sistemas Inteligentes (SI). Conforme Solange (2003, p.52), há muitas controvérsias sobre a definição de SI. No entanto, em sua maioria elas acabam por convergir para um mesmo ponto: SIs são sistemas de computador que têm a capacidade de realizar operações que se fossem realizadas por humanos seriam consideradas inteligentes. Os SEs diferem de outros SIs por possuírem a capacidade de gerar explicações sobre a linha de raciocínio por trás de cada uma de suas decisões.

Os SEs são programas de computador que manipulam conhecimento e informações de forma inteligente. Eles são desenvolvidos para resolver problemas que requerem uma grande quantidade de conhecimento humano e de especialização. Por isso todo SE possui uma Base de Conhecimento (BC) onde é armazenado conhecimento humano sobre uma determinada área na qual o sistema será usado. Um exemplo de SE seria o de um sistema de apoio à tomada de decisões empresariais: este sistema deveria armazenar informações relevantes ao bom desempenho empresarial de um determinado empreendimento. Assim os empresários poderiam consultá-lo sempre que tivessem a necessidade de tomar uma decisão.

3. ESTRUTURA DE UM SISTEMA ESPECIALISTA

Nem todos os SEs apresentam a mesma estrutura, no entanto, em sua grande maioria ela é semelhante a da figura 1.

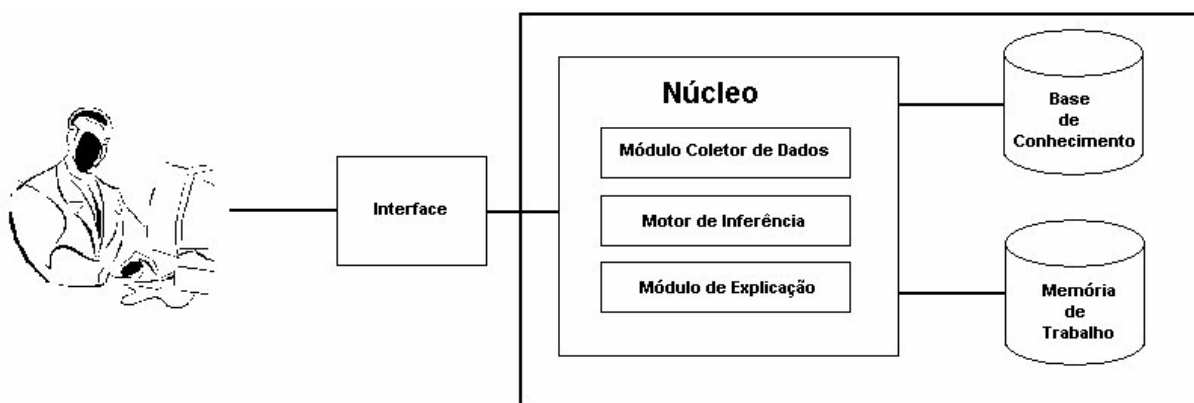


Figura 1. Estrutura básica dos sistemas especialistas.

3.1. BASE DE CONHECIMENTO (BC)

A BC é o local onde está armazenado o conhecimento sobre o domínio de atuação do SE. Esse conhecimento pode ser representado em várias linguagens de representação de conhecimento como: regras de produção, lógica ou redes semânticas.

Muitas vezes o conhecimento existente na BC não é totalmente suficiente para chegar a uma conclusão e fornecer uma resposta plausível, assim os SEs devem possuir mecanismos para lidar com esses problemas, por exemplo, solicitar ao usuário uma

determinada informação adicional para poder chegar a um resultado, ou fornecer uma resposta intermediária.

3.2. NÚCLEO DO SE

O núcleo do SE é composto basicamente por três módulos que podem ser independentes ou não.

- 1) Módulo Coletor de Dados (MCD):** é responsável pela interação com o usuário. Ele faz perguntas necessárias ao processo de raciocínio (inferência). Como perguntas sobre sintomas de doenças, na tentativa do diagnóstico correto;
- 2) Motor de Inferência (MI):** é responsável pelo desenvolvimento do raciocínio baseando-se nas informações obtidas pelo MCD e armazenadas na BC. É neste módulo que são processadas informações para se obter respostas aos usuários do SE;
- 3) Módulo de Explicação (ME):** tem a função de fornecer explicações, em alto nível, sobre o porquê de ter chegado à determinada conclusão ou o porquê de está fazendo uma determinada pergunta ao usuário.

3.3. MEMÓRIA DE TRABALHO (MT)

A Memória de Trabalho pode ser comparada com a memória RAM de um computador, pois a MT é uma área de armazenamento de informações temporárias. Informações que são necessárias apenas no momento em que um determinado trabalho está sendo realizado. Nela são registradas todas as respostas fornecidas pelos usuários em uma interação com o sistema, são armazenados, também, todos os passos executados em uma interação, as conclusões intermediárias e conclusões finais. A sua utilização traz algumas vantagens como: 1) evitar a repetição de uma mesma pergunta ao usuário; 2) possibilita fornecer ao usuário toda a linha de raciocínio responsável pelas conclusões obtidas; 3) evita a realização de buscas de informações repetidas na BC para a obtenção de conclusões intermediárias.

3.4. INTERFACE

A interface é o meio de interação entre os usuários e o SE. Ela permite a comunicação entre ambas as partes. É por meio dela que o sistema faz perguntas aos usuários e recebe as respectivas respostas. Também é através dela que o sistema expõe as conclusões e explicações. As interfaces devem apresentar linguagens bastante abstratas, próximas da linguagem natural. Para isso são utilizados recursos como texto, vídeo, gráfico, imagem, áudio ou animação.

4. REPRESENTAÇÃO DO CONHECIMENTO NA BASE DE CONHECIMENTO

Segundo Solange (2003, p.29), representar um conhecimento pode ser entendido como uma forma sistemática de estruturar e codificar o que se sabe sobre determinado

assunto. No entanto, em detrimento de uma codificação qualquer uma representação de conhecimento deve apresentar as seguintes características:

- 1) Ser compreensível ao ser humano, pois, como já sabemos, os SEs devem permitir explicações sobre seus raciocínios, essas explicações provêm do conhecimento armazenado na BC, portanto, esse conhecimento deve ser interpretável facilmente.
- 2) Ser robusta, ou seja, permitir a sua utilização mesmo que não aborde todas as situações possíveis de um determinado assunto.
- 3) Ser generalizável. Assim deve possuir conhecimento sobre vários pontos de vista de determinado assunto. Assim será possível ser atribuída a diversas situações e interpretações diferentes.

Há várias técnicas de representação de conhecimento como: representação Lógica, Redes Semânticas, Frames, Orientação a Objetos e Regras de Produção.

4.1. REGRAS DE PRODUÇÃO

Por serem de fácil entendimento e não necessitarem de conhecimento específico veremos, com mais detalhes, as Regras de Produção.

Os primeiros SEs foram sistemas baseados em regras. Sistemas assim trabalham com a idéia de que seres humanos tomam decisões baseando-se em regras do tipo *SE* <condição> *ENTÃO* <conclusões e ações> (NEGHEVITSKY, 2002, p.30).

Uma simples regra de produção seria: *SE* está chovendo *ENTÃO* não vou sair hoje. As condições podem encontra-se armazenadas na BC ou ser fornecidas pelos usuários do sistema como respostas às perguntas do próprio SE. Algumas regras podem produzir conclusões intermediárias as quais serão usadas como condições de outras regras para se chegar a uma conclusão final. Vejamos o exemplo seguinte:

SE <tenho dinheiro> *ENTÃO* <posso ir à festa>
SE <posso ir à festa> *ENTÃO* <convidarei minha namorada>

Como podemos observar, no exemplo acima, as regras são uma forma natural de representação de conhecimento.

5. AQUISIÇÃO DE CONHECIMENTO (AC)

Segundo Mastela (2004), AC pode ser definida como o processo de compreender e organizar o conhecimento de várias fontes. Esse conhecimento deverá ser codificado e armazenado em uma BC para posterior resgate por um SE.

O agente responsável por todas as atividades de construção de um SE é o Engenheiro de Conhecimento. Ele deve entender o domínio de aplicação do sistema, interagir com as fontes de informações (conhecimento que será incorporado ao SE) e,

ainda, possuir um conhecimento razoável de computação, linguagens e ferramentas de Inteligência Artificial a fim de escolher o melhor para implementação do sistema.

A AC é a atividade inicial do processo de Engenharia de Conhecimento e a mais difícil (MASTELA, 2004), por isso diz-se que ela é o “gargalo” da construção dos SEs. Essa dificuldade se dá em virtude da inexistência de uma metodologia eficiente, confiável e padrão para extração e organização do conhecimento das várias fontes. A dificuldade existe de fato, no entanto a Aquisição de Conhecimento é inerente à construção de uma Base de Conhecimento, sendo essa última uma componente básica de um SE é pertinente dizer que a AC é um processo pelo qual todo Sistema Especialista deverá enfrentar quando de sua concepção.

As fontes de conhecimento usadas no processo de AC podem ser documentadas ou não. As fontes documentadas são: livros, filmes, gráficos, diagramas, manuais, entre outras. Já as fontes não documentadas residem na mente das pessoas que atuam no domínio de aplicação do sistema (Especialistas que detêm o conhecimento que se quer armazenar no SE).

Ao utilizar fontes documentadas o Engenheiro de Conhecimento esbarra na dificuldade relativa a ter que conhecer o domínio do problema para poder compreender o conteúdo da documentação, por exemplo, não é possível para um leigo, adquirir um livro de cardiologia e conseguir extrair conhecimentos importantes. Assim o Engenheiro de Conhecimento acaba tendo que recorrer aos agentes do domínio, os quais são pessoas que trabalham diretamente com o assunto em questão. Mas em quase todos os domínios há um agente que se destaca por desenvolver métodos particulares e eficientes de lidar com os problemas e ainda por possuir um conhecimento profundo sobre o assunto. Esse agente é o Especialista. Especialistas desenvolvem estruturas de armazenamento de informações particulares que são mais enxutas e eficazes para lidar com informações de um domínio específico. Dessa maneira um Especialista é a melhor fonte de informações para um processo de AC.

Há muitos fatores que causam problemas na AC, quando se tenta extrair conhecimento de Especialistas. Um problema inicial é a comunicação entre o Especialista e o Engenheiro de Conhecimento. O Engenheiro pode começar uma série de entrevistas sem ter boa noção do assunto em questão. Assim o Especialista para ser compreendido deverá usar termos triviais, mas esses não são termos que expressam a maneira natural de trabalho do Especialista, logo há uma considerável perda de informação na sua transmissão. O ideal é que ambos, o Engenheiro de Conhecimento e o Especialista, aprendam um pouco sobre o trabalho do outro para que a conversa flua mais naturalmente, no entanto, até que isso ocorra já se terá perdido muito tempo.

Mesmo havendo uma boa interação entre Especialista e Engenheiro de Conhecimento ainda há alguns problemas no processo de AC.

Algumas estratégias de resolução de problemas usadas por Especialistas estão tão arraigadas na mente desses profissionais que só são ativadas em situações reais de aplicação de um trabalho, por isso elas quase sempre não são transmitidas para os Engenheiros de Conhecimento nas entrevistas.

Outra dificuldade é que em alguns casos o Especialista sabe resolver um problema, mas não sabe explicar como ou porque resolveu daquela maneira. Isso ocorre porque o conhecimento utilizado para resolver o problema é tácito, ou seja, é um conhecimento implícito em sua mente que para ser relatado deve ser explicitado. No entanto, explicitar um conhecimento que está arraigado em uma mente é uma tarefa que oferece bastante dificuldade para os seres humanos em geral, pois, muitas vezes, as pessoas possuem conhecimentos que nem mesmo sabem que o têm.

Há ainda o fato de que algumas pessoas sentem-se avaliadas no processo de entrevistas e acabam omitindo algumas informações importantes para a construção da BC.

Um outro problema é que Especialistas são profissionais extremamente valiosos e requisitados em uma empresa, assim é bastante difícil tê-los comprometidos plenamente com o desenvolvimento de um SE.

6. TÉCNICAS DE AQUISIÇÃO DE CONHECIMENTO

Muitos esforços têm sido feitos para sistematizar ou até mesmo automatizar o processo de AC. Estes esforços resultaram em várias técnicas. Essas técnicas podem ser classificadas em manuais, semi-automáticas e automáticas. As técnicas manuais são as mais utilizadas, as semi-automáticas geralmente são utilizadas em conjunto com as técnicas manuais, já as técnicas automáticas dizem respeito ao processo pelo qual o conhecimento é adquirido automaticamente, ou seja, sem a interferência humana, ou com pouca interferência. As técnicas automáticas são uma área tão vasta que não entram no escopo desse trabalho. Como exemplo de AC automático temos a mineração de dados e o aprendizado de máquina (redes neurais, árvores de decisões, entre outros).

6.1. TÉCNICAS MANUAIS

Segundo Solange (2002, p. 69), a maioria das técnicas manuais fundamentam-se na Psicologia e Análise de Sistemas. Nessas técnicas o Engenheiro de Conhecimento é responsável por adquirir o conhecimento do Especialista e outras fontes e codificá-lo na BC.

6.1.1. BASEADAS EM DESCRIÇÕES OU IMERSÃO NA LITERATURA

Nessa técnica o Engenheiro de Conhecimento faz um pequeno estudo sobre o assunto do problema em algumas literaturas com o objetivo de adquirir algum conhecimento inicial sobre o domínio. Isso é importante para que nas futuras entrevistas com o Especialista a conversa possa fluir a um nível um pouco mais adiantado, pois o simples uso de alguns jargões pelo poder vir a impossibilitar o Engenheiro de Conhecimento de adquirir determinadas informações.

6.1.2. BASEADAS EM ENTREVISTAS

Nessa técnica são realizadas entrevistas com o Especialista. As informações podem ser coletadas com o auxílio de gravadores filmadoras questionários, ou qualquer outro meio possível. Essas informações são posteriormente analisadas para se extrair o conhecimento desejado. Há dois tipos de entrevistas: as não estruturadas e as estruturadas.

- **Entrevistas Não Estruturadas**

O objetivo dessas entrevistas não é a adquirir conhecimento sobre um tópico específico, mas sim obter uma visão geral do domínio em questão. Assim as entrevistas são mais livres e propiciam um bom relacionamento entre o Engenheiro de Conhecimento e o Especialista.

Por permitir uma conversa mais livre este tipo de entrevista pode se estender a tópicos irrelevantes para o desenvolvimento do sistema. Dessa maneira ela não é recomendada para fases além das iniciais, de identificação do conhecimento e ambientação do Engenheiro com o domínio.

- **Entrevistas Estruturadas**

Essas entrevistas são mais formais. As perguntas são elaboradas com cuidado e tendem a levar a conversa à aquisição de conhecimentos específicos sobre o domínio. Segundo (MASTELA, 2004), ela força o Especialista a ser sistemático quando participa da entrevista. Geralmente as respostas são gravadas ou anotadas para posterior análise. O material coletado nas entrevistas pode ser transformado em descrições que facilitam o processo de análise. Em alguns casos essas descrições podem gerar um glossário de definições de termos usados no domínio de trabalho do Especialista.

Um problema encontrado nesse tipo de entrevista é que ela pode produzir resultados influenciados pelo entrevistador, por isso ela deve ser planejada e revisada antes de ser aplicada. O entrevistador deve ser indiferente.

6.1.3. BASEADAS EM ACOMPANHAMENTO

Esse método consiste em acompanhar o processo de raciocínio do Especialista em casos reais. Assim são pesquisados casos em prontuários ou antigos projetos desenvolvidos anteriormente por esse profissional. Isso estimula sua descrição sobre o assunto em questão.

Essa técnica por utilizar casos reais evita que o Especialista seja direcionado a responder questões irrelevantes, no entanto, nem sempre se consegue uma amostragem de casos realmente representativa.

6.2. TÉCNICAS SEMI-AUTOMÁTICAS

As técnicas de AC manuais podem ser bastantes problemáticas devido ao grande número de pessoas envolvidas, Especialistas, Engenheiros e Programadores, (SOLANGE, 2002, p.78). Esse processo de adquirir o conhecimento do Especialista e transmiti-lo aos

programadores acaba por gerar ruídos de comunicação entre as partes. Uma alternativa para minimizar esse tipo de problema é a Aquisição de Conhecimento semi-automática, que consiste na utilização de ferramentas computacionais que auxiliam ao Engenheiro de Conhecimento a codificação da BC. Além de reduzir o contingente de pessoas envolvidas e por consequência os problemas de comunicação, a aquisição semi-automática também acelera o processo de construção da Base de Conhecimento permitindo que o Engenheiro e o Especialista tenham respostas mais rápidas, pois à medida que a Base vai sendo construída ela pode ser testada. Assim possíveis erros serão detectados prematuramente.

7. CONCLUSÃO

A sabedoria vem com o uso bem sucedido do conhecimento, entendendo conhecimento como sendo o conjunto de informações necessárias a realização de determinadas tarefas é irrefutável que a sabedoria é o bem mais precioso da humanidade, pois somente com o saber é possível sobreviver e se desenvolver. A preciosidade da sabedoria também provém do conceito de que ter saber é ter poder, ou seja, poder para realizar atos que só são possíveis quando se sabe.

O conhecimento é inerente a qualquer área de atividade em nossa sociedade, assim existem conhecimentos dos mais diversos possíveis. As informações que compõem um determinado conhecimento podem está localizadas em diversos meios como: livros, anotações, registros em uma empresa, banco de dados, mente de pessoas, entre outros. Mas estas informações nem sempre podem ser facilmente resgatadas ou manipuladas. Informações em livros, por exemplo, são de difícil acesso quando não se sabe realmente o que se quer ou quando não se tem o livro certo. Conhecimento na mente de pessoas também é outro problema, já que essas pessoas nem sempre estão disponíveis.

Pessoas com um grande conhecimento específico em determinado assunto são denominados de Especialistas. Esses profissionais são de extrema importância em certas áreas de trabalho e são bastante requisitados. Assim é de interesse que o conhecimento da mente desses profissionais possa ser perene e facilmente acessível. Em virtude da necessidade do conhecimento uma das áreas de estudos da informática, a Inteligência Artificial, vem despendendo esforços para a construção de sistemas informatizados que facilitem o armazenamento e resgate de conhecimento de maneira inteligível ao ser humano. Esses sistemas são os Sistemas Baseados em Conhecimento ou Sistemas Especialistas. Esse tipo de sistema vem se mostrando bastante eficiente no apoio a tomada de decisões em diversas áreas de aplicação como, medicina, psicologia, investimentos financeiros, administração empresarial e outras.

Os Sistemas Especialistas surgem então como uma solução para suprir as necessidades em um mundo que necessita de informações cada vez mais precisas e rápidas. No entanto, o processo de Aquisição de Conhecimento é um dos maiores problemas na concepção desse tipo de sistema. O sucesso de um sistema dessa natureza depende da fidelidade das informações coletadas sobre o domínio de interesse.

Há várias técnicas de AC: as técnicas manuais que dependem muito do Engenheiro de Conhecimento, as técnicas semi-automáticas que fornecem ferramentas que ajudam ao Engenheiro otimizar e acelerar o processo de AC; e, as técnicas automáticas (que não foram objetivos desse trabalho) que dispensam completamente a necessidade de um

Engenheiro de Conhecimento. As técnicas automáticas são uma enorme área de estudo na Inteligência Artificial, o aprendizado de máquina, elas permitem construir programas de computador que adquirem conhecimento a partir de exemplos. Elas são eficientes no que diz respeito ao reconhecimento de padrões (imagens, voz, seqüências), mas não podem adquirir conhecimento complexo de uma mente humana (um Especialista), elas ainda não podem explicar as linhas de raciocínio usadas para resolver determinado problema, portanto são limitadas em alguns aspectos importantes.

8. REFERÊNCIAS

MASTELA, L. S. Técnicas de Aquisição de Conhecimento para Sistemas Baseados em Conhecimento. Disponível em:

<http://www.inf.ufrgs.br/gpesquisa/bdi/publicacoes/files/TI1LSM.pdf>. Acesso em 30/6/2004.

NEGHEVITSKY, M. **Artificial Intelligence: a Guide to Intelligent Systems**. Adilson Wsley Editora, 2002.

RUSSELL, S; NORVIG, P. **Inteligência Artificial**. Campus, 2004.

SOLANGE, O et.al. **Sistemas Inteligentes: fundamentos e Aplicações**. Manole, 2003.