

UNIVERSIDADE FEDERAL DE SANTA CATARINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

PROPOSTA DE UMA METODOLOGIA DE  
AVALIAÇÃO PARA SISTEMAS ESPECIALISTAS NA  
ÁREA MÉDICA

DISSERTAÇÃO SUBMETIDA À UNIVERSIDADE FEDERAL DE  
SANTA CATARINA PARA A OBTENÇÃO DO GRAU DE MESTRE EM  
ENGENHARIA ELÉTRICA

GISELE FAFPE PELLEGRINI

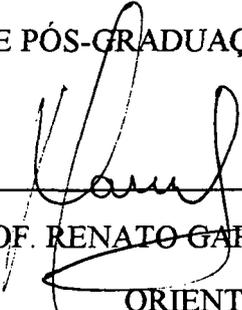
FLORIANÓPOLIS, 29 DE SETEMBRO DE 1995

PROPOSTA DE METODOLOGIA PARA AVALIAÇÃO DE SISTEMAS ESPECIALISTAS  
PARA A ÁREA MÉDICA

GISELE FAFTE PELLEGRINI

ESTA DISSERTAÇÃO FOI JULGADA ADEQUADA PARA OBTENÇÃO DO TÍTULO DE  
MESTRE EM ENGENHARIA

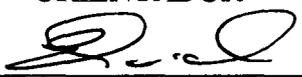
ESPECIALIDADE ENGENHARIA ELÉTRICA, ÁREA DE CONCENTRAÇÃO:  
ENGENHARIA BIOMÉDICA, E APROVADA EM SUA FORMA FINAL PELO  
PROGRAMA DE PÓS-GRADUAÇÃO



---

PROF. RENATO GARCIA OJEDA, D. Sc.

ORIENTADOR

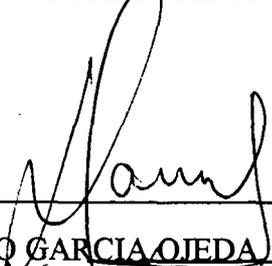


---

PROF. ENIO VALMOR KASSICK, D. Sc.

COORDENADOR DO CURSO

BANCA EXAMINADORA:



---

PROF. RENATO GARCIA OJEDA, D. Sc.



---

PROF. WALTER CELSO DE LIMA, D. Sc., L.D.



---

PROF. JORGE MUNIZ BARRETO, D. Sc.



---

PROF. FERNANDO MENDES DE AZEVEDO, D. Sc.

aos meus pais e minhas irmãs.

## Artigo

**"Avaliação de Sistemas Especialistas para área médica", Gisele Faffe Pellegrini, Renato Garcia Ojeda, Anais do XI Congreso Chileno de Ingenieria Eléctrica, Punta Arenas, Chile, p. H13-H17, 1995.**

### AGRADECIMENTOS:

Ao Prof. Renato Garcia Ojeda pela orientação, persistência e amizade.

Ao Grupo de Pesquisas em Engenharia Biomédica (GPEB), representado pelo Prof. Carlos I. Zanchin pelo apoio operacional pelo qual foi possível o desenvolvimento deste trabalho.

À CAPES pelo apoio financeiro.

Ao Dr. Li Shih Min pelas constantes sugestões e incentivo.

A todos os amigos que fiz durante o período do curso.

A todos os amigos do GPEB, em especial, Sílvia e Fernando que foram mais que amigos.

# Sumário

Capítulo 1 - Introdução .....	10
1.1 - Justificativa .....	15
Capítulo 2 - Fundamentação Teórica .....	18
2.1 - Arquitetura de Sistemas Baseados em Conhecimento .....	18
2.1.1 - Abordagem Simbólica .....	21
2.1.2 - Sistemas Conexionistas .....	25
2.1.3 - Sistemas Baseados em Conhecimento e a Aplicação na Medicina .....	28
2.2 - Tomada de Decisão Médica .....	30
2.3 - Aquisição de Conhecimento .....	34
Capítulo 3 - Modelos de Avaliação dos Sistemas Especialistas.....	39
3.1 - Proposta de Gaschnig.....	40
3.2 - Proposta de Lundsgaarde .....	45
3.3 - Proposta de Hart-Wyatt .....	46
3.4 - Proposta de Rossi-Mori .....	50
3.5 - Proposta de Nykänem .....	51
3.6 - Proposta de Clarke .....	54
3.7 - Proposta para Sistemas Conexionistas.....	57
Capítulo 4 - Proposta de uma Metodologia para Avaliação de SE na Área Médica .....	60
4.1 - Justificativa da Proposta.....	60
4.2 - Detalhamento da Proposta .....	63
4.3 - Verificação da Proposta num SE Desenvolvido pelo GPEB.....	69
Capítulo 5 - Discussão e Conclusão.....	75
Referências Bibliográficas.....	80

## Resumo

A avaliação dos Sistemas Especialistas (SE) para apoio a tomada de decisão na área médica vem se tornando necessária para que os SE sejam mais aceitos. Existem várias propostas para formalizar o processo de avaliação. Neste trabalho, propõe-se uma metodologia para avaliação de SE utilizando a metodologia de desenvolvimento por protótipo. Esta metodologia consiste em um sistema que é continuamente re-implementado. Através de um protótipo operacional, os constantes retornos obtidos pelos usuários irão garantir um SE de melhor qualidade e maior aceitabilidade, mais voltado para as reais necessidades dos usuários.

## Abstract

The evaluation of clinical decision support systems has long been recognized as an important part of the overall field of medical computing. There are a number of purposes for which such an evaluation might be performed. We propose the evaluation of medical expert system focuses a system development by prototyping. This technique consist of a system which is usually re-implement. Through an operating prototype the user feedback is continually obtained and this will result a better medical expert system, closely to the real user needs.

## Glossário

**GPEB** - Grupo de Pesquisa em Engenharia Biomédica

**HU** - Hospital Universitário

**IA**- Inteligência Artificial

**KADS** - Knowledge Aquisition Development Systems

**RN** - Redes Neurais

**SE** - Sistemas Especialista

**UFSC** - Universidade Federal de Santa Catarina

# Capítulo 1- Introdução

O Grupo de Pesquisas em Engenharia Biomédica (GPEB), da Universidade Federal de Santa Catarina, criado oficialmente em 1974, desenvolve pesquisas que se concentram principalmente nas áreas de informática médica e instrumentação biomédica (LIMA, 1993).

Por informática médica se entende a combinação da ciência médica com as várias tecnologias e disciplinas das ciências de informação e de computação, que resultam em novas metodologias que podem contribuir para o melhor uso dos conhecimentos médicos (SYMPOSIUM MEDICAL INFORMATICS, 1985). Compreende sistemas de informações que auxiliam na administração hospitalar, sistemas de cálculos laboratoriais e sistemas de apoio à tomada de decisão, entre outros.

Na área de informática médica, uma das linhas de pesquisa do GPEB é a inteligência artificial, onde o grupo vem desenvolvendo aplicativos como sistemas especialistas (SE) para diagnóstico de icterícia (NIEVOLA, 1988), para o ensino na área de treinamento de infecção hospitalar (RAMIREZ, 1991), um SE híbrido em anestesiologia (GARCIA, 1992) que foi ampliado para casos dos pacientes críticos e problemáticos do ponto de vista anestesiológico (PASSOLD, 1995), um SE estatístico para apoio a pesquisas médicas (NASSAR, 1994), um SE para aquisição de conhecimento aplicado ao diagnóstico de epilepsia (BRASIL, 1994) dentre outros.

Nos SE para apoio a tomada de decisão no diagnóstico médico vem se observando que apesar de um grande empenho durante seu desenvolvimento, poucos são realmente utilizados, e em alguns casos, quando o são, é por uma determinação superior, e não por sua ajuda efetiva (SHORTLIFE, 1991).

Todo processo de inovação científica passa por 3 fases. Primeiro um grande entusiasmo, seguido de um período de dúvidas e finalmente o último estágio onde a inovação se estabelece já como uma rotina ( de DOMBAL, 1987). Na informática médica, em especial nos sistemas de apoio à tomada de decisão para o diagnóstico médico, ocorreu o mesmo. Houve um grande entusiasmo na década de 70 quando os primeiros sistemas especialistas (SE) para apoio a tomada de decisão foram desenvolvidos mas ainda não se pode afirmar que estes sistemas já estão na terceira fase, ou seja, estão incluídos na rotina das clínicas médicas (HEATHFIELD, 1993).

Para justificar esta fase de dúvidas que impede o aperfeiçoamento da informática médica para os SE para apoio ao diagnóstico médico, de DOMBAL (1987) analisa algumas afirmações que representam obstáculos na aceitação dos SE, destacando:

1- Sistemas de apoio à decisão são desnecessários nas clínicas médicas.

Este é um argumento não mais aceito, pois a maioria dos clínicos atuais aceitam que a introdução da informática nas clínicas médicas tem apresentado bons resultados diminuindo o tempo gasto nas situações que envolvam tomadas de decisão e facilitando a investigação e organização de dados de pacientes (CLIFFORD, 1986, SHORTLIFFE, 1987).

2- Não se pode aceitar a decisão tomada pelo SE.

Também é um argumento não mais aceitável atualmente. Como existem muitas ferramentas que facilitam o desenvolvimento de SE, existem muitos sistemas que se dizem capazes de apoiar a tomada de decisão. O problema do usuário é saber distinguir quais desses sistemas foram cuidadosamente **testados e avaliados** e baseado em qual metodologia.

3- A ajuda do SE é confusa.

Como a área médica não é determinística, a experiência adquirida na prática é um fator de peso na tomada de decisões. Alguns médicos afirmam que os SE apresentam muitos caminhos alternativos gerando incerteza nos diagnósticos, o que torna o apoio à tomada

de decisão ainda mais confusa. Uma cuidadosa **verificação** deve ser feita para evitar esta característica.

4- A solução apresentada pelo SE não pode ser aplicada.

Este argumento não é justificado pelos fatos, pois os SE para dor abdominal, dor torácica e infecções gastro-intestinais (HECKERLING, 1991) já implementados têm apresentado resultados coerentes com o conhecimento de um especialista da área, facilitando o trabalho em centros médicos.

5- Médicos não aceitam as decisões tomadas por um SE.

Existe uma certa resistência em relação ao uso rotineiro dos SE, porque ainda não existe uma forma padronizada para liberar o uso desses sistemas. O problema recai mais uma vez na necessidade de **avaliação** dos SE, de forma que possibilite o médico discernir em qual conhecimento se baseia o SE para inferir suas conclusões. Mas não se pode afirmar que de um modo geral as decisões tomadas pelo SE não são aceitas, apenas se discute sobre a **validade** dessas soluções apresentadas.

6- Existem problemas relacionados com aspectos legais.

Se o SE apresenta uma solução inadequada e o médico a utiliza, quem deverá ser responsabilizado pelas consequências desta decisão? E quem deve responder legalmente se o SE apresentar uma solução que é considerada a mais usual, mas uma vez aplicada percebe-se que faltou informar com mais detalhes a situação atual do paciente, o que tornaria o procedimento apresentado inadequado? Estas questões sobre negligência, responsabilidade, ainda geram muitas dúvidas na classe médica ( BAINBRIDGE, 1991). Numa tentativa de evitar ao máximo estas situações, os responsáveis pelo desenvolvimento e construção de SE ressaltam a necessidade de uma metodologia que **avalie** consistentemente os sistemas.

7- O uso dos SE é prejudicado pela falta de incentivo.

Apesar da informática apresentar muitas vantagens no seu uso diário, alguns centros, devido a problemas de distância, têm dificuldade de atendimento pelos técnicos ou pessoal de suporte e manutenção do sistema para fazer ajustes, alterações necessárias para o centro específico onde o sistema foi implementado. A falta de manuais ou explicações de fácil entendimento também retardam a adaptação de um centro médico aos recursos oferecidos pelos sistemas computacionais (WYATT, 1989a).

Dentre os argumentos apresentados, destaca-se a característica **avaliação** como a causa mais frequente para o não uso rotineiro dos SE.

Feito um estudo do estado da arte em metodologia para avaliação de SE na área médica, constatou-se que vários grupos de desenvolvimento de sistemas especialistas para a área médica enfrentam este problema. Estima-se que 90% de todos os sistemas especialistas em medicina não foram totalmente avaliados (WYATT, 1991a).

Entretanto antes de discutir avaliação de sistemas computacionais (*software*), a própria noção de avaliação tem que ser definida e tornada mais precisa. Avaliar é dar valor com base em certos critérios que devem ser objetivos, e não puramente pessoais (WYATT, 1990).

É possível avaliar software do ponto de vista de quem desenvolveu e do ponto de vista do usuário. No primeiro caso, as características do software que são avaliadas se referem às qualidades intrínsecas, como estrutura modular, confiabilidade, desempenho, robustez e portabilidade. Na avaliação feita pelo usuário, as características observadas são aquelas relacionadas a aplicabilidade, utilidade, abrangência e facilidade de uso.

Alguns autores atribuem a falta de um "padrão ouro" (*Gold Standard*) para avaliação de SE o maior obstáculo para a determinação de uma metodologia aceita por toda a comunidade médico-científica (MILLER, 1991, NYKÄNEN, 1991, HART, 1990, SHORTLIFFE, 1987).

Atualmente, existem diversas discussões publicadas sobre a importância da avaliação mas nenhuma metodologia pode ser considerada a melhor para os sistemas especialistas na área médica, e isto vem se tornando um obstáculo para o uso mais constante dos sistemas especialistas (WHITEBECK, 1983). A importância da avaliação está na credibilidade que o sistema alcança com o usuário, na descrição das conclusões apresentadas e os critérios para decidir sobre uma tomada de ação. É importante que o usuário, geralmente um médico ou enfermeiro, saiba como o sistema aplicou o raciocínio clínico, para que possa considerar as consequências se estiver tomando uma decisão errada (WHITEBECK, 1983).

LUNDSGAARDE (1987), faz uma distinção bem clara entre os grupos envolvidos no desenvolvimento de um sistema especialista para a área médica. Existe o grupo de desenvolvimento subdividido entre os técnicos em computação e engenharia e os médicos que dominam o conhecimento especialista, e o grupo de usuários que utiliza o sistema.

É importante salientar a estreita relação que deve existir entre os técnicos de computação e/ou engenharia com os médicos especialistas, pois os técnicos devem ser capazes de representar com exatidão, através de recursos computacionais, o conhecimento e o raciocínio médico.

A avaliação desses sistemas deve ser feita primordialmente em relação ao ponto de vista do usuário, pois é a partir da interação do sistema que o usuário obtém suas conclusões. O usuário desses sistemas não é necessariamente um médico especialista no assunto do sistema, pode ser uma enfermeira ou um estudante de medicina.

A definição dos limites de atuação desses dois grupos também representa um problema quando se faz uma avaliação, pois o grupo que desenvolve não pode especificar quando se

alcança um ponto final definitivo e o grupo do usuário exige saber até que ponto pode confiar no sistema implementado (LUNDSGAARDE, 1987).

## 1.1 - Justificativa

A ausência de uma metodologia padronizada tornou os critérios de avaliação muito subjetivos. Para ilustrar destaca-se cinco sistemas que sofreram as primeiras avaliações por grupos de pesquisadores diferentes com critérios diferentes. Esses sistemas são o MYCIN, INTERNIST, PUFF, HELP e ILIAD.

O MYCIN (SHORTLIFFE, 1976), foi desenvolvido no início da década de 70 na Universidade de Stanford. Foi projetado para investigar e recomendar terapias apropriadas para pacientes com infecções bacterianas (LUNDSGAARDE, 1987). O trabalho de avaliação foi feito na Faculdade de Matemática da Open University em Standford, baseado na interface homem-máquina e deixou de considerar a performance do sistema (atuação). Para WHITEBECK (1983) a avaliação do MYCIN não leva em conta a restrição feita para a seleção do grupo de pacientes que deveriam ser testados.

INTERNIST (MILLER, 1986) é um programa de diagnóstico computadorizado projetado para o uso na medicina interna mas ainda não foi implementado e testado num ambiente clínico (LUNDSGAARDE, 1987). Sua avaliação foi feita para mostrar os pontos fortes e fracos, com a intenção de obter um retorno dos prováveis usuários para outros desenvolvimentos de projetos. Esta avaliação não se preocupou com a performance do sistema.

O PUFF (AIKINS, 1983) é um programa que interpreta funções pulmonares a partir de dados colhidos em teste. O PUFF passou a ser operacional em 1979 no Pacific Medical Center. Em 1983 já tinha interpretado cerca de 400 casos vistos no laboratório de função pulmonar, mas

LUNDSGAARDE (1987) ressalta que é um sistema bem limitado no campo da atuação se for comparado com outros sistemas.

O HELP (WARNER, 1979) , Health Evaluation Through Logical Processing, desenvolvido no Hospital de Salt Lake City, passa por todas as etapas proposta por GASCHNIG (1983) sendo inclusive comercializado. O HELP, sistema para solucionar problemas relativos aos diagnósticos e prognósticos de doenças congênitas do coração, a partir de dados armazenados, foi desenvolvido como um sistema modular e cada módulo foi conceituado como uma entidade clínica.

O ILIAD (HECKERLING, 1991), é um sistema especialista para consulta médica que auxilia os diagnósticos para medicina interna. Foi desenvolvido pelo departamento de Informática Médica da Universidade de Utah, a partir do banco de dados formado pelos dados dos pacientes admitidos no Hospital da Universidade de Utah e dos procedimentos adotados pelos médicos especialistas. A avaliação dos seus resultados foi feita através das comparações dos seus diagnósticos com os de médicos intensivistas, através de análises estatísticas.

Sendo um grupo de pesquisa, o GPEB tem interesse que seus aplicativos para apoio à tomada de decisão médica no diagnóstico ou terapia, sejam aceitos e utilizados pelos usuários. Assim, este trabalho tem por objetivo mostrar as propostas de avaliação de sistemas existentes para a área médica e propor uma metodologia de avaliação para os projetos desenvolvidos no GPEB, levando em consideração suas características específicas, como localização, perfil dos componentes do grupo.

No Capítulo 2 é feito um levantamento sobre os itens mais importantes que envolvem a construção de um SE para a área médica, como uma abordagem sobre a arquitetura de sistemas simbólicos e conexionistas, uma visão do processo de tomada de decisão médica e da aquisição de conhecimento de um especialista para o desenvolvimento de um SE.

No Capítulo 3 é apresentado um estudo do estado da arte das metodologias de avaliação que são utilizadas no desenvolvimento de um SE, analisando as principais características para ser apresentada uma proposta de metodologia para o GPEB, sendo este o conteúdo do Capítulo 4.

Finalmente, no Capítulo 5 são levantadas as discussões e conclusões para avaliação de sistemas especialistas para a área médica.

## Capítulo 2 - Fundamentação Teórica

Sistemas de apoio à decisão médica receberam atenção significativa nos encontros de Sistemas de Engenharia e Medicina em 1989, em Stanford, e em Viena no Medical Informatics Europe em 1991, onde começaram a ser levantadas as questões sobre a falta de padronização para a aceitação dos sistemas para área médica, resultando num baixo índice de utilização.

No domínio médico, bases de julgamento imprecisas ou informações incompletas são comuns e conseqüentemente o problema da responsabilidade legal na medicina emergiu como um estímulo para a discussão sobre a metodologia de avaliação.

Grande parte dos sistemas médicos são aplicativos da área de IA, baseados em conhecimentos, formado pelos Sistemas Especialistas e pelos Sistemas Conexionistas ( Redes Neurais). Avaliar estes sistemas implica em avaliar também o especialista que participou da construção da base de conhecimentos, tornando mais abrangente o problema da avaliação (LAWRENCE, 1992). Neste capítulo são apresentados alguns conceitos básicos sobre sistemas especialistas, processo de tomada de decisão médica e a aquisição de conhecimento.

### 2.1-Arquitetura de Sistemas Baseados em Conhecimento

Inteligência Artificial (IA) é a parte da ciência da computação que cria estruturas computacionais inteligentes, isto é, sistemas que exibem características que podem ser associadas com a inteligência humana, como a linguagem natural, aprendizado, raciocínio e resolução de problemas. Os sistemas decisórios, robótica, desenvolvimento de linguagem natural, percepções visuais, prova de teoremas, banco de dados inteligente e diagnóstico médico são algumas das áreas de aplicação da Inteligência Artificial (RICH, 1993).

Uma técnica em IA é um método de explorar os conhecimentos que devem ser representados de tal modo que possam (RICH, 1993):

- 1- capturar generalizações, ou seja, agrupar situações que partilhem propriedades importantes;
- 2- ser de fácil compreensão para os usuários que farão a entrada dos dados para a aprendizagem. Em alguns programas de IA os dados são obtidos automaticamente (leitura de vários instrumentos), em outros domínios, a maioria dos conhecimentos que o programa possui deve ser fornecida através da interação com o usuário;
- 3- ser facilmente modificado para corrigir erros e para refletir mudanças (flexibilidade);
- 4- ser utilizado em muitas situações, mesmo quando não totalmente preciso ou completo;

Além de projetar sistemas que realizam tarefas seguindo uma lógica que qualquer pessoa poderia executar, há um grande número de tarefas que exigem muito conhecimento especializado. Essas tarefas só podem ser realizadas por especialistas que combinam conhecimento e experiência. Sistemas que realizam este tipo de tarefa são chamados de sistemas especialistas. Estes sistemas são desenvolvidos para analisar problemas e apoiar o usuário na tomada de decisões em questões complexas como problemas de matemática simbólica como o MATHEMATICS (WOLFRAN, 1988) ou análise química como o DENDRAL (LINDSAY, 1980).

Historicamente pode-se dizer que a IA reuniu os esforços de pesquisas associando os conceitos da lógica formal, linguagens de processamento e a psicologia cognitiva. Por outro lado, possibilitou um crescimento da computação interativa criando ambientes de programação interativos. Pode-se subdividir ainda o estudo da IA em três áreas de pesquisa relativamente independentes, como mostra a figura 1 (HARMON, 1988).

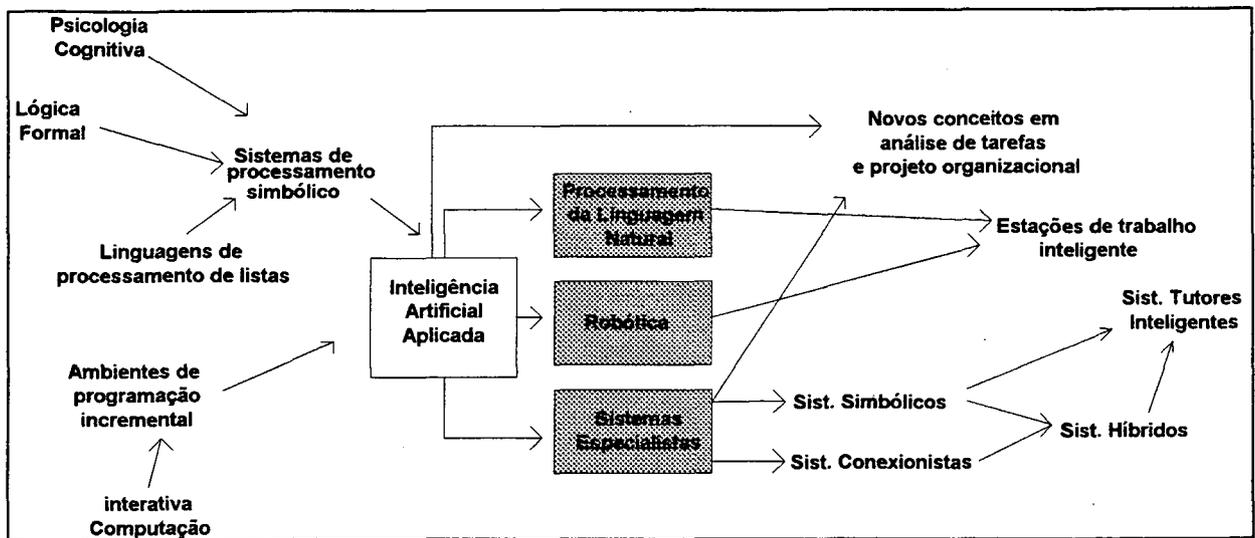


Figura 1. Evolução dos Sistemas baseados em conhecimento (HARMON, 1988).

Os aplicativos em IA são em geral desenvolvidos utilizando-se uma abordagem simbólica ou conexionista. Os sistemas que utilizam a combinação das duas técnicas são denominados de Sistemas Híbridos, que tentam absorver a capacidade dos sistemas simbólicos e sistemas conexionistas, e minimizar suas deficiências. (LOPES et al, 1993).

Através das pesquisas de processamento da Linguagem Natural e da Robótica são criadas cada vez mais estações de trabalho inteligente. A *Natural I/O* representa um avanço das pesquisas em Linguagem Natural para o processamento da voz, sendo uma interface capaz de reconhecer a voz humana como entrada de dados, combinando a tecnologia de sistemas híbridos com as técnicas de processamento de sinal digital (CAUDILL, 1992). Já no início da década de 90, Teuvo Kohonen da Helsinki University construiu um sistema que simula uma máquina de escrever utilizando os recursos da *Natural I/O*, ou seja, este sistema é capaz de escrever cada palavra dita.

Os fatores limitantes destes sistemas ainda são a quantidade de dados necessários para definir um único fonema e a grande variação de velocidade, tom de voz, pronúncia e dialética humana (CAUDILL, 1992).

### 2.1.1- Abordagem Simbólica

Um sistema especialista (SE), ou sistema cognitivo, é um programa que representa o mecanismo do pensamento do especialista relativo as questões profissionais da sua área (FEIGENBAUM,1988). O computador deve agir num mundo complexo usando o modelo computacional do raciocínio humano, alcançando as mesmas conclusões que o especialista humano. FEIGENBAUM (1988) define um SE como um programa inteligente que usa conhecimentos e procedimentos inferenciais, para resolver problemas difíceis exigindo muita habilidade e perícia.

A abordagem simbólica é baseada em questões onde a atividade humana é o raciocínio e não o cálculo exato, onde a habilidade em tomar atitudes corretas em decisões difíceis está baseada na inferência, julgamento simbólico e experiências passadas. Os programas tradicionais com a lógica de procedimentos não são capazes de fazer estas associações, pois seguem sempre uma mesma sequência (HARMON, 1988; SABBATINI, 1993).

A estrutura de um SE apresenta uma interface de aquisição que interage com o especialista, uma base modificável de conhecimento (regras e fatos), uma interface com o usuário e um mecanismo de inferência que faz a lógica das tomadas de decisões (LAWRENCE, 1992; SABBATINI, 1993), como mostra a figura 2.:

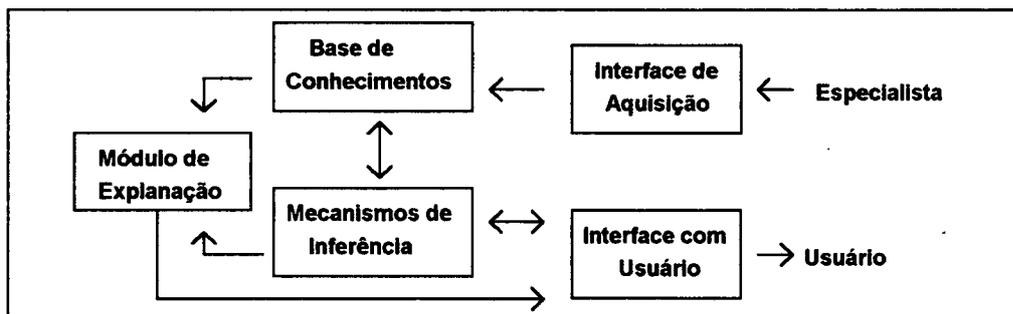


Figura 2. Estrutura típica de um SE (LAWRENCE, 1992, SABBATINI, 1993).

A interface de aquisição é responsável pela interação com o especialista que pode modificar e adicionar novo conhecimento. Os especialistas humanos determinam as regras e inserem fatos novos ao SE.

Os fatos descrevem um corpo de informação que é amplamente conhecido, e geralmente aceito pelos especialistas em um campo, como trechos de um livro de medicina. Os fatos são manipulados pelo sistema durante sua execução (LAWRENCE, 1992).

As regras (heurísticas) são em sua maioria restritas, regras pouco discutidas que caracterizam a tomada de decisão a nível de especialista na área. O nível de desempenho de um sistema especialista é função principalmente do tamanho e da qualidade da base de conhecimento que possui (WYATT, 1990; SHORTLIFE, 1991). As regras (heurísticas) descrevem o que fazer com os fatos e podem ser declarações do tipo SE ENTÃO, como por exemplo:

SE o paciente tem dor torácica;

E a dor é opressiva;

E a dor inicia rapidamente;

E a dor aumenta com o tempo

ENTÃO o diagnóstico é Infarto Agudo do Miocárdio.

As regras de produção (heurísticas) não garantem a solução mais eficiente, mas são úteis quando não existe nenhum algoritmo conhecido. São regras práticas (*rules of thumb*) que formalizam a tomada de decisão.

HARMON (1988) destaca o uso de heurísticas, uma base de conhecimento estruturado simbolicamente e programação orientada para processamento simbólico, como estrutura de um processo altamente interativo.

A base de conhecimento é a estrutura de dados onde os fatos e as regras de produção estão armazenadas.

O projetista do SE é chamado de "engenheiro de conhecimento" fazendo uma referência à tecnologia de "engenharia do conhecimento". O engenheiro de conhecimento é o responsável por transformar o conhecimento do especialista humano numa representação possível de implementação computacional, como por exemplo, em regras que determinarão as decisões do SE (KELSEY, 1991; LAWRENCE, 1992; RICH, 1993). Esta aquisição do conhecimento humano exige muito critério para que não ocorram inconsistências, pontos cegos, erros e contradições no conhecimento fornecido pelos especialistas humanos.

Um fato importante nos SE é que o raciocínio seguido deve ser o mais transparente possível e capaz de explicar a qualquer momento porque usou determinado procedimento. Esta tarefa é normalmente realizada pelo Módulo de Explicação do SE.

Um SE inicia interagindo com o usuário que propõe um caso para que se inicie o processo de raciocínio. Em alguns casos, esta etapa pode ser feita através de leitura direta de instrumentos destinados a controle de equipamentos industriais (LAWRENCE, 1992). Tendo todas as informações iniciais o sistema manipula os conhecimentos e os fatos, fazendo inferência aos novos fatos controlando o fluxo de dados. Esse mecanismo denomina-se motor de inferência. É a parte central de um SE, que interage com o usuário através de uma interface de comunicação, no modo de consulta e acessa a base de conhecimento para fazer inferências sobre o caso proposto pelo usuário.

A interface com o usuário apresenta também um módulo de explicação que deve informar, sempre que acionado, porque o sistema optou por uma decisão em particular. Esta possibilidade de se obter uma explicação durante a execução de uma tarefa é uma característica de um SE (HARMON, 1988).

O mecanismo de inferência deduz a resposta utilizando a técnica de encadeamento direto (*forward chaining*) ou encadeamento reverso (*backward chaining*).

O método de encadeamento direto trabalha em seqüência, analisando os fatos, aplicando as regras, produzindo mais fatos, aplicando mais regras, até que uma conclusão seja alcançada (LAWRENCE, 1992). É mais usado nos casos em que haja um grande número de possíveis soluções.

O método de encadeamento reverso faz o caminho inverso. Trabalha com uma hipótese como se esta fosse uma conclusão possível e vai interagindo com o usuário para que este confirme ou desconsidere esta hipótese. É também conhecido por sistema orientado por metas.

Nos SE é importante observar o grau de certeza que existe nas produções das regras. Esses fatores de certeza são fatores aproximados, baseados mais na pragmática do que na estatística. Uma regra de inferência pode ter um fator de certeza associado, expressando o grau de confiabilidade que é atribuído a uma conclusão. A teoria da probabilidade, baseada no teorema de Bayes, também fornece uma abordagem para calcular fatores de certeza, onde os números associam a probabilidade de cada evento dentro de um conjunto de possíveis ocorrências.

Uma outra forma de representação de conhecimento é a que um objeto é representado por um *frame* e seus atributos através de *slots*, contendo o nome do objeto, suas propriedades, valores, valores padrões (*default*), limites para valores, procedimentos (algoritmos ou telas associadas), regras para determinados valores, apontadores para outros *frames* e qualquer outro atributo relacionado ao objeto (TOWNSEND 1987), utilizando a técnica de programação similar à orientada por objetos (HELLER, 1991). Os *frames* individuais são interligados como nós numa estrutura de grafo ou redes semânticas, onde cada nó representa um fato e as regras são

representadas pelos arcos que ligam os diversos nós da rede. A solução apresentada pelo sistema é o resultado de uma exploração da base de conhecimento através da rede montada a partir dos nós e arcos.

Neste trabalho focaliza-se principalmente a abordagem simbólica. Em consequência, para simplificar a terminologia, sempre que for feita menção a SE (Sistemas Especialistas) entenda-se que a implementação é feita pelo paradigma simbólico a menos que seja explicitamente dito o contrário.

## 2.1.2- Abordagem Conexionista

As Redes Neurais (RN) representam uma outra técnica de representação do conhecimento em IA. Os SE dependem da forma de representação do conhecimento e da sequência lógica dos fatos. As Redes Neurais que formam parte dos Sistemas Conexionistas (CAUDILL, 1991), procuram imitar a organização paralela do cérebro, com as propriedades computacionais de paralelismos real e processamento distribuído.

As RN encontram soluções fazendo associações entre os dados de entrada e de um grande número de observações passadas, sendo capazes de resolver problemas não-lineares e complexos, através de generalizações, onde a precisão não é tão importante quanto a velocidade de resposta, como por exemplo, o movimento do braço de um robô.

As RN são capazes de se auto-organizarem, isto é, elas podem aprender novas tarefas de associações e classificações através de um mecanismo embutido de adaptação.

Inicialmente, acreditava-se que o desenvolvimento de um sistema de RN era bem mais complexo do que um SE, mas já foram publicados trabalhos demonstrando que em alguns casos isto não ocorre ( de AZEVEDO, 1991a, 1991b).

Uma RN adquire conhecimento diretamente de uma base de exemplos com suficiente número de casos significativos do domínio do conhecimento selecionado por um especialista. Nas RN a representação do conhecimento é feita automaticamente durante o seu processo de aprendizado pelas conexões sinápticas que representam matematicamente o conhecimento abstraído pelas redes para o problema (LAWRENCE, 1992).

As RN são utilizadas quando as regras não são determinísticas ou demasiadamente complexas, como por exemplo problemas estocásticos ou não-lineares, pois os neurônios artificiais das RN são capazes de dar respostas não lineares e são altamente interconectados.

Entretanto as RN não têm facilidade de explanação de seu raciocínio, ficando associado seu funcionamento ao de uma caixa-preta (HART, 1990). Com efeito, como o conhecimento fica distribuído pelas conexões se torna mais complexo que no caso simbólico. Entretanto existem algumas propostas para minimizar este problema (GALLANT, 1988).

Na modelagem do conhecimento nestes Sistemas Conexionistas pode-se utilizar algumas técnicas como o algoritmo de retropropagação, Lógica Difusa ("Fuzzy Logic") e Algoritmos Genéticos.

Uma das redes mais utilizadas na representação e processamento de informação médica é a rede neural multicamada que utiliza o algoritmo de treinamento *backpropagation*. Esta rede é não-linear, onde nenhuma informação é retropropagada durante sua execução. Seu esquema de aprendizado é supervisionado, onde um sinal de erro de saída é retropropagado pela rede, modificando o peso das conexões minimizando este erro (LIPPMAN, 1987). A rede com o aprendizado através do algoritmo *backpropagation* tem três ou mais camadas que são denominadas de entrada, invisíveis ou intermediárias, e de saída. Os nós da camada de entrada se conectam somente aos nós da camada invisível, e estes se conectam com os da camada de saída.

Esta rede possui estágios separados para aprendizado e operação e uma vez treinada o processo de aprendizagem se encerra. Durante a operação, nenhuma informação é realimentada.

A Lógica Difusa representa melhor o comportamento do cérebro humano do que a lógica binária, pois admite números fracionários entre 0 e 1, para descrever conceitos que são vagos em alguma área de diferenciação, o que não ocorre na álgebra booleana, onde os valores assumidos como lógicos, verdadeiro e falso, são representados numericamente apenas por 0 e 1 (JOHNSON, 1992; GARCIA, 1989, 1993; BARRETO, 1995). Assim para cada operação é avaliado todas as regras aplicáveis ao caso, combinando seus resultados.

O Algoritmo Genético está baseado em pesquisas genéticas, de forma semelhante à teoria de Darwin de seleção natural ou evolução das espécies. A estrutura de busca para solução de um problema, sofre modificações como corte, alteração e adição de novos caminhos. O processo se resume basicamente em três etapas: seleção randômica do caminho que será o criador da próxima estrutura, determinação das características da nova estrutura e mutação que trabalha com modificações dentro da nova estrutura, culminando com uma nova solução (JOHNSON, 1993; SPILLMAN, 1993).

É possível também trabalhar conjuntamente com a abordagem simbólica e conexionista, formando os Sistemas Híbridos, que são construídos para prover solução que cada abordagem isoladamente não seria capaz de solucionar (BARRETO, 1990; CAUDILL, 1991; de AZEVEDO, 1991, 1991b; GARCIA, 1992). Esta integração pode ocorrer a partir de várias técnicas, sendo cada uma aplicada à uma situação (LOPES, 1993):

- 1- Particionamento do problema: técnica que divide um problema complexo em pequenos problemas, utilizando o método mais apropriado para resolver cada item. É uma abordagem de arquitetura paralela, onde o fluxo de informações é gerenciado.

- 2- Redes neurais embutidas: é uma abordagem que a RN constitui uma parte do SE, que tem a função de gerenciamento.
- 3- Sistema Especialista Gerador de Explicações: mecanismo que permite que um SE seja ativado a partir das RN, gerando explicações adequadas para cada situação. Também é conhecido por confabulação.
- 4- Sistema Especialista Conexionalista: arquitetura que utiliza conceitos de Redes Semânticas, Redes Neurais e outras estruturas conexionalistas para implementar SE (MACHADO, 1990; de AZEVEDO, 1991b, GALLANT, 1993).

### 2.1.3- Sistemas Baseados em Conhecimento e a Aplicação na Medicina

Os sistemas especialistas, como uma parte dos sistemas baseados em conhecimento, vem sendo utilizados em todas as áreas de pesquisa. Na medicina estes sistemas são bastante utilizados para apoiar o médico a tomar decisões. Podem ser classificados em sistemas de gerenciamento de dados e sistemas com capacidade de raciocínio automático e de inferência. No primeiro grupo concentram-se os sistemas de recuperação de dados sobre pacientes, cálculos matemáticos assistidos por computador, análise e interpretação primária de dados. No segundo destacam-se os sistemas de classificação de doenças, os SE baseados em consulta e os SE baseados em crítica (SABBATINI, 1993).

Para áreas mais específicas, onde se utiliza uma base de conhecimento médico mais restrita, os SE são bem definidos em termos de representação do conhecimento, regras de decisão, dados para apoiar a decisão, padronização de nomenclatura e concordância entre os especialistas, e são capacitados para produzir resultados úteis, como apoio à resolução de problemas de diagnósticos complexos. Estes sistemas restritos são considerados como especialistas confiáveis (van BEMMEL, 1986).

Na década de 70 , quando iniciou-se o primeiro trabalho de aplicação das técnicas de IA em medicina, foi o construído o MYCIN, que hoje, após exaustivos testes apresenta resultados satisfatórios (CLANCEY, 1984; LUNDSGAARDE, 1987, ). Existem vários SE aplicados à medicina como por exemplo:

DTA - consultas sobre terapias com digitais;

PIP - consulta e diagnose na medicina interna;

CASNET/GLAUCOMA - para o diagnóstico e aconselhamento terapêutico nos caso de glaucoma;

PUFF- diagnóstico de distúrbios respiratórios;

ABEL - identificação de distúrbios eletrolíticos e ácido-básicos e aconselhamento terapêutico;

VM - interpretação e consulta sobre a terapia intensiva em relação a pacientes em estado grave;

ONCOCIN- seleção de protocolos oncoterápicos;

ILIAD- consulta e diagnóstico para medicina intensiva.

Uma aplicação também muito interessante e bem aceita pelos profissionais da saúde é o uso de SE para alerta automático de riscos de saúde, procedimentos básicos e pesquisa de dados laboratoriais a partir dos dados armazenados no registro clínico de um hospital. O sistema HELP realiza este tipo de tarefa apresentando um desempenho aceitável.

Existem grandes perspectivas para a área médica e computacional, onde a integração gera cada vez mais recursos, como por exemplo, a facilidade de fazer relações estatísticas entre variáveis de um banco de dados médicos ou a utilização de RN em várias pesquisas médicas possibilitando o armazenamento de milhares de informações, simulando o processo dos neurônios humanos.

## 2.2 - Tomada de Decisão Médica

Os SE de apoio à tomada de decisão simulam o raciocínio dos especialistas para determinar um diagnóstico médico ou resolver um problema clínico.

A resolução do problema clínico é um processo metódico que pode ser aprendido e ensinado (JOHNS , 1988). É um procedimento interativo, onde o paciente apresenta sinais e sintomas, além de testes laboratoriais. Também é um processo contínuo onde nenhuma hipótese é irrefutável (FLETCHER, 1991).

O processo de diagnosticar é dinâmico e inicia no primeiro contato com o paciente (FLETCHER,1991). Pode ser dividido em três fases distintas : ativação de uma hipótese, avaliação da hipótese e coleta de informações. Essas fases são iterativas e para cada nova hipótese deve-se refazer a avaliação e coleta de informações necessárias (KASSIRER,1978).

Estabelecer um diagnóstico é um processo imperfeito, resultando em probabilidade e não em certeza de se estar correto. No processo de obtenção de informação clínica para o diagnóstico ressalta-se a importância de alguns itens como (JOHNS , 1988; FLETCHER, 1991) :

### **1- Exatidão**

Determina quanto a informação representa a realidade. Por exemplo se for um teste laboratorial de um paciente com suspeita de alguma doença, faz-se uma comparação com um "padrão ouro", ou seja com um teste de uma pessoa sem suspeita da mesma doença. Este "padrão ouro" pode ser qualquer parâmetro considerado perfeito, mas como na prática este padrão é difícil de ser encontrado, escolhe-se como "padrão ouro", um que sabidamente não é o perfeito, mas é considerado o melhor possível.

### **2- Ocorrência precisa**

Medida da extensão da capacidade da reprodução da informação. Em situações semelhantes as informações recolhidas devem ser as mesmas.

### **3- Variação**

Refere-se as possíveis variações que a informação pode sofrer dependendo do modo que o fato foi observado. É comum encontrar variações no método de observação, no próprio observador e no próprio fato a ser observado.

### **4- Sensibilidade e especificidade**

A sensibilidade representa a proporção de indivíduos com uma doença que apresenta um teste positivo para esta doença.

Especificidade determina a proporção dos indivíduos sem a doença que tem um teste negativo. Um teste específico raramente erra (FLETCHER, 1991).

O uso do teste sensível é indicado para confirmar a presença de uma doença que não se pode correr o risco de não se diagnosticar.

O teste específico é mais necessário para se ter certeza de um diagnóstico surgido por outros dados. Este dado é importante quando o paciente pode se sentir lesado física, emocional ou financeiramente por um diagnóstico errado para a presença de uma doença.

Associado ao teste de sensibilidade e especificidade tem-se uma probabilidade da doença. Este valor chama-se valor preditivo. O valor preditivo positivo é a probabilidade de uma doença em um paciente com o resultado de um teste positivo. O valor preditivo negativo é a probabilidade de não ter a doença quando o resultado do teste é negativo. O valor preditivo é a característica mais relevante para os que interpretam resultados dos testes.

### **5- Validade**

Está intimamente ligado aos testes de sensibilidade e especificidade. As consequências da não validade das informações, testes laboratoriais ou respostas a questionários, podem ser graves como:

- i) o custo do erro cometido pode ser alto;

ii) a pesquisa exaustiva de um fato isolado e específico que não representa exatamente o problema;

iii) direcionamento do tratamento de acordo com um fato que não tem muita importância no processo da doença.

#### **6- Risco**

Risco refere-se à probabilidade de algum evento indesejado predizendo a ocorrência de uma doença. Esses eventos são os fatores de risco. A presença de um fator de risco aumenta a probabilidade de que uma doença esteja presente, portanto esta informação pode ser usada no processo diagnóstico. Para a avaliação do risco não existe uma escala exata, e além disto pode-se enumerar mais alguns fatores como:

- i) o desconhecimento da probabilidade de um resultado indesejado;
- ii) a dificuldade para ponderar a gravidade de resultados desfavoráveis.

#### **7- Custos e Benefícios**

Custos freqüentemente estão relacionados aos custos financeiros da coleta de informações. Mas também incluem os efeitos colaterais, a redução do tempo de vida e a perda de outros recursos como tempo, energia e disponibilidade.

Os benefícios resultantes das informações são a possibilidade de aliviar o sofrimento do paciente, restaurar uma função e prevenir a morte prematura. Um problema comum relacionado ao item benefício é a incapacidade de distinguir entre o interesse clínico e o benefício do paciente.

Hierarquicamente pode-se dizer que os médicos se preocupam com os benefícios da informação clínica, os riscos associados e, por último, com os custos (FLETCHER, 1991).

A tomada de decisões clínicas requer a consideração dos benefícios e custos de uma ação. Tanto os benefícios como os custos devem ser amplamente definidos, de modo incluir todas as consequências importantes da decisão.

Freqüentemente é necessário fazer um balanço entre os diferentes benefícios e custos de uma ação, por exemplo, tempo versus qualidade de vida na quimioterapia. Este balanço é feito sem regras formais ou sem o uso de probabilidades explícitas. É o que se chama de julgamento clínico.

Para a tomada de decisão propriamente dita, ainda pode-se classificar os modelos de decisão usados no raciocínio médico após os primeiros contatos com o paciente e os testes laboratoriais (quando necessário). Estes modelos misturam o raciocínio médico com o conhecimento adquirido dos estudos e pesquisas bibliográficas. A tomada de decisão pode ser caracterizada por quatro modelos segundo van BEMMEL (1986):

### **1- Modelos Estatísticos**

É utilizado para o diagnóstico diferencial, pois é possível determinar limites de confiança para a doença ou para a normalidade. Através da curva de distribuição das probabilidades da doença pode-se aplicar o teorema de Bayes.

### **2- Modelos Lógicos**

São baseados nos resultados de decisões sim/não, formalizados como decisões consecutivas em forma de árvores ou tabelas verdade. Teoricamente não existem diferenças entre os modelos estatísticos e lógico, a não ser pelo fato de que o modelo estatístico usa o conjunto total de dados para determinar a solução e o lógico vai decompondo a classificação em uma série de pequenas decisões.

### **3- Modelo Inverso**

É uma categoria diferente de modelo pois pressupõe uma descrição quase perfeita do problema. Exige um nível muito alto de abstração. Raramente se tem conhecimento científico suficiente para desenvolver os modelos inversos.

#### 4- Modelo de Raciocínio

São também conhecidos como raciocínio heurístico e é muito mais baseado na prática do especialista do que no conhecimento teórico. É o tipo de conhecimento que é comumente usado na prática médica seguindo uma linha de raciocínio, numa seqüência de perguntas e respostas onde gradativamente a solução surge.

Finalizando, pode-se dizer que o processo de tomada de decisão é um processo dinâmico, interativo e contínuo, sempre fazendo uma revisão das possibilidades, mas o encontro com o paciente é o que realmente capacita o médico a obter a informação que vai levá-lo à conclusão mais apropriada (JOHN, 1988a; KASSIRER, 1993).

### 2.3 - Aquisição de Conhecimento

Para o desenvolvimento de um sistema baseado em conhecimento, é necessário que o engenheiro de conhecimento trabalhe junto com o especialista da área que se deseja construir o sistema especialista, para que se efetue a fase de aquisição de conhecimento (PEREIRA, 1993; BRASIL; 1994).

Existem vários métodos de aquisição de conhecimento mas, de um forma geral, este processo apresenta as seguinte fases, como apresentado na figura 3:

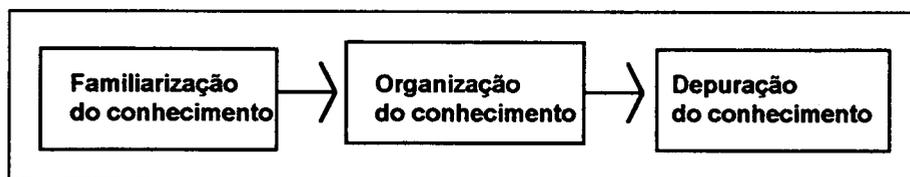


Figura 3. Esquema geral de aquisição de conhecimento (BOTTEN, 1989; PEREIRA, 1993).

### **1- Familiarização do engenheiro do conhecimento com a área do sistema**

Nesta fase o engenheiro de conhecimento deve estudar e discutir com o especialista os aspectos teóricos relevantes no domínio do conhecimento, tendo então um primeiro contato com os elementos, relações e procedimentos. Este processo pode ocorrer através de entrevistas informais ou da observação quando o especialista humano está resolvendo um problema.

### **2- Organização do conhecimento**

Feita a coleta de informações deve-se definir a base de conhecimento relacionando quais os fatos que podem ser unificados, quais as informações significativas e quais as que podem ser eliminadas. A interação com programas computacionais permite depurar e/ou modificar uma base de conhecimento. O especialista humano deve interagir junto com o engenheiro de conhecimento, tomando o cuidado de selecionar amostras que não deixem margens a um eventual questionamento estatístico.

As regras também são montadas nesta fase, onde o engenheiro de conhecimento deve explicitar todas as possíveis relações de causa e efeito dentre os elementos observados, através das cláusulas SE...ENTÃO. Um nível adicional na organização do conhecimento que está sendo gerado, diz respeito a um agrupamento das regras em função da contribuição de cada uma no processo de diagnóstico, na identificação da causa geral ou na identificação da causa básica dos sintomas observados. BYLANDER (1987) sugere que em uma área de conhecimento, diferentes problemas podem ser tratados, cada qual com regras apropriadas.

### **3- Depuração do conhecimento**

Esta etapa compreende a verificação e validação da base de conhecimento. BOTTEN (1989) lembra que verificar e validar vêm do latim *veritas* e *valere* significando respectivamente verdade e valor. Logo um SE está verificado se sua base de conhecimento for consistente e validado se o sistema oferece soluções adequadas para os casos propostos.

Para verificar a base de conhecimento, é desenvolvida uma análise lógica procurando identificar inconsistências na base de conhecimento, como regras conflitantes, redundantes e que também estejam faltando. A análise de uma regra feita juntamente com o especialista humano, pode levar à análise de outras regras relacionadas. Isto pode seguir uma linha de raciocínio completa ou então partir para a análise de cadeias similares. Na discussão de uma regra o especialista pode acabar associando informações paralelas pertinentes à situação.

Estas etapas são geralmente seguidas na construção de SE para qualquer área, mas na área médica, os engenheiros de conhecimento enfrentam um problema adicional. A tomada de decisão médica é baseada em duas correntes que são o conhecimento empírico e o conhecimento específico (ou teórico). A formalização do conhecimento empírico, isto é, a formalização do conhecimento adquirido através da experiência durante os tratamentos com pacientes, é uma tarefa complexa (van BEMMEL, 1986).

Para a aquisição do conhecimento médico tem-se também o problema de adquirir conhecimento de diferentes especialistas da mesma área. Um ponto crucial é determinar se o conhecimento presente já está na base, quando existe conflitos internos e quando a base já está completa o suficiente (van BEMMEL, 1986).

Na tentativa de solucionar este problema ROCHA (1990) apresenta uma técnica de aquisição de conhecimento utilizando RN difusas, onde especialistas expressam seu conhecimento a respeito de um problema agrupando um conjunto apropriado de evidências, na forma de um grafo composto por arcos de ligações ponderadas "E" e "OU", partindo das evidências até as hipóteses. O sistema permite armazenar diferentes grafos de conhecimentos (redes semânticas), representando o ponto de vista de cada especialista na área. No momento de consulta o usuário pode especificar por qual ponto de vista deseja que o processo de dedução de uma hipótese seja orientada. O sistema pode ainda, a partir dos diferentes grafos de conhecimento que têm

armazenado, criar sua própria representação do conhecimento, realizando um consenso dos diferentes pontos de vista dos especialistas da área (MACHADO, 1990; ROCHA, 1990).

Por conhecimento específico entende-se o entendimento do problema e do processo total de uma doença, e o conhecimento empírico engloba a experiência adquirida durante o tratamento do paciente. O primeiro tipo de conhecimento é obtido no processo da pesquisa médica, o segundo já tem uma característica fenomenológica.

Os dois tipos de conhecimento são bem documentados em artigos e revistas médicas, mas o conhecimento empírico se apresenta de uma forma mais informal. Neste tipo de conhecimento o que se valoriza é a experiência adquirida em tratamentos, representando uma parte importante do conhecimento especialista.

Para formalizar o conhecimento de um médico com o intuito de montar um SE, o conhecimento empírico é o que causa maior transtorno aos engenheiros de conhecimento, porque nem mesmos os médicos sabem explicar como usam o conhecimento na prática diária (van BEMMEL, 1986).

Na tentativa de formalizar o conhecimento médico vários pontos devem ser considerados. O primeiro é a **especificação da área** do problema, onde deve-se ter a noção exata da abrangência da área em estudo, verificando as restrições e limites do sistema a ser construído.

A documentação obtida a partir das fichas dos pacientes de um determinado hospital ou de um estudo cooperativo forma a **base de dados**, formando a base de conhecimento do sistema.

Em seguida deve-se decidir qual **modelo de decisão** será utilizado para a construção do sistema. Estes modelos podem ser baseados em estratégias matemáticas e/ou heurísticas.

A relação **custo/benefício** deve ser otimizada. Esta relação não deve ser expressa em termos financeiros mas sim em qualidade de vida para o paciente.

Finalmente, é necessário **testar** o sistema, utilizando casos selecionados ou aleatórios.

O processo de aquisição da informação médica e a tomada de decisão é um círculo fechado onde se enfatizam as etapas da coleta de informações, diagnósticos e o retorno dos especialistas sobre o resultado das terapias. Nota-se que a decisão é baseada no conhecimento específico ("porquê faz") e empírico ("como se faz"), onde geralmente o conhecimento específico é traduzido pelas regras e o empírico pelas bases de conhecimento.

As técnicas de IA abriram um novo horizonte na medicina, permitindo a resolução de problemas médicos, tirando o abismo que existia entre medicina e computação através dos sistemas especialistas que une as duas áreas (van BEMMEL, 1986), mas como não existe uma padronização para avaliação desses sistemas, esta integração ainda não alcançou o sucesso esperado.

Mais do que um "padrão ouro" para comparação, a avaliação deve ser entendida como um processo que faz parte do desenvolvimento do sistema, cobrindo a fase de planejamento, aquisição de conhecimento e finalmente a implementação do SE. Mesmo depois da aceitação do sistema pelo usuário, a avaliação deve continuar auxiliando o usuário a julgar a expectativa do tempo de vida do sistema, de acordo com a manutenção necessária e as novas idéias para aprimoramento do sistema e para a especificação de futuros projetos. A avaliação exige um constante acompanhamento do desempenho e das mudanças no ambiente de trabalho.

## Capítulo 3- Modelos de Avaliação dos Sistemas Especialistas

Uma revisão na literatura sobre o estado da arte mostrou que avaliação para SE não tem um conceito universal de escalas ou critérios de medidas. Avaliação não é só evitar o erro nem saber como um programa de computação vai funcionar em situações imprevisíveis. LUNDSGAARDE (1987) contribui para as pesquisas de avaliação fazendo a seguinte definição: "avaliação é a coleta e a análise da informação de várias estratégias metodológicas que determinam a importância, abrangência, eficiência e impacto sobre o uso dos programas computacionais".

Sendo assim, a avaliação não deve ser entendida como uma medida de comparação entre o conhecimento do médico e do SE, mas sim uma medida entre os custos e benefícios de cada um. Os SE não foram projetados para substituir humanos, mas para fornecer informações e apoiar a tomada de decisões para usuários da área médica (WHITEBECK, 1983).

Os especialistas humanos também sofrem uma avaliação no seu desempenho, geralmente baseado em critérios subjetivos, pois a própria ética médica inibe certas comparações (WHITEBECK, 1983). Se é feita uma consulta ao especialista e ele aponta o caminho certo, fica evidente sua base de conhecimento na área, refletindo seu conhecimento e experiência profissional.

Já o desempenho dos sistemas computacionais não tem nenhuma implicação ética, e seus resultados são comparados com as opiniões dos especialistas humanos, geralmente através de análise dos resultados. Porém se a conclusão do especialista não coincidir com a do sistema, isto já é suficiente para não ser aceita a sugestão do SE. O responsável sobre a tomada de decisão sempre é o médico que faz uma avaliação da sugestão do SE (BAINBRIDGE, 1991).

Mesmo que os projetistas de um SE façam um estudo estatístico cobrindo as ocorrências dos casos mais comuns de combinações de sinais e sintomas, os especialistas sempre encontrarão

combinações que não estão presentes nos dados estatísticos. Também não se pode afirmar que se para uma determinada combinação de casos o desempenho foi satisfatório, o SE será sempre bem sucedido em qualquer combinação de sintomas.

Dado esse enfoque no problema de avaliação, WHITEBECK (1983) propôs que para um SE ser avaliado é importante que qualquer resposta venha acompanhada de um explicação e possibilidades, demonstrando sua linha de raciocínio, habilitando o usuário a avaliar a sugestão sob três diferentes aspectos:

- 1- Verificação das particularidades no mecanismos de inferência;
- 2- Julgamento da exatidão e abrangência da base de conhecimento do SE;
- 3- Exame das informações relevantes da base de conhecimento, verificando se o sistema usa consistentemente as informações que tem armazenadas.

WHITEBECK (1983) mostra que a abrangência deve ser considerada como um dos tópicos na avaliação. É importante ressaltar a diferença que existe entre exatidão (*accuracy*) e abrangência, pois um sistema desenvolvido baseado nos sinais e sintomas de um determinado grupo racial e étnico pode apresentar resultados imprecisos se for aplicado em outro grupo, mas neste caso não se pode dizer que o SE tem um baixo grau de exatidão.

Quando se utiliza dados de pacientes para avaliar o desempenho de um SE, deve-se levar em conta também para que tipo de usuário foi desenvolvido o SE, ou seja a avaliação está intimamente ligada ao escopo dos objetivos do SE (WHITEBECK, 1983).

### 3.1- Proposta de Gaschnig

Considerando estas questões GASCHNIG (1983), propôs uma escala para que todos os SE na área médica fossem avaliados. Essa escala consiste nos seguintes passos:

- 1- Numa visão geral definir os objetivos a serem alcançados pelo SE;
- 2- Implementação de um protótipo para demonstrar sua utilidade;
- 3- Refinamento do sistema;
  - a) testes com casos aleatórios para obter um retorno (*feedback*), fazendo um primeiro refinamento do protótipo;
  - b) liberar cópias refinadas para usuários menos exigentes solicitando retorno;
  - c) liberar novas cópias para usuários finais e esperar o retorno;
- 4- Avaliação do desempenho;
- 5- Avaliação da aceitação pelos usuários;
- 6- Funcionamento dos protótipos em tempo integral;
- 7- Fazer uma pesquisa para demonstrar a utilidade do sistema em larga escala;
- 8- Fazer alterações necessárias garantindo a ampla distribuição do sistema;
- 9- Liberação definitiva do sistema e definição de uma estratégia de *marketing* garantindo a manutenção e atualização do sistema para os usuários;

Em 1987, LUNDSGAARDE faz uma análise dos SE desenvolvidos e constata que apenas o sistema HELP chegou a passo 9 da escala de GASCHNIG , sendo que a maioria não ultrapassa o passo 3.

LUNDSGAARDE (1987) mostra que INTERNIST, um programa de diagnóstico computadorizado projetado para uso na medicina interna, ainda não foi implementado e testado num ambiente clínico. Entretanto foi feita uma avaliação empírica, subjetiva pela equipe que desenvolveu o sistema, apenas comparando os resultados que o sistema apresentava com a opinião dos especialistas. Foi testado cerca de 900 casos e o INTERNIST teve um acerto de 50%. Foi considerado um bom desempenho, embora os especialistas são cuidadosos em afirmar que este sistema não faz diagnóstico e nem pretende tirar o julgamento do médico, apenas auxilia nas análises estatísticas de um grande volume de dados.

Um outro sistema, o PUFF, passou a ser operacional em 1979. Sua base de conhecimento foi testada em 107 casos escolhidos de um arquivo do laboratório de doenças pulmonares. De acordo com os resultados apresentados, foram feitas modificações no sistema. Esta interação continuou até que os especialistas consideraram que o sistema estava respondendo como um próprio médico. Depois, um outro grupo de especialistas, que não participou do seu desenvolvimento, fez mais uma série de testes. Os dados mostram que no primeiro grupo houve um acerto de 96% dos casos, já no segundo este índice passou para 82%.

JENKINS (1981), especialista em cardiologia, mostra que a maioria dos casos são avaliados apenas por comparação com a opinião dos especialistas, e afirma que este procedimento não estabelece nenhum padrão criterioso para avaliar o desempenho do sistema. Suas observações criticam os itens metodológicos para avaliação como padrão, escala de medida, propósito de estudo da avaliação e questiona o critério para julgar a inteligência dos especialistas humanos. Mas o sistema DIAGNOSER, projetado para apresentar diagnósticos referentes a doenças congênitas no coração, também teve sua avaliação baseada na comparação dos resultados, porém foi escolhido para avaliar quatro especialistas com diferentes graus de experiência na área.

VANKER (1984) desenvolveu a primeira base de conhecimento de consulta médica para dermatologia na Escola de Medicina da Universidade de Missouri, que resulta num SE para representar conhecimento específico com a intenção de ajudar nas consultas médicas identificando tumores de pele. Após várias interações com a equipe médica o modelo final aponta um diagnóstico com precisão de 84%, baseado nas comparações dos resultados.

HABBEMA (1985) contribui relatando o desenvolvimento, implementação e avaliação do DIAG, um SE desenvolvido no Hospital Western em Toronto. Seu desenvolvimento levou 4 anos. Um protótipo do DIAG foi instalado na unidade de dermatologia que atende cerca de 100 pacientes por dia. Através de um constante trabalho de interação com a equipe médica e a

diretoria do hospital, as recomendações sugeridas (*feedback*) fizeram com que os analistas do DIAG modificassem várias partes do sistema, aumentando a eficiência dos diagnósticos e sua utilização por toda equipe médica.

ONCOCIN, um SE de gerenciamento de terapias de câncer não traz muita discussão sobre sua avaliação (LUNDSGAARDE, 1987). Um outro SE na mesma área foi desenvolvido através do Breast Cancer Project em Alabama permitindo que uma comunidade restrita de médicos e os que não têm acesso nem a especialista, nem a pesquisas em Universidade tivesse contato com um SE em oncologia. Os projetistas afirmam que a consulta pelo sistema habilitou a comunidade médica, em cooperação com outros especialistas regionais, a administrar a quimioterapia. O sistema se encarrega de protocolar formas e indicar dados para serem coletados para sua administração. A avaliação deste sistema é descrita como um processo "passo a passo" com o médico acompanhando todo o raciocínio apresentado. O sistema foi testado com os dados em andamento de cada paciente, ao invés de dados de arquivos, como a maioria dos outros casos de avaliação. Segundo os projetistas isto foi um ponto positivo, sendo o retorno das informações mais rápido, tornando o relacionamento com a equipe médica mais freqüente. Os projetistas concluem que 5 anos de estudos deveriam ser conduzidos para tornar o sistema completamente avaliado, mas perguntam-se se este não é um tempo muito longo. Este é um sistema que alcançou o estágio 6 da escala de GASCHNIG.

Na área cirúrgica existem vários SE, mas muito pouco se fala sobre as avaliações. O Medical College da Georgia, fez um estudo para a avaliação de um SE partindo desde a fase inicial do desenvolvimento. Foi feito um algoritmo para atendimento e testado num grupo de enfermeiros responsável que passou a apresentar melhores resultados no tempo e na qualidade do atendimento do que um outro que não havia recebido o algoritmo. Este estudo mostrou como um simples protótipo pode ser implementado trazendo benefícios para os pacientes e sem um custo muito grande de recursos ou tempo. A avaliação final, neste caso é a comparação do desempenho obtido por enfermeiras utilizando o SE com os procedimentos sugeridos pelos

especialistas. O grande benefício neste tipo de avaliação foi a melhora no cuidado ao paciente, através da sistematização do tratamento emergencial, mas levantou uma séria questão entre os médicos, que é a dúvida se o uso contínuo do SE reduzirá o *status* profissional passando os clínicos a serem supervisionados apenas por administradores e programas computacionais. A esta questão de DOMBAL (1979) demonstra através de pesquisas estatísticas que os SE tem a função primordial que é a de auxílio e não de substituição ao médico.

O SE DIAGNO II, projetado para simular julgamento clínico psiquiátrico, foi avaliado através da comparação do seu desempenho com 9 psiquiatras de clínicas particulares e 13 que estavam no segundo ano de residência em psiquiatria. Foram estudados 100 casos reais e 46 hipotéticos. A validade foi estabelecida mostrando que o programa gerava julgamentos similares aos clínicos. Apesar disto os especialistas consideram que um SE para a psiquiatria ainda está numa fase inicial pois apresenta uma abrangência limitada.

O sistema HELP é o único a alcançar o passo 9 da escala proposta por GASCHNIG. Mas apesar de existirem muitos artigos sobre técnicas e campos de atuação é difícil descrever exatamente como o sistema foi avaliado. Ressalta-se que o HELP foi desenvolvido como um sistema modular onde cada módulo (como por exemplo patologia laboratorial, radiologia clínica, cardiologia) foi conceituado como uma entidade independente. Assim, cada módulo foi avaliado como um trabalho de cooperação entre os especialistas e os projetistas. O uso do SE pelo staff médico era voluntário e isto proporcionou uma maior integração entre os médicos e projetistas.

A estratégia de avaliação feita quando o HELP estava na fase de implementação no Hospital Arnot Ogden em Utah foi diferente. Cada grupo de especialista do desenvolvimento perguntava ao grupo correspondente na especialização "o que o sistema faz para/por você?". Todos os grupos responderam que o HELP aumentava a qualidade dos cuidados e facilitava o trabalho.

Para a diretoria deste hospital, o retorno favorável das impressões obtidas justificava o investimento na aquisição do HELP. Entretanto este tipo de validação informal não é sistematizada nem publicada oficialmente. É mais fácil saber como e porquê um grupo de médicos e administradores compram um sistema computadorizado do que ter relatórios mostrando como um SE apresenta resultados similares aos de 2 especialistas em 95% dos casos (LUNDSGAARDE, 1987). A partir de 1981 o HELP passou a ser comercializado, chegando a última etapa da escala proposta por GASCHNING.

### 3.2- Proposta de Lundsgaarde

LUNDSGAARDE (1987) faz algumas críticas ao processo de avaliação afirmando que as técnicas existentes são precárias e primitivas. Assim, propõe uma nova escala, que está diretamente relacionada ao estágio de desenvolvimento e a relativa complexidade do sistema. Acreditando que as estratégias de avaliação, boas ou ruins, influenciam no desenvolvimento e no uso clínico de futuros sistemas, o autor enfatiza que os sistemas devem ter as mesmas escalas para a avaliação. Sua proposta consiste em separar o sistema em 3 níveis:

#### **Nível 1- Estágio conceitual**

O SE trabalha como um protótipo. Nesta fase a equipe de desenvolvimento precisa ter bem claro os objetivos e os testes de desempenho do sistema. Faz o refinamento dos algoritmos em resposta ao retorno (*feedback*) de outros laboratórios ou de outras partes interessadas. Aplicação em clínica "limitada" pode ser julgada como base experimental. Nenhuma saída ou avaliação formal do protótipo do sistema pode prejudicar o avanço do desenvolvimento do sistema. Ainda não traz nenhum benefício para os usuários do SE.

### **Nível 2- Estágio intermediário**

O SE está pronto para ser usado num ambiente clínico de cuidado com a saúde. O projetista deve ter um pequeno contato com o usuário em potencial. No local do desenvolvimento e nas clínicas dos usuários as avaliações são mais práticas e econômicas. O objetivo de tais avaliações é prover direta e continuamente o projetista com o retorno (*feedback*) do usuário. Algumas funções do sistema podem ser eliminadas e outras incluídas, de acordo com o interesse dos especialistas. Neste ponto sugere-se a criação de um documento formal contendo os testes realizados e as alterações necessárias.

### **Nível 3 - Estágio de maturidade**

O SE alcançou maturidade. Ainda está sujeito a modificações e trocas mas o essencial está pronto para ser usado nas clínicas. A avaliação do sistema agora necessita de julgamento clínico e de usuários experientes que não participaram dos testes nos níveis anteriores.

A noção de escalas de desenvolvimento com avaliação pode auxiliar para distinguir entre custos e benefícios no ambiente de desenvolvimento e os esforços externos de avaliação. A idéia de avaliação em estágios não é nova já tendo sido sugerida por Arnstein e Christakis em 1975 (LUNDSGAARDE, 1987).

## **3.3- Proposta de Hart e Wyatt**

HART & WYATT (1990) consideram que a avaliação deve ser vista sob os aspectos ético, legal e intelectual e sugere uma metodologia de avaliação dividida em 3 fases acompanhando o desenvolvimento do sistema.

A primeira fase é a **definição do sistema**, onde se identifica qual o problema que o SE deve solucionar, o tipo de usuário e o ambiente que atuará. Nem sempre é possível definições completas e sugere-se a construção de protótipo para obter *feedback*, formando o ciclo construir-testar-refinar. Para facilitar o trabalho futuro de avaliação do processo, deve-se exigir um especialista na área médica. Ao final desta fase o protótipo está concluído.

Em seguida passa-se a fase de **teste laboratorial**. Do mesmo modo que um farmacêutico hesitaria em prever os efeitos de uma nova droga sem testes clínicos, o SE deve passar por exaustivos testes laboratoriais (YOUNG, 1987). Os usuários do sistema (enfermeiras, médicos residentes) e os especialistas representam os dois grupos de maior interesse para esta fase de teste, onde o sistema já é visto como uma ferramenta para auxiliar na tomada de decisão. Ressalta-se que o sucesso de um SE está ligado ao fato de ser desejado, usável no ambiente clínico e apresente conclusões razoáveis para o usuário. Para observar estas condições HART & WYATT (1990) sugerem questões específicas cobrindo os parâmetros estrutura, processo e resultado, apresentadas na tabela 1.:

ASPECTOS	USUÁRIOS	ESPECIALISTAS
Estruturas	O sistema é desejado? Tem uma aplicação específica?  As respostas podem ser utilizadas corretamente e em tempo hábil?	O sistema é de boa qualidade? O conhecimento apresentado é apropriado? A parte central do conhecimento está representada?
Processo	A interface é amigável e apresenta opções coerentes? As telas são claras?  Apresenta um diálogo agradável com o usuário? O sistema é rápido o suficiente?	A lógica é consistente?  O método para representar incerteza está correto? O sistema é robusto para a variação dos dados? O sistema de controle definido está claramente representado?
Resultados	O sistema apresenta respostas coerentes? Suas conclusões parecem com as de um especialista? As explicações são aceitáveis?	Consegue detectar casos fora do escopo? Comete erros dentro do seu domínio?  Comparando com sua prática, qual a exatidão dos resultados apresentados?

Tabela 1. Aspectos avaliados.

A aplicação deste ou outro questionário pode ser considerada uma estratégia de avaliação subjetiva, mas é difícil formalizar o nível de satisfação humana perante um sistema computacional.

A seleção de casos para os testes laboratoriais não deve ser puramente randômica nem devem ser reutilizados os dados da fase de depuração do sistema. O ideal é que seja elaborado um conjunto de dados considerados difíceis por uma equipe de médicos independentes, que não tenham participado do desenvolvimento. A esse conjunto deve-se misturar uma pequena proporção de casos aleatórios. Assim testa-se a abrangência do conhecimento da base do sistema.

É raro encontrar uma única resposta certa para um diagnóstico clínico, tornando a avaliação por comparação dos resultados um problema crítico. A formalização de um padrão de resposta também deve ser obtida através de uma equipe de especialistas que devem julgar os mesmos casos utilizando os mesmos critérios. É importante ressaltar que os casos testados não devem ter sido coletados por nenhum destes especialistas.

Finalmente, a terceira fase é a de **estudos de campo** onde se observa as influências do sistema no ambiente em que foi instalado. Um questionário também pode servir de guia para esta fase como mostra a tabela 2.

ASPECTOS	USUÁRIO
Estrutura	Os usuários aceitam o sistema? Os dados são válidos para o sistema? O sistema apresenta mensagens significativas ? É possível integração com os outros da mesma área?
Processo	O sistema é usado frequentemente ou existe alguma objeção? Existem algum tópico do sistema que não seja usado? O sistema tem algum efeito no atendimento médico como tempo de consulta, número de atendimentos e mudança no comportamento do usuário? O sistema tem alguma influência subjetiva como satisfação dos usuários, interação entre usuários, pacientes e administradores?
Saída	Melhorou a qualidade dos serviços para o paciente? O índice de mortalidade diminuiu? Qual a relação custo/ benefício do uso do sistema?

Tabela 2. Aspectos para avaliação de campo

Depois de passar por estes estágios, todos os resultados negativos e positivos devem ser publicados para que os erros cometidos não se repitam na construção de um sistema futuro. HART & WYATT (1990) afirmam que este método ainda é subjetivo e sugere que mais esforços no sentido de definir uma metodologia ainda devem ser feitos.

FIESCHI (1990) divide a avaliação dos SE em duas etapas: **estática** e **dinâmica**. Avaliação estática se refere apenas ao conteúdo da base de conhecimento e não ao modo de usar suas informações, e por isto, nesta fase, prefere usar o termo verificação. Esta abordagem permite observar a coerência, abrangência (*completeness*), e a detecção de erros de conteúdo, mas não acusa erro por falta de informação. Abrange os diferentes tipos de verificação:

- 1- lógica: detenção de contradição e conflitos;
- 2- estrutura: existência de duplicação de regras, redundância, objetos não utilizados nas regras, regras circulares, objetos irrelevantes;
- 3- semântica: valores limites, incompatibilidade;

Para abordagem dinâmica prefere o termo teste ou validação e direciona as observações para os resultados obtidos pelo sistema. Os resultados devem ser analisados por especialistas que não participaram do desenvolvimento, fato este que FIESCHI (1990) chama **princípio de independência**. Como a análise de um diagnóstico envolve graus de subjetividade, FIESCHI (1990) afirma que quando os resultados são acompanhados dos fatores de certeza, explicações e coeficientes estatísticos, são recebidos com mais confiança pelo usuário. O confronto dos resultados com outras bases de conhecimento na mesma área pode ser encarado como uma abordagem complementar.

Em seus trabalhos FIESCHI (1990) diz que o perfil dos especialistas também influencia na definição do SE, já que os médicos mais novos demonstram mais expectativas. Observou também que se o SE introduz mudanças mínimas nas rotinas clínicas, ele é mais bem aceito.

### 3.4- Proposta de Rossi-Mori

ROSSI-MORI (1990) afirma que o processo de avaliação não pode se limitar ao estágio de comparação da resposta final com um "padrão ouro" dado por especialistas, se este existir. O conhecimento médico não é estático e freqüentemente especialistas discutem sobre novos métodos apresentando várias soluções plausíveis. Defende a importância da definição clara dos itens de avaliação para estabelecer guias para o próprio projeto do SE. Apresenta um ciclo de decisão, acompanhando todo o desenvolvimento de um sistema, partindo do geral para o específico. Este ciclo é dividido em:

#### 1- Definição do problema

É a etapa em que o campo de atuação do sistema é definido. É indispensável a presença do clínico para relacionar os jargões médicos com a estrutura da interface, definir os limites do sistema e montar o primeiro conjunto de hipóteses. Um protótipo deve ser construído sendo capaz de:

- 1- oferecer respostas instantâneas e compatíveis com o processo mental;
- 2- apresentar um conhecimento geral sobre o assunto;
- 3- ter habilidade para manejar situações que não estão completamente definidas;

#### 2- Exploração das hipóteses

Assim como FIESCHI (1990), sugere duas verificações: estática e dinâmica. A construção da base de conhecimentos é feita de acordo com o conhecimento adquirido dos especialistas. Como nem todos os procedimentos clínicos são únicos e exclusivos, para o sistema ser considerado como apoio à tomada de decisão, deve-se fazer uma distinção dos casos para teste entre típicos e atípicos.

O processo de avaliação deve centralizar a atenção nos casos em que o usuário não terá uma resposta simples (trivial). Nos casos considerados típicos (mais frequente) é possível obter uma padronização através de um grupo de especialistas, e nos atípicos, quando não há um padrão formal, sugere-se que se use a técnica de *brain storm*, onde

cada especialista apresente sua idéia. Neste caso, o SE deverá informar ao usuário que o caso em questão é atípico e apresentar as possíveis alternativas, deixando que este faça a escolha final.

### **3- Impacto nos tratamento de saúde**

A partir da implementação de um SE é possível explorar mais interações com o ambiente de trabalho, avaliar os "efeitos colaterais", obtidos diariamente no tratamento com o paciente, para promover o melhor atendimento. Um aspecto importante na difusão dos SE é que estes contribuem para o progresso da ciência médica e para a melhoria do atendimento dos pacientes, aumentando os bancos de dados, estabelecendo estratégias explícitas para as situações típicas, trazendo um conforto ao clínico permitindo o acesso rápido a um determinado conhecimento médico.

## **3.5- Proposta de Nykänen**

NYKÄNEN (1991) diz que o único modo de obter a aceitação completa do usuário é através da avaliação formalizada do SE, e faz uma distinção entre os termos verificação, validação e avaliação:

**Verificação:** demonstra a consistência, o quanto a base é completa.

**Validação:** determina os acertos das soluções encontradas com respeito as necessidades do usuário e seus requisitos. Se refere também à dimensão do especialista.

**Avaliação:** refere-se ao julgamento do sistema, se é benéfico, se faz o que foi solicitado. É a determinação da importância, eficiência e o impacto do SE no ambiente.

NYKÄNEN (1991) enfatiza a avaliação como o aspecto mais complexo, sendo uma interação natural e deve acontecer nos diferentes estágios do desenvolvimento, incluindo a

aquisição do conhecimento. O sistema deve ser avaliado como uma parte do ambiente (condições do organismo) e não como uma parte separada.

Deste modo, mostra uma metodologia que deve cobrir 3 aspectos:

### **1 - Elicitação do conhecimento**

É a fase de modelagem do conhecimento. Devem ser consideradas as seguintes questões:

- 1- Qual conhecimento deve ser elicitado?
- 2- Como o conhecimento deve ser organizado?
- 3- Como o conhecimento pode ser validado?

Nesta fase o especialista externa seu conhecimento e o projetista tenta elicitá-lo o que é relevante, descobrindo os elementos chave do domínio e o método usado pelo especialista, resultando um modelo conceitual. A validação desse modelo abrange a qualidade do conhecimento e seu modo de integração com os itens de um problema proposto.

### **2- Modelo de desenvolvimento de ciclo de vida**

Neste aspecto NYKÄNEM (1991) critica o tempo utilizado para se alcançar todos os passos da escala de GASCHNIG. A abordagem do ciclo de vida para a avaliação é dividida em 4 fases:

- 1- Exploração preliminar da especificação do sistema : é uma conferência baseada nos objetivos e requisitos funcionais documentados pelo usuário
- 2- Validação do sistema na prática: significa julgar os resultados obtidos. Pode ser feita pelo usuário final ou pela equipe de desenvolvimento, que é responsável pela manutenção e atualização. Este desempenho está associado ao conhecimento dos especialistas
- 3- Funcionalidade e utilização: o foco principal está na interface com o usuário, na interação homem-máquina. Deve-se observar se:

- as opções desejadas foram implementadas;
  - o *lay-out* da tela de apresentação é explicativo;
  - o diálogo com o sistema é aceitável;
  - as respostas são consistentes, bem descritas;
  - o tempo de resposta é satisfatório;
  - os limites são definidos;
  - como se comporta numa situação que está fora dos seus conhecimentos;
  - o sistema pode ser usado em outros ambientes da mesma área sem sofrer grandes modificações.
- 4- Impacto do sistema: avalia os efeitos que tem sobre os usuários. A questão é saber se o uso do sistema melhorou a qualidade do trabalho, através da observação de longa duração. Verifica a relação custo/benefício, o gerenciamento do serviço de saúde, a educação e treino dos operadores, avanços em pesquisa médica e os aspectos legais e éticos, tempo de consulta e a comparação do número e tipo de exames pedidos antes e depois da implementação do SE.

### **3- Sistema como uma parte integrada do ambiente**

Deve-se ter um entendimento da estrutura onde o sistema deve ser integrado. Observa-se se o sistema pode efetivamente ajudar o usuário. No setor clínico é importante saber se a base de conhecimento está relacionada com as condições locais da população.

Os responsáveis pela avaliação são divididos em grupos (NYKÄNEM, 1991):

#### **1- Equipe de desenvolvimento**

Responsável pelos testes básicos de desempenho. Deve apresentar uma documentação objetiva da construção do projeto. Os problemas comuns neste grupo é a falta de tempo e uma certa dificuldade para detectar as falhas do sistema em funcionamento.

Dentro desta equipe, se inclui o gerente de projeto que deve avaliar o processo de desenvolvimento através da administração e pesquisas de consumo e perspectivas.

## 2- Usuário

Deve avaliar o sistema antes de aceitá-lo em uso. É de sua responsabilidade introduzir o sistema na prática clínica e mostrar de que forma o sistema em particular vai atender suas necessidades. Sugerem-se as perguntas:

- O que desejo alcançar com o sistema?
- Qual a relação custo/benefício?
- Como o SE engloba nossos requisitos?
- Quais as consequências de utilizar o sistema?

## 3- Especialista

Tem seu papel na decisão da qualidade do sistema e na liberação para uso.

## 4- Terceiros

É um grupo formado por clínicos que não estão envolvidos nem no desenvolvimento nem no local de implementação. Uma das dificuldades deste grupo, é padronizar uma resposta certa (*Gold Standard*), pois estes podem ter linhas diferentes de pensamentos.

### 3.5- Proposta de Clarke

CLARKE (1994) apresenta uma proposta para avaliação baseada no modelo de ciclo de vida proposto por NYKÄNEN (1991), porém enfatizando a necessidade de *feedback*, num processo contínuo de interações com os especialistas, engenheiros de conhecimento e usuários. Através do ciclo de vida, a razão, o tipo e a consequência de uma falha apontada no processo de avaliação/construção do SE são identificados e modificados. Essa metodologia é dividida em 4

fases sendo guiada também para determinar quando a mudança de tecnologia e mudanças na prática médica podem causar alterações nos resultados como mostra a figura 4.

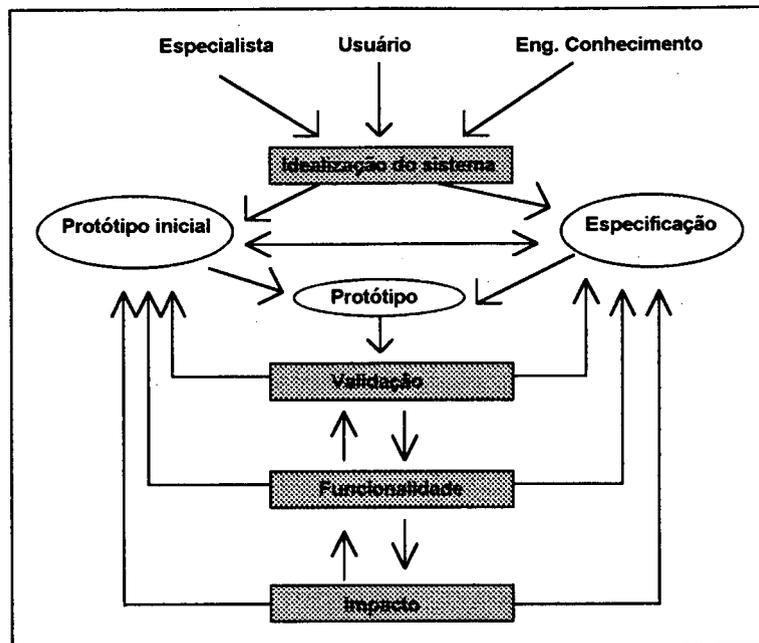


Figura 4. Avaliação por fases (CLARKE, 1994)

### Fase I - Idealização do sistema

As linhas mestre nesta etapa são:

- 1- Existe uma necessidade para o sistema?
- 2- Qual o papel do sistema em todo o empreendimento?
- 3- Qual o tipo de usuário que o sistema exige?

Depois da idealização do sistema, duas abordagens para definição do sistema têm lugar: o protótipo inicial e a especificação detalhada dos requisitos. O protótipo é o resultado dessas duas abordagens, representando um progresso na formalização da estrutura de avaliação, permitindo que se veja a funcionalidade do sistema no seu ambiente e o impacto que este causa.

## **Fase II- Validade do Sistema**

Esta fase pode ser dividida em mais 4 etapas:

**1- Validade estrutural:** faz a validade da base de conhecimento envolvendo testes para decidir se a base satisfaz as condições de consistência. Isto pode ser feito rodando simples casos ou usando ferramentas projetadas para validar a base de conhecimento do sistema. As propriedades do mecanismo de inferência do domínio particular de um problema também deve ser avaliada.

**2- Validade semântica:** está relacionada com a aplicação do conhecimento e tenta medir o grau da confiança da base. Um padrão deve ser definido para servir de comparação com a atuação do sistema. Existem dificuldades quando não se encontra o melhor padrão, tendo que definir o que é aceitável.

**3- Funcionalidade do sistema:** refere-se a avaliação do sistema já no ambiente do usuário. As interações homem-máquina são de crucial importância, determinando quando o usuário é capaz de operar o SE corretamente e com confiança. Para facilitar a aceitação do sistema é necessário que este seja de fácil uso e que não se exija muito aprendizado especial (EDMONDS, 1990). O tempo de demanda da interação humana deve ser minimizado e o ambiente de trabalho não deve sofrer muita alteração com a implementação do sistema. O usuário deve estar ciente dos limites do conhecimento do sistema. O *lay-out* das telas deve ser de fácil visualização, sendo bem próximo da versão final. O sistema obtém um alto grau de confiança quando é aplicado em outro local de domínio semelhante, apresentando um bom desempenho.

**4- Impacto do sistema:** O SE permite mudança de atitude em toda a organização do ambiente de trabalho. O campo de ação pode ser positivo, como uma aplicação de atualização educacional permanente, ou negativo, gerando constantes críticas. A relação custo/benefício deve analisar hardware, software, tempo de aprendizagem do usuário e a

manutenção para confrontar com a possibilidade de recuperação rápida e precisa de dados e melhora na eficiência ao atendimento aos pacientes. Os valores monetários, éticos e legais também são enfocados.

CLARKE (1994) conclui seu trabalho dizendo que essas 4 fases da metodologia não podem ser apresentadas isoladamente, pois a avaliação deve acompanhar todo o desenvolvimento do sistema, evitando erros futuros e economizando tempo e esforço para a equipe de desenvolvimento.

### 3.6- Proposta para Avaliação de Sistemas Conexionistas

Os Sistemas Conexionistas, especialmente as Redes Neurais, representam um desafio maior para avaliação, pois apesar de terem um grande potencial de representar funções nem sempre fornecem informação para explicar seu raciocínio, sendo chamados então de sistemas caixa-preta (HART & WYATT, 1990; WYATT, 1991a; GALLANT, 1993). A base de conhecimento desses sistemas é representada matematicamente e de difícil entendimento. A questão fundamental é como decidir entre exatidão (*accuracy*) e pouca informação. Para a avaliação dos sistemas caixa-preta, o critério se concentra na qualidade do treinamento e na aceitação do desempenho. As questões sobre transparência como "A representação do conhecimento é adequada?" não se faz a um sistema caixa-preta, e os avaliadores devem ter a noção de que nenhum teste de dados é capaz de cobrir todos os casos relevantes.

O difícil é saber quando parar de testar as Redes Neurais, sendo comum uma rede funcionar bem num conjunto de dados de treinamento e funcionar mal num conjunto de testes (HART, 1990). WYATT (1991a) propõe que para atingir uma alta qualidade é preciso usar os critérios:

- 1- Repetibilidade: um mesmo caso é apresentado mais de uma vez com intervalos de tempo para um mesmo especialista, e observar se as soluções apresentam uma variação mínima.
- 2- Usuário compatível: os dados devem ser obtidos num ambiente próprio, por usuários que serão os eventuais usuários do sistema.
- 3- Representação da população: os dados devem garantir a abrangência de uma população bem definida apresentando uma distribuição correta de critérios.
- 4- Qualidade no "padrão ouro": é mais perigoso um erro na definição do "padrão ouro" do que um erro nos dados. A classificação de um padrão deve ser baseada em conjuntos de dados e não apenas na opinião de um especialista..
- 5- Exatidão: sempre que possível todo o caso deve ser gravado para a base se tornar mais eficiente. Deve-se cuidar para que os dados sejam os mais exatos possíveis.

Um grande problema é a dificuldade de trocar dados entre centros diferentes e os casos com muitas variações. Para um sistema caixa-preta é essencial saber as condições sobre as quais os dados foram coletados. Mas mesmo assim, muitos pesquisadores acham que os testes de treinamento devem ser coletados em centros de estudos diferentes dos de desenvolvimento.

As conclusões da avaliação dos usuários e especialistas dos sistemas caixa-preta podem ser divididas em 3 categorias:

- 1 - Rejeição: quando a solução apresentada é totalmente inaceitável para o clínico.
- 2- Precaução: quando o sistema atingiu um nível de desempenho melhor do que os sistemas que apresentam transparências, para um conjunto específico de dados.
- 3- Aceitação: quando o sistema apresenta um bom desempenho para qualquer conjunto de dados.

O ideal para o apoio à decisão médica é construir modelos transparentes que possam prover os usuários de explicações (WYATT, 1991a).

Algumas equipes de pesquisa argumentam que se o sistema caixa-preta contém um certo nível de performance, é porque já está embutida um certo nível de estatística básica associado a base de conhecimento, o que já pode ser definido como uma avaliação durante a fase de desenvolvimento.

A comunidade médica e informática ainda está discutindo a avaliação para os sistemas caixa-preta com mais critérios, que conseqüentemente beneficiará também à avaliação dos SE.

Deve-se ressaltar que outras formas de metodologia também são utilizadas para desenvolvimento de SE como por exemplo o KADS, mas neste trabalho o enfoque primordial foi para metodologias de desenvolvimento voltadas para avaliação dos SE para apoio a tomada de decisão médica.

# Capítulo 4- Proposta de uma Metodologia para Avaliação de SE na Área Médica

## 4.1- Justificativa da Proposta

Para formalizar uma proposta de metodologia para avaliação de SE na área médica, especialmente os desenvolvidos pela área de Engenharia Biomédica, analisou-se as questões sugeridas por WYATT (1991):

### **1 - Somos capazes de avaliar nossos produtos?**

A área de Engenharia Biomédica é formada basicamente por engenheiros que são responsáveis pelo planejamento, organização e desenvolvimento de SE para área médica. Este grupo representa o engenheiro de conhecimento (LAWRENCE, 1992, KELSEY, 1991, RICH, 1993). Se os projetos desenvolvidos fossem direcionados para áreas determinísticas, como por exemplo, algumas áreas da engenharia, com certeza a Engenharia Biomédica teria condições de oferecer informações adicionais sobre a qualidade dos seus produtos. Mas o principal objetivo da área é desenvolver trabalhos conjuntamente e para a área médica. Sendo assim, os engenheiros de conhecimento precisam da avaliação do grupo médico para que seus projetos obtenham um crédito científico na classe médica. O grupo de engenheiros é capaz de fazer a verificação ou teste estático do produto, ou seja, se responsabiliza apenas pelos erros de lógica, de programação ou erros no acesso à base de conhecimento (NIKÄNEM, 1991, FIESCHI, 1991).

### **2- Quando deve ser feita a avaliação?**

SHORTLIFFE (1993) afirma que para se obter um ótimo desempenho num SE deve-se começar a avaliação desde a fase inicial de desenvolvimento do projeto. Como a Engenharia Biomédica trabalha conjuntamente com a parte médica, esta sugestão deve ser aplicada na prática. Como o tempo de desenvolvimento é geralmente longo, em torno de dois anos, é fundamental que as duas partes tenham avaliações constantes com retorno para a parte médica e de desenvolvimento, verificando se os objetivos iniciais estão realmente sendo atingidos.

Como o projeto de desenvolvimento pode ser dividido por fases, o nível das avaliações difere para cada fase. Para um nível inicial, observa-se o problema de uma forma geral, mas já com a indicação de uma solução correta. No processo de construção do sistema, avaliações também são necessárias, mas concentram-se nas ramificações dos problemas, indicando áreas que necessitam de refinamento, detalhamento. Quanto mais próximo do final do projeto, mais subjetiva fica a avaliação, concentrando-se mais nos aspectos do conhecimento. Os custos de uma avaliação podem ser mais controlados se integrados ao processo de desenvolvimento, contudo deve-se ter a idéia clara de que a avaliação está diretamente associada :

- 1- ao valor do sistema para o usuário;
- 2- aos riscos se for feita uma avaliação precária.

### 3- Porque se preocupar com a avaliação?

Sistemas de apoio a decisão médica são projetados para auxiliar o especialista humano, mostrando alternativas a serem utilizadas em situações complexas como fazer um diagnóstico ou planejamento de terapias. Como a tomada de decisão humana ainda não é um processo bem entendido, alguns usuários, tem um sentimento negativo sobre o suporte computacional para a tomada de decisão. Assim, o único meio de remover este sentimento é obter aceitação do usuário através de uma avaliação formalizada. O mais importante para esses sistemas é avaliar o tipo de resposta que o sistema oferece, e as questões sobre os recursos computacionais devem ser analisadas com menos rigor. Os riscos de se aceitar um resposta do SE sendo esta uma resposta errada também deve ser levado em consideração (WHITEBECK, 1983), como mostra a tabela 3:

Resultado da Avaliação	Estados do SE	
	Sistema é válido	Sistema inválido
Aceita-se a validação	Decisão correta	Erro. O usuário está com um sistema incorreto.
Declarado não válido	Erro. Os testes não foram bem aplicados.	Decisão correta.

Tabela 3. Comparando os resultados da avaliação com o SE.

Como os produtos da Engenharia Biomédica são SE para a área médica, uma vez terminado seu desenvolvimento, sente-se a necessidade de colocá-los em prática, mas porém devem passar por uma avaliação criteriosa.

#### **4 - Como avaliar?**

Se o SE deve trabalhar no apoio a tomada de decisão como um especialista humano, a primeira coisa a fazer é comparar a resposta do SE com um especialista humano. Essas comparações podem ser cruzadas ou cegas, respectivamente o especialista sabe que suas respostas estão sendo comparadas com as de um SE, ou não sabe. Este é um item muito delicado pois nem todo especialista está disponível para testes, sendo o tempo um fator importante na rotina médica. Os especialistas sentem dificuldade em interromper suas atividades para ficar testando por muito tempo os SE. Também é difícil determinar quando deve-se interromper os testes, formando uma opinião sobre o desempenho do sistema. Um outro fator subjetivo que representa mais um obstáculo para a avaliação dos SE, é que, assim como qualquer profissional de qualquer área, alguns médicos sentem-se constrangidos para fazer comparações de seus diagnósticos com as de outro profissional, ainda mais quando este "outro profissional" é um sistema computacional (WYATT, 1991; SHORTLIFFE, 1991; BERNER, 1994).

Se considerarmos todas as propostas existentes, verifica-se que a que reúne os principais itens e que pode ser facilmente implementada é a avaliação por protótipo que vem acompanhando o desenvolvimento do produto. A proposta de CLARKE é a mais atual e apresenta uma abordagem deste tipo abrangendo muitas contribuições de outros autores, que já vêm discutindo a necessidade da avaliação há algum tempo. A idéia é que se o produto está sendo constantemente avaliado desde sua fase de planejamento, o resultado final será um produto confiável.

Pela experiência do GPEB, percebe-se que desde o primeiro contato com o especialista que existe uma certa expectativa em relação ao produto final. Como o tempo de desenvolvimento

é longo, nota-se que alguns especialistas perdem o entusiasmo porque não é apresentado nenhum resultado parcial. Este é mais um motivo para se implementar a avaliação por fases, trazendo o especialista mais vezes para o laboratório de desenvolvimento, onde este começa a sentir as dificuldades de formalização do produto.

Mesmo estando localizado dentro do Hospital Universitário (HU) da UFSC são poucos os médicos que participam efetivamente no processo de avaliação. Percebe-se que os SE desenvolvidos no GPEB estão sendo visto apenas como projetos acadêmicos, talvez porque depois do primeiro contato, leva-se muito tempo para dar um retorno ao especialista. Com a idéia de avaliação por protótipo, provavelmente, os projetos serão mais aceitos para serem utilizados na prática, pois o especialista estará participando de todas as etapas de desenvolvimento, trabalhando junto com o engenheiro de conhecimento, tornando familiar as terminologias, rotinas diárias e preocupações de ambos os grupos. A avaliação por protótipo também se apresenta como o caminho mais viável para direcionar pesquisas *fuzzy* e problemas mal-definidos.

## 4.2- Detalhamento da Proposta

A proposta de avaliação para o GPEB apresenta três fases como mostra a figura 5:

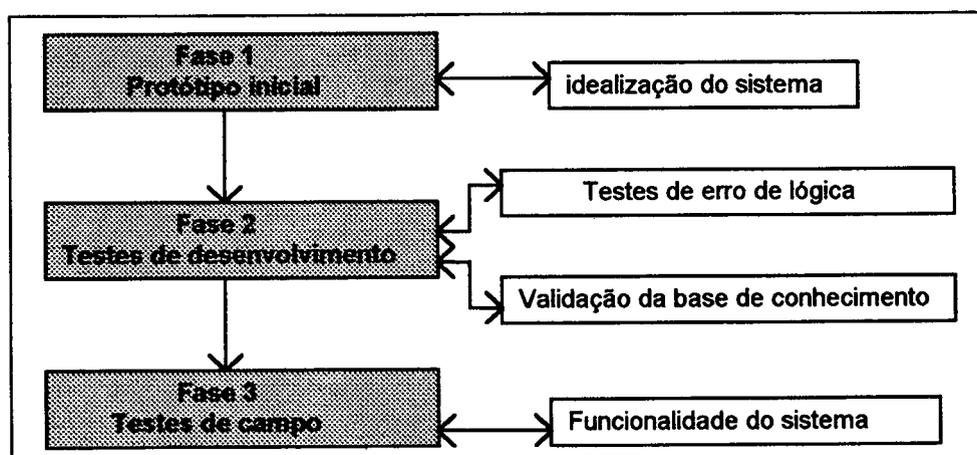


Figura 5. Proposta de avaliação.

### **Fase 1- Protótipo inicial**

É a definição do projeto. Nesta etapa deve-se ter com exatidão o que se deseja alcançar, qual o produto resultante no final do desenvolvimento. Como esta etapa inicial exige uma familiarização entre as duas áreas, o desenvolvimento de um protótipo à curto prazo ajuda o especialista a ter uma visão melhor de todos os recursos computacionais que podem ser utilizados, facilitando a tarefa de especificar o projeto. É importante ter documentado qual o objetivo final, tendo a dimensão exata do domínio do SE, ou seja, o que ele se propõe a fazer. Este protótipo não precisa ser único, podendo ser dois ou três, ou até que se tenha uma visão completa dos objetivos que se deseja alcançar. Da avaliação desta fase pode surgir um documento formal com as especificações dos requisitos, tendo um noção real da necessidade do sistema, quais os benefícios que o sistema trará quando implementado e o quem utilizará o sistema.

Esta etapa equivale aos dois primeiros passos da escala de GASCHINIG, ao estágio conceitual da proposta de LUNDSGAARDE, a etapa de definição do sistema de HART & WYATT, a etapa de definição do problema de ROSSI-MORI, a fase de elicitação do conhecimento de NYKÄNEM, e a fase de idealização do sistema de CLARKE.

### **Fase 2- Testes de desenvolvimento**

Uma vez definido o que se pretende alcançar, o engenheiro de conhecimento começa a estruturar o projeto com uma visão mais detalhada. Cada item do protótipo deve ser refinado para ser implementado. A base de conhecimento já começa a ser montada. Nesta fase a avaliação pode ser dividida em duas etapas:

#### **i) Testes de erro lógica.**

Esses testes são feitos basicamente pelos engenheiros de conhecimento, ou pela equipe de desenvolvimento do projeto. São detectados erros de lógica de programação, erros de acesso à

informação, lógica de operação, *lay-out* das telas, sequência das telas de operação, robustez do sistema, dentre outros que se refiram exclusivamente a parte computacional. Normalmente, esta fase ocorre em paralelo com o desenvolvimento. O ideal é que estes tipos de erros não aconteçam quando se começa a testar a base de conhecimento, mas nem sempre isto ocorre.

## ii) Validação da base de conhecimento

Para validar o SE é necessário que o especialista trabalhe junto com o engenheiro de conhecimento, pois o conhecimento médico precisa ser consolidado. São feitos testes para consistência e abrangência, detectando contradições, duplicação e redundância das regras, regras circulares e condições irrelevantes. É importante que o especialista participe ativamente desta fase, pois a base de conhecimento deve estar formalizada de acordo com o raciocínio clínico. Os especialistas também podem chamar um provável usuário do sistema para se certificar que as soluções apresentadas são facilmente entendidas.

Nesta etapa da avaliação, deve-se observar se os requisitos definidos estão sendo atendidos, ou seja qual a extensão do projeto. Se for necessário partir para novas definições deve-se recomeçar todo o processo. Alguns itens merecem especial atenção nesta fase:

### a) Montagem do conjunto de testes para a base de conhecimento

Recomenda-se que para testar a base de conhecimento, seja selecionado casos que apresentem uma certa dificuldade, incluindo casos atípicos, de pacientes com doenças raras, ou situações complexas com muitas variáveis. A montagem deste conjunto de testes deve ser feita por outros especialistas que não contruíram na construção do sistema para evitar vícios o que FIESCHI (1990) chama de **princípio de independência**. Uma boa alternativa é a proposta de BERNER (1994) que sugere a formação de uma equipe de especialistas que não tenham participado do desenvolvimento, que comporiam um conjunto representativo de testes.

**b) Entrada de dados**

Nesta etapa a entrada de dados é avaliada, e o especialista deve observar se as informações pedidas pelo sistema são suficientes para uma busca correta na base de conhecimento. É bom lembrar que especialistas mais experientes necessitam de menos informações para buscar uma solução do que médicos menos experientes na área.

**c) Soluções apresentadas**

Além de procedimentos corretos, deve-se ter em mente que o usuário sempre apresenta um nível de expectativa em relação ao SE, e que se decepciona se o sistema não for capaz de realizar sugestões relevantes para um caso. Uma resposta óbvia não invalida um sistema, mas também não o qualifica como tendo um bom desempenho. Para as soluções apresentadas, de forma geral, deve-se observar as seguintes características:

- 1- Eficácia, ou a capacidade de produzir os resultados pretendidos e desejados.
- 2- Confiabilidade, ou a capacidade de produzir os resultados pretendidos e desejados sem erros.
- 3- Exatidão, ou a capacidade de lidar com o grau de precisão, em representação numérica, e correção, nos cálculos, que a tarefa em questão requerer.
- 4- Consistência, ou a capacidade de sempre produzir os mesmos resultados quando os mesmos dados são processados.
- 5- Eficiência, ou a capacidade de operar sem problemas, exigindo o mínimo de recursos para produzir os resultados pretendidos e desejados.
- 6- Transparência, ou a capacidade de refletir, apropriadamente e sem distorções, o mundo real através do modelo conceitual adotado.
- 7- Segurança, ou a capacidade de preservar a integridade dos dados na eventualidade de falha de hardware, erro ou eventos inesperados.
- 8- Expansabilidade, ou a capacidade de absorver mais funções ou maior quantidade de dados, sem necessidade de alterações estruturais.

#### **d) Interface**

O apelo visual de um sistema, facilidade de operação, e uma sequência de entrada de dados conveniente para o usuário são pontos fundamentais para a aceitação de um SE. FIESCHI (1990) afirma que se o SE introduz mudanças mínimas nas rotinas clínicas, ele é bem mais aceito. Como características da interface a serem observadas destaca-se:

- 1- Facilidade de instalação.
- 2- Facilidade de uso e operação, incluindo opções para interatividade, orientação para objetos, utilização de mouse, menus, janelas e ajuda *on-line* (*help* interativo), dentre outros.
- 3- Capacidade de lidar com monitores de alta resolução e em cores, quando essas características se fazem necessárias ou recomendáveis.
- 4- Documentação correta e de excelente qualidade, tanto em termos técnicos como na apresentação visual, e com exemplos relacionados ao campo de aplicação.

Esta fase corresponde aos passos 3, 4 e 5 da escala de GASCHNIG, ao estágio intermediário de LUNDSGAARDE, ao teste laboratorial de HART & WYATT, a exploração das hipóteses de ROSSI-MORI, ao modelo de desenvolvimento de ciclo de vida de NYKÄNEM e a fase de validade do sistema de CLARKE.

#### **Fase 3 - Testes de Campo**

Esta fase compreende a avaliação da funcionalidade do sistema, seu impacto no ambiente de trabalho. Significa avaliar os efeitos que o sistema tem sobre seus usuários, seus comportamentos, e todo processo de cuidado a saúde. A questão é se o uso do sistema provoca melhorias no trabalho do usuário e em toda sua organização. Esta fase é a que requer mais tempo, pois as características observáveis são a qualidade sobre cuidado com a saúde, relações de custo benefício que só podem ser avaliadas durante a rotina de uso do sistema. Os termos que podem ser usados para avaliar o impacto do sistema para área da saúde podem ser: provisão de

cuidado clínico para um único paciente, rotina de gerenciamento de serviço de saúde e educação e treinamento dos operadores de saúde e avanços em pesquisas médicas.

É importante que o usuário do sistema tenha um conceito exato das particularidades do sistema, como:

- 1- Características sociais, culturais e políticas que podem torná-lo tendencioso, especialmente no caso de bases de conhecimento e gráficos ou desenhos utilizados para finalidades educacionais.
- 2- Pressuposições acerca da possibilidade e praticabilidade de acesso a locais, equipamentos e recursos, que podem não ser os mais apropriados para a boa utilização do software.
- 3- Descrição do contexto em que se espera que o software seja utilizado, se em um hospital, numa clínica especializada, um centro de saúde, um consultório, etc.
- 4- Especificação acerca dos principais beneficiários do software, como médicos, atendentes em centros de saúde, pesquisadores, professores, e também, quando necessário, indicação acerca dos prováveis operadores do sistema, se profissionais da saúde, pessoal administrativo, etc.

É fundamental que a manutenção do sistema seja eficiente, pois além de demonstrar interesse no funcionamento do sistema, no teste de campo o usuário final contruibui com muitas informações sobre melhoria para o atual sistema e ainda ajuda com idéias para um projeto futuro. Como o GPEB funciona no local de implementação dos seus projetos, no HU, com certeza esta tarefa de amparar o usuário, sempre que necessário, será realizada sem dificuldade.

É difícil determinar com exatidão quanto tempo é necessário para esta fase, pois o que realmente importa é o volume de trabalho, o quanto o SE foi utilizado. Se o SE apresentar um desempenho satisfatório nesta fase, pode-se reunir um grupo de clínicos, chamado de terceiros por NYKÄNEM (1991), que não estão envolvidos nem com o desenvolvimento nem com o local

de implementação para chegar a um consenso sobre o desempenho do SE, podendo então ser difundido entre outros centros que tenham um trabalho semelhante.

Esta fase corresponde aos passos 7, 8 e 9 da escala de GASCHNIG, ao estágio de maturidade da proposta de LUNDSGAARDE, ao estudo de campo de HART & WYATT, ao impacto nos tratamentos de saúde de ROSSI-MORI, ao sistema como uma parte integrada do ambiente de NYKÄNEM e ao impacto do sistema de CLARKE.

### 4.3 - Verificação da Proposta num SE Desenvolvido pelo GPEB

Os projetos desenvolvidos pelo GPEB ainda não apresentam uma metodologia formalizada para a avaliação. De um modo geral, os projetos chegam até a fase de validação da base de conhecimento (Fase 2), como por exemplo o projeto Provanes (GARCIA, 92), que chegou a ser testado por especialistas que não participaram do desenvolvimento do sistema, mas as sugestões que foram dadas ainda não foram implementadas. A montagem da base de dados se restringiu a um único especialista o que deixou a explanação das soluções encontradas muito restritas a uma única abordagem do problema. Mesmo assim, foi feita uma avaliação a nível laboratorial, com a participação de 6 alunos residentes na especialidade de anestesiologia e 6 médicos especialistas em anestesiologia. Depois de testes com 124 pacientes, os usuários verificaram mais de 90% de respostas certas para o sistema. Foram encontradas 2 contra-indicações absolutas de pacientes classificados com alguma característica de limitação. Mas o sistema fica restrito ao conhecimento do especialista que participou do desenvolvimento do sistema. Depois desta avaliação, o sistema não foi re-implementado para novos testes.

No projeto Sistema Inteligente para Treinamento do Controle de Infecção Hospitalar (Ramirez, 1991), pode-se dizer que os testes para a mesma fase, nem saíram do próprio ambiente de desenvolvimento. Para a validação deste sistema, foram chamados voluntários da área da

saúde, que demonstraram interesse pelo sistema desenvolvido. A validação ocorreu sob a forma de uso do sistema, onde estes usuários apenas testavam a interface de entrada de dados. A avaliação não se preocupou com a qualidade das respostas oferecidas pelo sistema. A montagem do banco de dados foi feita de modo aleatório, com dados de arquivos passados. Cada resposta oferecida pelo sistema, foi confrontada com resultados fornecidos pela comissão de controle de infecção hospitalar. Não foi estabelecido um percentual de acertos do sistema. Este tipo de avaliação não traz nenhuma garantia para o usuário, pois não se tem nada documentado, mostrando qual o percentual de acerto, nem a opinião dos usuários. Seria necessário que um especialista da área desse o seu parecer sobre o sistema para dar mais credibilidade ao sistema. Como não houve um uso contínuo do sistema, não houve um interesse maior para modificá-lo.

O Sistema Inteligente de Auxílio ao Ensino em Traumatologia Crânio-Encefálica foi desenvolvido para possibilitar o treinamento de residentes para atender pacientes em ambiente hospitalares (NIEVOLA, 95), teve uma avaliação mais voltada para a base de conhecimento que foi montada com casos exemplos obtidos a partir de fichas de pacientes atendidos no Hospital Celso Ramos, situado em Florianópolis. No relatório do desenvolvimento deste sistema, não é mencionado como a avaliação foi feita, nem a qualificação do usuário, mas ressalta-se que em média, o sistema apresentou 90% de acertos para todas as perguntas formuladas.

O mais recente projeto desenvolvido no GPEB foi o sistema especialista híbrido em anestesiologia para paciente crítico/problemático (PASSOLD, 1995), que é uma expansão do sistema Provanes, um sistema para proposta e avaliação de procedimentos anestésicos (GARCIA, 1992). Por ter sido o mais recente, apresenta-se uma análise através das fases propostas para avaliação para este SE.

### **Fase 1 - Protótipo inicial**

Por ser expansão do sistema Provanes, esta fase de definição foi bem detalhada, pois já se tinha um sistema em funcionamento. Os especialistas responsáveis já haviam trabalhado no

desenvolvimento do Provanes, o que contribuiu para facilitar o entendimento entre os dois grupos. A proposta inicial de desenvolvimento foi formalizada e apresentava os seguintes itens:

- 1- definição de paciente crítico/problemático
- 2- montagem da base de dados anestesiológicos para pacientes críticos/problemáticos
- 3- análise do sistema original
- 4- montagem dos conjuntos de treino e teste para as redes neurais
- 5- integração das redes ao sistema original
- 6- verificação do sistema expandido

### **Fase 2- Validação da base de conhecimento**

A etapa de testes de erros de lógica, foi realizada paralelamente com o desenvolvimento do sistema. A verificação lógica da estrutura semântica da base de conhecimento permitiu a correção dos erros de lógica do programa e detecção de conflitos e duplicação nas regras.

A etapa de validação da base de conhecimento também foi realizada, embora tenha sido por pouco tempo. A maior dificuldade encontrada nesta fase foi a falta de tempo dos especialistas para validar a base.

Respeitando o princípio da independência, foram chamados outros especialistas para contribuírem nesta etapa da avaliação. Talvez por não terem acompanhado o trabalho de desenvolvimento, os especialistas de um modo geral, acharam a tarefa de avaliar os resultados complexa, devido a não existência de um "padrão ouro" para a área de anestesiologia.

Seguindo a proposta de HART (1990), o engenheiro de conhecimento formulou algumas perguntas para que guiassem os especialistas:

- 1- Existe uma ordem de importância na sequência de entrada dos dados para a solução do problema?
- 2- Existe alguma limitação na entrada de dados que impeça o SE de apresentar uma solução?

- 3- As soluções apresentadas são relevantes?
- 4- As soluções apresentadas são completas?
- 5- Qual sua sugestão para melhorar o SE?
- 6- Este sistema é realmente utilizável, trazendo benefícios para os usuários?

Comparando com a proposta de HART & WYATT (1990), as perguntas formuladas se referem ao processo (1,2), à estrutura (3,4) e aos resultados (5,6). Para ficar mais completo poderia ter sido chamado um médico, que não fosse especialista em anestesiologia pra verificar se este compreenderia as soluções sugeridas pelo sistema.

Passando pelos demais itens observou-se:

### **1- Montagem do conjunto de teste para o sistema**

De acordo com a proposta sugerida foi montado um conjunto de teste numericamente representativo com dados do Hospital Celso Ramos e Hospital Universitário de Florianópolis (HU-UFSC). Os especialistas participaram da seleção tentando reunir uma proporção numérica representativa da maioria dos casos atendidos nesses hospitais. A montagem dos dados teve maior representatividade dos casos típicos. Seria desejável mais casos de pacientes atípicos com doenças raras, e casos complexos, onde seria exigido mais conhecimento do sistema.

### **2- Entrada de dados**

Esta parte foi bem explorada pelos especialistas, que concluíram que a ausência da temporalidade na entrada dos dados é um fator relevante para alguns procedimentos. Por exemplo, é comum durante o atendimento de um paciente em estado de emergência, colher amostras de sangue no momento do seu atendimento e minutos antes do início do procedimento cirúrgico. Esta diferença no tempo não é enfatizada pelo sistema.

### **3- Soluções apresentadas**

Os especialistas concluíram que nem sempre o sistema foi capaz de apresentar sugestões relevantes para o caso, ou seja, para alguns casos o usuário esperava respostas mais úteis do sistema. Esta observação não torna o sistema inválido, pois este foi desenvolvido tendo como objetivo a expansão do Sistema Provanes (GARCIA, 1992) para extrair e acumular conhecimento da base de dados anestesiológica, para os casos de pacientes críticos e/ou problemáticos, sendo esta tarefa cumprida.

O perfil do usuário deste sistema não é necessariamente um médico especialista em anestesiologia, podendo ser um médico residente ou estudante no final do curso. Mesmo que as respostas apresentadas não sejam consideradas relevantes pelos especialistas, o sistema apresenta respostas aceitáveis, que com certeza colaboram para o aprendizado de médicos não especialistas em anestesiologia e estudantes.

### **4- Interface com o usuário**

Os especialistas concluíram que a interface não é muito flexível, obrigando o usuário a manter uma sequência na ordem dos dados. Esta sequência é um reflexo do raciocínio do especialista que participou do desenvolvimento do sistema. Os especialistas acham melhor ter uma maior liberdade na entrada dos dados, de modo que cada especialista utilizasse a sua própria sequência de raciocínio para a entrada de dados.

É bom mencionar que apesar do usuário ter sempre razão (WYATT, 1991), alguns pontos de formalização não podem ser flexíveis demais.

### **Fase 3 - Teste de campo**

O sistema híbrido para anestesiologia ainda não foi testado efetivamente em campo. O processo de avaliação foi interrompido na fase 2.

Talvez este seja o maior problema do GPEB, pois dificilmente um projeto consegue ultrapassar a fase 2. Como a maioria dos projetos são elaborados para a aquisição do título de Mestrado ou Doutorado, depois que se termina a programação e a fase de validação para base de conhecimento, o engenheiro de conhecimento apresenta seu trabalho, atinge o grau merecido, e o projeto não chega a ser modificado de acordo com o *feedback* dos especialistas para passar para o teste de campo.

O ideal seria que o GPEB fizesse as modificações sugeridas e levasse adiante os projetos. É bom lembrar que para que isto aconteça, o especialista deve realmente demonstrar interesse e colaborar com o grupo de pesquisa, para que os esforços de ambos os grupos sejam recompensados.

## Capítulo 5- Discussão e Conclusões

Desde a década de 70 já se tem noção do potencial que se pode alcançar com o apoio computacional para tomada de decisão na prática médica, mas atualmente são poucos os artigos que descrevem as aplicações desses recursos nas clínicas médicas. Numa recente pesquisa, foi apurado que poucos sistemas estão operando efetivamente como apoio a tomada de decisão fora dos seus domínios de origem.

Os maiores problemas apontados para esta baixa utilização são as dificuldades para representar o conhecimento e o raciocínio clínico, e ainda a interação do usuário com o computador. Mas os constantes congressos sobre informática médica revelam outros parâmetros, já que cada vez mais recursos técnicos estão disponíveis. Talvez seja uma atitude muito simplista generalizar em apenas dois tópicos a não utilização rotineira do apoio computacional na prática médica para tomada de decisão, deixando de analisar outros parâmetros, como por exemplo a própria filosofia do desenvolvimento dos projetos.

Fazendo uma análise da equipe de desenvolvimento e dos usuários pode-se dizer que a maioria dos que estão envolvidos na construção de sistemas para apoio a tomada de decisão médica são acadêmicos e estudantes de mestrado e doutorado, querendo demonstrar suposições teóricas. Se o sistema chega a ser implementado, geralmente é com a intenção de avaliar o desempenho do modelo teórico, e quando isto ocorre, normalmente é negligenciado a aceitação do usuário, o desempenho do sistema, a documentação e sua manutenção. Sendo assim, não se deve estranhar o fato de que estes sistemas fracassem quando aplicados na prática médica, já que seus objetivos iniciais eram o título acadêmico ou alguma publicação científica.

Além de grande parte dos projetistas dos sistemas de apoio à tomada de decisão médica não terem experiência suficiente no desenvolvimento de produtos computacionais, muitos não fazem uma pesquisa adequada sobre as necessidades reais do usuário e do potencial que este pode

oferecer durante a construção do sistema, fracassando na identificação dos aspectos da tomada de decisão do raciocínio clínico. Normalmente, quem desenvolve não está capacitado para investigar os aspectos da tomada de decisões, e por outro lado, os clínicos são muito hábeis em apontar diagnósticos alternativos com apenas alguns itens sobre o paciente, com dificuldade em explicar pequenos detalhes da aplicação e os aspectos diferentes de um diagnóstico. Este aspecto reforça a idéia de que o especialista deve trabalhar conjuntamente com o engenheiro de conhecimento desde a fase inicial de desenvolvimento, inclusive para diminuir a dificuldade de comunicação entre ambos.

Este é um aspecto importante porque a própria dificuldade de comunicação entre os dois grupos, faz com que muitos especialistas visualizem um sistema como algo muito complexo, e acabam fracassando em apontar os objetivos a serem alcançados a partir da construção de um SE.

Fazendo uma comparação com os sistemas de gerenciamento, o projetista do sistema é capaz de identificar com relativa facilidade quais os objetivos a serem alcançados e os parâmetros mais relevantes para serem utilizados com as técnicas de gerenciamento para a construção de um sistema, mas no desenvolvimento dos SE deve-se gerenciar o conhecimento, o que implica numa tarefa mais complexa.

A representação do conhecimento médico de forma explícita em processos de decisão deve ser desenvolvida por pessoas que tenham conhecimento em medicina. Quanto mais motivados pelos resultados da pesquisa, mais tempo e esforço será dedicado ao desenvolvimento do SE.

Assim, a proposta de desenvolvimento por protótipo permite que o usuário atue mais ativamente durante todo o processo de construção obtendo um SE com uma melhor definição da lógica médica, melhor qualidade da aquisição de dados e conseqüentemente mais critérios para

arquivos de dados poderão ser desenvolvidos para aumentar a base de conhecimento para a lógica de decisão em experiências passadas.

A baixa utilização dos SE para apoio à tomada de decisão também pode ser comparada com o extensivo uso de editores de texto, planilhas de cálculo e pacotes gráficos. Todas essas ferramentas envolvem um aprendizado para novas técnicas e novas interfaces, e os usuários estão sempre dispostos a aprender a utilizá-las, tão logo percebam as vantagens reais que o domínio dessas ferramentas traz. No domínio clínico, *scanner*, sistemas de ECG implementados, instrumentos laboratoriais e sistemas gerenciais alcançam o mesmo sucesso, porque suas vantagens são amplamente divulgadas, tornando-se conhecidas pelos usuários.

Conclui-se então que para os SE serem aceitos nas rotinas clínicas, os próprios usuários devem ser os primeiros a serem convencidos sobre a segurança, exatidão e eficiência do SE. Para isto é fundamental a introdução de metodologias de avaliação como uma parte integrante de todo o processo de desenvolvimento, convencendo os clínicos sobre o valor dos SE para a tomada de decisão.

Um outro aspecto paralelo à pouca utilização dos SE é a falta de direcionamento dos objetivos dos sistemas, que no final aparecem como um aglomerado de rotinas, sem avaliação e sem estarem atreladas as reais necessidades dos médicos, sendo recebidas por estes como uma ameaça a sua autonomia.

Fazendo uma análise sobre as dificuldades para implementação dos SE, destacam-se os itens:

### **1- Preocupação com técnicas computacionais**

Grande parte dos analistas envolvidos na construção de sistemas, perde muito tempo construindo recursos computacionais, como por exemplo, incrementar tempo de resposta e otimizar memória, esquecendo que o mais importante, no SE para a área médica é o valor das

informações obtidas através do conhecimento adquirido dos especialistas, e não os novos recursos computacionais.

## **2- Fracasso em adotar modelos para análise de problemas**

Não se pode adotar um modelo de formalismo universal na área médica, pois não se trata de situações determinísticas. Cada caso é um caso novo.

## **3- Falta de uma linguagem clara para comunicação.**

Não se pode adotar padrões rígidos para um formalismo particular pois os dois grupos responsáveis pela construção do SE apresentam muitas diferenças. A linguagem utilizada para o desenvolvimento do sistema deve ser compreendida por ambos com uma certa facilidade. Simples representações gráficas ou protótipos podem ser bem úteis nesta comunicação.

## **4- Muitos especialistas ainda têm dúvida sobre os benefícios e os limites de um sistema de suporte à decisão**

O fator limitante no crescimento de sistemas de computação para uso em clínicas médicas, não é nem o software, nem o hardware. É o "medical ware". É fundamental que o especialista compreenda que os SE são planejados para ajudar humanos nas tomadas de decisão, sendo ferramentas auxiliares, e não substitutas, e que para serem implementadas no ambiente de trabalho, os SE devem ser avaliados sendo observados os aspectos desempenho, utilidade, aceitabilidade pelo usuário, impacto do sistema e seu custo-benefício.

Nem todo o avanço tecnológico consegue tirar a necessidade da avaliação no tempo e lugar do SE para a área médica, talvez porque ainda não exista uma metodologia geral, um "padrão ouro" definido que possa ser utilizado sem a intervenção do especialista.

Talvez, trabalhando através de um desenvolvimento interativo com o usuário, como a construção de protótipos e os constantes retornos, se obtenha uma avaliação mais completa,

abrindo todas as etapas do desenvolvimento, tendo a relação custo-benefício reduzida se comparada com uma única avaliação no final do projeto

Com o desenvolvimento através de protótipos poderá ser estabelecido com mais clareza os objetivos do sistema. Assim, espera-se diminuir as dificuldades encontradas pelos pesquisadores em informática médica para eliciar seus reais problemas, evitando que sejam misturados os conceitos de uma pesquisa de seu próprio interesse com a verdadeira intenção de construir um sistema de apoio à tomada de decisão.

Uma sugestão é tornar os sistemas mais práticos do que teóricos, realmente voltados para a necessidade dos usuários, determinando a natureza e a qualidade das decisões e o impacto nos pacientes. O projeto deve ser totalmente voltado para os interesses do usuário. Com o ciclo protótipo-retorno-reimplantação é facilmente atingido a aceitação do usuário, pois este estará continuamente dando *feedback* para os engenheiros de conhecimento sobre a funcionalidade do sistema. Assim a versão final emergirá naturalmente.

Desta forma espera-se que com a motivação constante entre os engenheiros de conhecimento e os especialistas surjam sistemas mais compatíveis com a realidade e conseqüentemente mais utilizados na rotina da prática médica.

Para o GPEB, esta motivação é necessária para que os projetos desenvolvidos sejam passíveis de serem transformados em produtos de utilidade no próprio HU, de onde se tira grande parte de todo conhecimento médico, através dos especialistas e de consultas às fichas de pacientes para construção de bases de dados. Os projetos desenvolvidos no GPEB só fazem sentido se realmente puderem ser aproveitados em benefício da qualidade dos serviços prestados pelo HU para a comunidade.

## Referências Bibliográficas

1. AIKINS, J., KUNZ, J., SHORTLIFE, E.H., PUFF: an expert system for interpretation of pulmonary function data. **Comput. Biomed. Res.**, v. 16, p. 199-208, 1983.
2. de AZEVEDO, F. M., BARRETO, J. M., de BAENST-VANDENBROUCKE, A., EPPRECHT, E. K., ZANCHIN, C. I., Connectionist networks: insights towards a more operational technology, **IATED - International Conference**, Artificial Intelligence Application and Neural Networks, Zurich, Switzerland, julho 1-3, 1991.
3. de AZEVEDO, F. M., LIMA, W. C., ZANCHIN, C. I., EPPRECHT, L., A neural network implementation of case base reasoning in medical diagnostic, 43º Reunião Anual da SBPC, 11º Encontro Nacional de Automática, RJ, p. 14-19 julho 1991.
4. de AZEVEDO, "Identification of arm dynamics using neural networks", 46ª Reunião Anual SBPC, Anais, p. 97, Vitória, ES, 17-22 julho 1994.
5. de AZEVEDO, F. M., BARRETO, J. M., "A neural network approach for expert systems", **World Congress on Medical Physics and Biomedical Engineering**, Proceedings, Parte I, Apstract PS17-1.18, p. 466, 21-26 august 1994.
6. de AZEVEDO, F. M., BARRETO, J. M., " IMC scheme using neural networks for robot arm", X Congresso Brasileiro de Automática - VI Congresso Latinoamericano de Controle Automático, Anais p. 891-896, Rio de Janeiro, 19-23 setembro, 1994.
7. BAINBRIDGE, D. I., Computer-aided diagnosis and negligence, **Med. Sci. Law**, vol. 31, n. 2, p. 127-136, 1991
8. BANKOWITZ, R. A., LAVE, J.R., McNEIL, M. A., A Method for assessing the impact of a computer-based decision support system on health care outcomes, **Methods of Information in Medicine**, vol. 31, p. 3-11, abril, 1992.
9. BARRETO, J. M., Neural networks learning: a new programming paradigm?, **ACM International Conference: Trends and Directions in Expert Systems**, Florida, USA, 1990.

10. BARRETO, J. M., MACHADO, M. A. S., BRAGA, M. J. F., **Conceitos de Matemática Nebulosa na Análise Risco**, Ed. Artes e Rabiscos, 1995.
11. van BEMMEL, J.H., Formalization of medical knowledge, **Meth. Inform. Med.**, v.25, p.191-193,1986.
12. BERNER, ETA S., ALGINA, J., DENNIS V., MANGALL E., Performance of four computer-based diagnostic systems, **The New England Journal of Medicine**, p. 1792 - 1796, Junho, 1994.
13. BOITEN, N.; KUSUAK, A.; RAZ, T., Knowledge base: integration, verification and partitioning. **EJOR**, v. 42, p.111-128, 1989.
14. BUCHMAN, T., KUBOS, K., SEIDLER A., A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit, **Critical Care Medicine**, vol. 22, n. 5, Maio 1994.
15. BRASIL, L.M., **Aquisição de conhecimentos aplicada ao diagnóstico de epilepsia**. Florianópolis: UFSC, 1994. 302p. Dissertação (Mestrado no Grupo de Pesquisas em Engenharia Biomédica)- Pós-graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, 1994.
16. BYLANDER, T.; CHANDRASEKARAN, B., Generic tasks for knowledge based reasoning: the "right" level of abstraction for knowledge acquisition. **IJMMS**, v.26, p.231-243,1987.
17. CAUDILL, MAUREEN, Expert Networks; An emergin technology marries expert system and neural network to provide the best of blth worlds, **Byte**, October 1991.
18. CAUDILL, MAUREEN, Kinder, gentler computing, **Byte**, vol.17, v. 4, p.135-150, Abril 1992.
19. CLANCEY, W. J., LETSINGER, R., NEOMYCIN: Reconfiguring a rule-based expert system for application to teaching. **Medicine Artificial Intelligence: The First Decade**, p. 361-381, 1984.
20. CLARKE, K., O'MOORE, R., SMEETS, R., TALMON, J., BRENDER, J., McNAIR, P., NYKANEM, P., GRIMSON, J., BARBER, B., A methodology for evaluation of

- knowledge-based systems in medicine, **Artificial Intelligence in Medicine**, vol. 6, p. 107-121, 1994.
21. CLIFFORD, P. C., CHAN, M., HEWETT, D. J., The acute abdomen: management with micro-computer aid, England, 1986, apud DOMBAL, F. T. de, Computer-aided Decision Support - The Obstacles to Progres, **Methods of Information in Medicine**, vol. 26, n. 4, p. 183-184, 1987.
  22. de DOMBAL, F.T. , Computer-aided decision support - the obstacles to progres, **Methods of Information in Medicine**, vol. 26, n. 4, p. 183-184, 1987.
  23. EDMONDS, E., Human-computer interface evalutaion: not user friendliness but design for operation, **Med. Inform.**, vol. 15, n. 3, p. 253-260, 1990.
  24. FEIGENBAUM, E. A., Computer assisted decision-making in medicine. **J. Med. Philos.**, n. 9, p. 135-160, 1988.
  25. FIESCHI, M., Towards validation of expert systems as medical decision aids, **J. Biomed Comput**, p. 93-108, Département de l'informatique Médicale, Hopital de la Conception, Marseille, France, abril, 1990.
  26. FLETCHER, R. H., FLETCHER, S. W., WAGNER, E. H. Diagnóstico, **Epidemiologia Clínica**, Trad. Maria Inês Schmidt et al., Porto Alegre: Artes Médicas, 1991.
  27. GALLANT, S.I., Connectionist Expert Systems Communication of the ACM 31, n. 2, p. 152-169, 1988.
  28. GALLANT, S. I., **Neural Network Learning and Expert Systems**, Massachussetes Institute of Technology, MIT Press, Longon, England, 1993.
  29. GARCIA, R. O., **Técnicas de inteligência artificial aplicadas ao apoio à decisão médica na especialidade de anesthesiologia**. Florianópolis: UFSC,1992. 229p. Tese (Doutorado em Ciências) - Pós-graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, 1992.
  30. GARCIA, R. O., LIMA, W. C., DUARTE, D. F, Sistema experto para la especialidad de anesthesiologia usando redes neurales, **V Congreso Chileno de Ingenieria Electrica**, Concepcion, Chile, p. 511-514, 1989.

31. GARCIA, R. O., BARRETO, J. M., Training neural networks for medical knowledge representation, **Ninth International Conference on Systems Engineering**, University of Nevada, Las Vegas, USA, Julho 1993.
32. GASCHNIG, J., KLAHR P., POPLE H., SHORTLIFFE E. H., TERRY A., Evaluation of expert systems: issues and cases studies, **Addison-Wesley**, 1983, apud LUNDSGAARDE, H.P., Evaluating medical expert systems. **Soc. Sci. Med.**, v. 24, n. 10, p.805-819, 1987.
33. HABBEMA, J. D. F., HILDEN J., The measurement of performance in probabilistic diagnosis. General Recomendations. **Meth. Inform. Med.**, vol. 20, 97-100, 1985.
34. HARMON, P., DAVID KING, JOHN, Sistemas especialistas - A inteligência artificial chega ao mercado, **Ed. Campus**, 1988, 304 pag.
35. HART, A., WYATT, J., Evaluating black-boxes as medical decision aids: issues arising from a study of neural networks, **Med. Inform.**, vol.15, n. 3, p. 229-236, 1990.
36. HEATHFIELD, H. A., WYATT, J., Philosophies for the design and development of clinical decision support systems, **Methods of Information in Medicine**, vol. 32, p. 1-8, 1993.
37. HECKERLING, P., ELSTEIN, A., The effect of incomplete knowledge on the diagnoses of a computer consultant system, **Med. Inform.**, vol. 16, n. 4, p. 363-370, 1991.
38. HELLER, M., AI in Practice, **Byte**, p. 267-278, janeiro, 1991
39. HOLLAND, J. D., **Adaptation in Natural and Artificial systems**, Bradford Books, 1992.
40. JENKIS, J., Computerized electrocardiography. **Rev. Bioengng**, p. 307- 350, Novembro, 1981
41. JOHNS, R.J., FORTUIN, N.J. Clinical information and clinical problem solving. In: HARVEY, A.M., JOHNS, R.J., MCKUSICK, V.A., OWENS, JR A.H., ROSS,R.S. **The principles and practice of medicine**. 22 ed. Norwalk, Connecticut: Appleton & Lange, 1988. m
42. JOHNSON, R. COLIN, What is cognitive computing?, **Dr. Dobb's Journal**, p. 18 - 24, fevereiro, 1992.

43. KASSIRER, J.P., GORRY, G.A. Clinical problem solving: a behavioral analysis. **Ann. Inter. Med**, v.89, p. 245-255, 1978.
44. KASSIRER, J.P., A report card on computer-assisted diagnosis - the grade: C, **The New England Journal of Medicine**, p. 1824- 1825, Junho, 1994.
45. KELSEY, M.C., Intelligent systems: how can they help? , **Journal of hospital Infection**, Supplement A 418-423 , 1991.
46. LAWRENCE, J, Introduction to neural networks and expert systems, **California Scientific Software**, Nevada City, 260 p.,1992.
47. LIMA, W.C., **Grupo de Pesquisas em Engenharia Biomédica (GPEB)**. Florianópolis: Curso de Mestrado em Medicina Interna da UFSC, 1993. (Notas de aula).
48. LINDSAY, R. K., BUCHANAN, B.G., FEIGENBAUM, E. A., LEDEBRG, J., Applicatins of artificial itelligence for organic chemistry: the dendral project, **McGraw-Hill**, 1980.
49. LIPPMAN, R. P., An introduction to computing with neural nets, **IEE ASSP Mag.** p. 4-21, April, 1987.
50. LOPES, H. S., NASSAR S., RAMIREZ A., LIMA W.C., Arquiteturas Híbridas Integrando Paradigmas Conexionistas e Simbolistas, **I SBAI - UNESP**, Rio Claro, SP, Brasil, 1993,
51. LUNDSGAARDE, H.P., Evaluating medical expert systems. **Soc. Sci. Med.**, v. 24, n. 10, p.805-819, 1987.
52. MACHADO, R. J., ROCHA, A. F., LEÃO, B. F., Calculating the mean knowledge representation from multiple experts. in: KACPRZYK, J & FEDRIZZI, M. (eds.), **Multiperson Decision Making Using Fuzzy Sets and Possibility Theory**. (Netherlands): Kluwer Academic Publishers, p. 113-127, 1990.
53. MILLER, P., SITING, D., The evaluation of clinical decision support systems: what is necessary what is interesting, **Med. Inform.**, vol. 15, n. 3, p. 185-190, 1990.

54. MILLER, P., POPLE Jr., H. E., MYERS, J.D., INTERNIST-1, an experimental computer-based diagnostic consultant for general internal medicine. **N. Eng. J. Med.**, v. 307, n. 8, p. 468-476, 1982.
55. NASSAR, S., **Sistema Especialista Estatístico para Apoio a Pesquisas Médicas**, Universidade Federal de Santa Catarina - UFSC (Florianópolis). Projeto de Tese de Doutorado - Grupo de Pesquisas em Engenharia Biomédica - Departamento de Engenharia Elétrica - UFSC, 1994.
56. NASSAR, S. M., BARRETOM, J. M., LIMA, W.C. , "Distributed artificial intelligence on quantitative medical research", **World Congress on Medical Physics and Biomedical Engineering**, Rio de Janeiro, Proceedings, Parte I, Abstract PS17-1.5, p 460, 21-26 august 1994.
57. NIEVOLA, J. C., **Sistema especialista para auxílio ao diagnóstico médico de icterícia**. Florianópolis: UFSC, 1988, 90 p. Dissertação (Mestrado no Grupo de Pesquisas em Engenharia Biomédica) - Pós-graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, 1988.
58. NIEVOLA, J. C., **Sistema Inteligente para auxílio ao ensino em traumatologia crânio-encefálica**, Tese (Doutorado no Grupo de Pesquisas em Engenharia Biomédica) - Pós-graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, 1995.
59. NYKÄNEN, P., CHOWDHURY,S., WIGERTZ, O. Evaluation of decision support systems in medicine. **Computer methods and Programs in Biomedicine**, v.34,p.229-238,1991.
60. OLIVEIRA, J.P., GARCIA, R.O., LIMA, W.C., "Artificial neural networks for representing medical knowledge in anesthesiology", **World Congress on Medical Physics and Biomedical Engineering**, Proceedings, Parte I, Abstract PS17-1.19, p. 467, 21-26 august 1994.

61. PASSOLD, F. **Sistema especialista híbrido em anestesiologia para pacientes críticos/problemáticos**. Florianópolis, UFSC, 1995, 190p. Dissertação (Mestrado no Grupo de Pesquisas em Engenharia Biomédica).
62. PASSOLD, F., GARCIA, R. O, LIMA, W. C., "Artificial neural networks for representing medical knowledge in anesthesiology", **World Congress on Medical Physics and Biomedical Engineering**, Rio de Janeiro, Proceedings, Parte I, Abstract PS 17-2.15, p. 467, 21-26 august 1994.
63. PEREIRA, N. ALVES, Método de aquisição de conhecimento para sistemas especialistas para problemas de diagnose, **I SBAI, UNESP** , Rio Claro/SP , Brasil, Setembro de 1993.
64. RAMIREZ, A.T.P. **Sistema inteligente para treinamento do controle de infecção hospitalar**. Florianópolis, UFSC, 1991, 203p. Dissertação ( Mestrado no Grupo de Pesquisas em Engenharia Biomédica).
65. RICH, E., **Inteligência Artificial, McGraw-Hill** , 1993.
66. ROCHA, A. R., LAGINNA, M. P. R., MACHADO, R., SIGUIEV, D., ANÇÃO, M., Declarative and procedural knowledge: two complement thesis for expertise. in: VERDEGAY, J. L., DELGADO, M (eds.) **Approximate Reasoning Tools for Artificial Intelligence**. (Rheinland): Verlag TUV, 1990.
67. ROSSI-MORI, A., PISANELLI, D., RICCI F., Evaluation stages and design steps for knowledge-based systems in medicine, **Med. Inform.**, vol. 1, n. 3, p.191-204, 1990.
68. SABBATINI, R.M.E. O diagnóstico médico por computador. **Informédica**, v. 1, n.1, p. 5-10, 1993.
69. SHORTLIFFE, E.H. **Computer-Based Medical Consultation: MYCIN**. Amenterdam, **Elsevier Scientific Publ.**,1976
70. SHORTLIFFE, E.H. Medical expert systems - knowledge tools for physicians. **West J. Med.**,v. 145 [special issue], p. 830-839,1986.
71. SHORTLIFFE, E.H. Computer programs to support clinical decision making, **JAMA**, v.258, n. 1, p. 61-66,1987.

72. SHORTLIFFE, E.H., Knowledge-based systems in medicine, **Lectures Notes in Medicine Informatics MIE 91**, n. 45, p. 5-10, agosto, 1991.
73. SPILLMAN, Richard, Genetics Algorithms; Nature's way to search for the best, **Dr. Dobb's Journal**, p. 26-30, fevereiro, 1993.
74. SYMPOSIUM ON MEDICAL INFORMATICS, 1985, Washington, Medical education in the informatic age, 1985. Washington, DC.: **Association Of American Medical Colleges**, 1985.
75. VANKER A. D., van STEOCKER, W., An expert diagnostic program for dermatology, **Comput. Biomed. Res.**, 17, p. 241-247, 1984.
76. TOWNSEND, C, **Introduction to Expert Systems**, Oxford University Computing Laboratory Programming Research Group, Fevereiro, 1992
77. WARNER H. P., Computer-assisted medical decision-making, **Academic Press**, 1979.
78. WHITEBECK, C., BROOKS, R. Criteria for evaluating a computer aid to clinical reasoning. **J.Med. Philos**, v.8,p. 51-65, 1983.
79. WOLFAN, S., **Mathematica, A System for doing mathematics by computer**, Addison-Wesley Publishing Company, 1988.
80. WYATT, J., Building computer aids for medical decisions: a review and discussion of the techniques available. **Adv. Bioscienc.**, v. 74, p.381-396,1989a.
81. WYATT, J., Lessons learn from the field trial of ACORN, an expert system to advise on chest pain. **Proceedings MedInfo'89**, p.111-115, 1989b.
82. WYATT, J., SPIEGELHALTER, D., Evaluating medical expert systems: what to test and how? **Med. Inf.**, v.15,n. 3,p. 205-217,1990.
83. WYATT, J.,Computer-based knowledge systems. **Lancet**,v.338,p.1431-1436,1991a.
84. WYATT, J. Standford AI in medicine workshop. March 1990. **Meth. Inform. Med.**, v.30, p.65-67,1991b.
85. YOUNG, F.E. Validation of medical software- Present policy of Food and Drug Administration. **Ann. Intern. Med.**, v.106,n. 4, p. 628-629,1987.