UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

O Estudo e Desenvolvimento do Protótipo de uma Ferramenta de Apoio a Formulação de Consultas a Bases de Dados na Área da Saúde

por

CARINE GELTRUDES WEBBER

Dissertação submetida à avaliação, como
requisito parcial para a obtenção do grau de
Mestre em Ciência da Computação

Prof. José Mauro Volkmer de Castilho
Orientador

Porto Alegre, abril de 1997.

CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Webber, Carine Geltrudes

O Estudo e Desenvolvimento do Protótipo de uma Ferramenta de Apoio a Formulação de Consultas a Bases de Dados na Área da Saúde / por Carine Geltrudes Webber. — Porto Alegre: CPGCC da UFRGS, 1997.

100f.: il.

Dissertação (mestrado) — Universidade Federal do Rio Grande do Sul. Curso de Pós-Graduação em Ciência da Computação, Porto Alegre, BR-RS, 1997. Orientador: Castilho, José Mauro Volkmer de.

1. Recuperação de Informações. 2. Formulação de Consulta. 3. *Thesaurus*. 4.Terminologias Médicas I.Castilho, José Mauro Volkmer de. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Reitora: Prof. Wrana Maria Panizzi
Pró-Reitor de Pesquisa e Pós-Graduação: Prof. José Carlos Ferrz Hennemann
Diretor do Instituto de Informática: Prof. Roberto Tom Price
Coordenador do CPGCC: Prof. Flávio Rech Wagner
Bibliotecária-Chefe do Instituto de Informática: Zita Prates de Oliveira

Dedicatória

Aos meus pais

Agradecimentos

Neste oportunidade gostaria de prestar um agradecimento especial:

Ao meu orientador, Dr.José Mauro Volkmer de Castilho, inicialmente por ter me aceito como sua orientanda neste curso e, desta forma, me oferecido uma oportunidade de crescimento pessoal e científico imensurável. A sua convivência me acrescentou muitos exemplos de vida e amizade dos quais eu jamais me esquecerei.

À Dra.Beatriz de Faria Leão cuja contribuição e incentivo foi de grande importância para a realização deste trabalho.

Aos colegas do grupo de BDI: Álvaro, Juliano, Miguel, Palazzo, Rafael, Paulo, Carolina, Rodrigo, Fabiana e, em especial, a Fabiane e a Gisele.

A toda a minha família por me aguentar todo este tempo e me aceitar em casa de novo.

Aos meus amigos Alexandre Ribeiro e Heitor Strogulski pelo incentivo na realização deste curso.

A todos os meus amigos de Caxias que mesmo longe sempre me apoiaram e não se esqueceram de mim. Em especial à minha grande amiga Cristina, pela sua compania, incentivo e amizade.

A todos os meus amigos e colegas do CPGCC, pela amizade e troca de experiências.

Ao pessoal da secretaria do CPGCC, do Instituto de Informática e da Biblioteca do Instituto de Informática por todo auxílio eficientemente prestado.

À CAPES pelo incentivo financeiro.

À Deus por toda luz e serenidade nos momentos difíceis.

Sumário

Lista de Abreviaturas

CID9	Classificação Internacional de Doenças Nona Edição
DeCS	Descritores Médicos da Saúde
INN	International Nonproprietary Name
MEDLINE	Medical Literature Analysis and Retrieval System On Line
MeSH	Medical Subject Headings
MLDIP	Medical Letter's Drug Interaction Program
NLM	National Library of Medicine
SGN	Standard Generic Name
SNOMED	Systematized Nomenclature of Medicine
SRI	Sistema de Recuperação de Informações
UMLS	Unified Medical Language System
USAN	United States Adopted Names
USP	United States Pharmacopeia

Lista de Figuras

Lista de Tabelas

Resumo

O objetivo deste trabalho é, através do estudo de diversas tecnologias, desenvolver o protótipo de uma ferramenta capaz de oferecer suporte ao usuário na formulação de uma consulta à MEDLINE (Medical Literature Analysis and Retrieval System On Line).

A MEDLINE é um sistema de recuperação de informações bibliográficas, na área da biomedicina, desenvolvida pela National Library of Medicine. Ela é uma ferramenta cuja utilização tem sido ampliada nesta área em decorrência do aumento da utilização de literatura, disponível eletronicamente, por profissionais da área da saúde.

As pessoas, em geral, buscam informação e esperam encontrá-la exatamente de acordo com as suas expectativas, de forma ágil e utilizando todas as fontes de recursos disponíveis. Foi com este propósito que surgiram os primeiros Sistema de Recuperação de Informação (SRI) onde, de forma simplificada, um usuário constrói uma consulta, a qual expressa sua necessidade de informação, em seguida o sistema a processa e os resultados obtidas através dela retornam ao usuário.

Grande parte dos usuários encontram dificuldades em representar a sua necessidade de informação de forma a obter resultados satisfatórios em um SRI. Os termos que o usuário escolhe para compor a consulta nem sempre são os mesmos que o sistema reconhece.

A fim de que um usuário seja bem sucedido na definição dos termos que compõem a sua consulta é aconselhável que ele conheça a terminologia que foi empregada na indexação dos itens que ele deseja recuperar ou que possa contar com um intermediário que possua esse conhecimento. Em situações em que nenhuma dessas possibilidades seja verdadeira recursos que viabilizem uma consulta bem sucedida se fazem necessários.

Este trabalho, inicialmente, apresenta um estudo geral sobre os Sistemas de Recuperação de Informações (SRI), enfocando todos os processos envolvidos e relacionados ao armazenamento, organização e à própria recuperação. Posteriormente, são destacados aspectos relacionados aos vocabulários e classificações médicas em uso, os quais serão úteis para uma maior compreensão das dificuldades encontradas pelos usuários durante a interação com um sistema com esta finalidade. E, finalmente, é apresentado o protótipo do Sistema para Formulação de Consultas à MEDLINE, bem como seus componentes e funcionalidades.

O Sistema para Formulação de Consultas à MEDLINE foi desenvolvido com o intuito de permitir que o usuário utilize qualquer termo na formulação de uma consulta destinada a MEDLINE. Ele possibilita a integração de diferentes terminologias médicas, originárias de vocabulários e classificações disponíveis em língua portuguesa e atualmente em uso. Esta abordagem permite a criação de uma terminologia biomédica mais completa, sendo que cada termo mantém relacionamentos, os quais descrevem a sua semântica, com outros.

Palavras-chave: Recuperação de Informações, Formulação de Consultas, Terminologias Médicas, Thesaurus.

Title: "The Study and Development of the Prototype of a Tool for Supporting Query Formulation to Databases in the Health Area"

Abstract

The goal of this work is, through the study of many technologies, to develop the prototype of a tool able to offer support to the user in query formulation to the MEDLINE (Medical Literature Analysis and Retrieval System On Line).

The MEDLINE is a bibliographical information retrieval system in the biomedicine area developed by National Library of Medicine. It is a tool whose usefulness has been amplifyed in this area by the increase of literature utilization, eletronically available, by health care profissionals.

People, in general, look for information and are interested in finding it exactly like their expectations, in an agile way and using every single information source available. With this purpouse the first Information Retrieval System (IRS) emerged, where in a simplifyed way, a user defines a query, that expresses an information necessity and, one step ahead, the system processes it and returns to the user answers from the query.

Most of the users think is difficult to represent their information necessity in order to be succesful in searching an IRS. The terms that the user selects to compose the query are not always the same that the system recognizes.

In order to be successfull in the definition of the terms that will compose his/her query is advisable that the user know the terminology that was employed in the indexing process of the wanted items or that he/she can have an intermediary person who knows about it. In many situations where no one of these possibilities can be true, resources that make a successfull query possible will be needed.

This work, firstly, presents a general study on IRS focusing all the process involved and related to the storage, organization and retrieval. Lately, aspects related to the medical classifications and vocabulary are emphasized, which will be usefull for a largest comprehension of the difficulties found by users during interaction with a system like this. And, finally, the prototype of the Query Formulation System to MEDLINE is presented, as well as its components and funcionalities.

The Query Formulation System to MEDLINE was developed with the intention of allowing the user to use any term in the formulation of a query to the MEDLINE. It allows the integration of different medical terminologies originated from classifications and vocabulary available in Portuguese language and in use today. This approach permits the creation of a more complete biomedical terminology in which each term maintains relationships that describe its semantic.

Keywords: Information Retrieval, Query Formulation, Medical Terminology, *Thesaurus*

1 Introdução

Há algum tempo a ciência da computação vem se preocupando com o armazenamento e a recuperação de informações, desenvolvendo e aprimorando ferramentas com estas funcionalidades.

Com a crescente substituição do papel pela a mídia eletrônica o número de usuários que acessam sistemas de computação atingiu grandes proporções. Aliado a isto, a popularização da Internet deu um grande impulso à disseminação das informações, anteriormente restritas a um grupo ou ambiente, e favoreceu o acesso direto à informação pelos usuários.

Atualmente nós, seres humanos, sofremos com o excesso de informação que nos é disponibilizada. Diariamente mais e mais informação é produzida e consumida. Sua origem pode ser tão variada quanto seu formato e propósito. O fato é que a quantidade de informação, que de alguma forma está ao nosso alcance, extrapola enormemente a capacidade humana de manipulação satisfatória.

Em geral, as pessoas buscam informações e esperam encontrá-la exatamente de acordo com as suas expectativas e de forma rápida utilizando as diversas fontes disponíveis. Com este propósito surgiram os primeiros Sistemas de Recuperação de Informações.

Passados muitos anos desde o surgimento do primeiro grande Sistema de Recuperação de Informações (SRI), o SMART [SAL 68], ainda se tem como um grande desafio a construção de um sistema realmente eficiente em todos os seus aspectos. Constantes avanços têm sido feitos no sentido de desenvolver novas técnicas ou aprimorar as existentes para a execução de tarefas, porém, ainda verifica-se uma certa distância da situação ideal.

Tradicionalmente, um SRI tem como meta recuperar documentos relevantes em resposta a uma consulta do usuário. Hoje, o que se percebe é que a funcionalidade dos SRI's tem aumentado. Isto porque os usuários não esperam mais apenas a recuperação mas também querem que o sistema sinta-se responsável pelo gerenciamento do texto, o que inclui disseminação, extração, categorização e roteamento de documentos e também um melhor "entendimento" da necessidade de informação do usuário. Os sistemas tradicionais não tem a capacidade de satisfazer estas necessidades e, por isso técnicas de inteligência artificial têm sido aplicadas a fim de suprir estas novas necessidades e apresentar soluções para novos problemas [LEE 93].

Separar o joio do trigo. Esta expressão tem uma conotação bastante realística em se tratando de um sistema de recuperação. Pode-se perceber dois tipos de informação: aquela que identifica-se como útil, pelo conteúdo que representa, e aquela que identifica-se como inútil e deve ser desprezada. Esta classificação não é tão fácil de ser feita. A determinação da utilidade de certa informação é dependente também do usuário que fez a consulta. Dependendo de fatores como a sua área de atuação, linha de pesquisa e interesse pessoal, o usuário pode estar interessado em determinada

abordagem da informação. O ideal seria que o sistema considerasse características do usuário como seu modelo mental e as diferenças individuais.

Esta dependência do sistema em relação ao usuário se inicia no momento em que ele define uma consulta, a qual deve expressar claramente uma necessidade de informação, e segue até o momento em que o usuário satisfeito encerra esta atividade de pesquisa. Os problemas gerados pela dependência do usuário afetam diretamente o desempenho do sistema como um todo e são, geralmente, causados pelas diferenças de linguagem entre o sistema e o usuário.

A tarefa de selecionar informação e classificá-la requer grande habilidade e depende das várias etapas de organização e manipulação de informações. Destaca-se a indexação, que é um fator de extrema importância para este processo, já que é a partir dela que as informações são identificadas. A pesquisa pelas informações também se constitui um fator problemático pois é sabido que o usuário acha complicado escrever o que ele deseja utilizando operadores booleanos e palavras chaves, e a utilização da linguagem natural também não é muito clara pelo vocabulário restrito a que está associada. Para que ele possa obter resultados significativos deve interagir com o sistema de forma a ir restringindo o número de respostas.

A maior restrição de um SRI é que a mesma linguagem deve ser utilizada tanto para descrever o conteúdo dos documentos quanto a requisição de informação. Uma linguagem é composta de um vocabulário baseado em uma visão particular ou enfoque específico. Para que uma consulta alcance resultados satisfatórios ela deve estar construída dentro desta mesma visão particular ou dentro deste mesmo enfoque. Se considera-se um sistema onde documentos estão indexados pelo seu próprio vocabulário e as consultas são feitas através de um vocabulário controlado o processo de comparação pode ser bastante dificultado, principalmente devido ao fato de que o sucesso, ou fracasso, da consulta vai depender grandemente da forma de expressão dos autores do documento e da consulta.

Uma das principais tarefas envolvidas na pesquisa em bancos de dados de informações é a de expressar a necessidade de informação do usuário na linguagem de consulta do sistema [PEN 93]. Em geral, o usuário produz uma consulta diretamente, mapeando sua necessidade de informação para uma linguagem de representação. O usuário expressa sua requisição em linguagem natural ou artificial, sendo esta restrita e definida para um sistema em particular. Uma boa formulação da consulta requer uma boa dose de compreensão e conhecimento da linguagem e das características internas do sistema. Com o aumento das bases de dados disponíveis, o aprendizado necessário para se conseguir fluência com mais de umas poucas fontes torna-se impraticável. A qualidade da formulação da consulta pode ser melhorada com o auxílio de um intermediário, cujo conhecimento sobre o banco de dados, a sintaxe da linguagem de consulta e dos termos indexadores exerce um papel fundamental.

Ao considerar-se um domínio específico e os termos utilizados nele, verifica-se graves problemas de linguagem. Estes problemas atingem tal nível que nem mesmo os profissionais de uma área específica concordam com determinadas denominações. Dentro desta problemática a área da biomedicina se encaixa perfeitamente. Nela existem múltiplos termos para denominar conceitos, nem todos aceitos pelos profissionais atuantes.

Como conseqüência direta deste fato verifica-se a dificuldade que estes profissionais apresentam na formulação da consulta para uma pesquisa em um sistema

de recuperação de referências bibliográficas. Os termos que o usuário escolhe para a consulta nem sempre são os mesmos que o sistema reconhece.

Com o passar dos anos tem aumentado a utilização da literatura na solução de problemas na área médica. Cada vez mais pesquisadores e, principalmente, os próprios médicos têm sentido necessidade de empregar a literatura disponível eletronicamente em situações em que é preciso solucionar problemas associados ao tratamento de pacientes ([FLO 95] e [GUI 93]). Por este motivo os sistemas que contribuem para esta tarefa não podem ser dispensados, porém, devem surgir ferramentas que apoiem esta busca de informações.

1.1 Objetivos do trabalho

O objetivo deste trabalho é estudar mecanismos que possam ser aplicados em um Sistema de Recuperação de Referências Bibliográficas para a área da Biomedicina com a finalidade de auxiliar o usuário nas suas atividades de pesquisa. Posteriormente, estes mecanismos serão aplicados na construção de um protótipo que vise atender estas necessidades. Para que se alcance este objetivo pretende-se também:

1. desenvolver um estudo sobre os SRI's e todos os processos que estão envolvidos desde sua concepção até sua utilização;
2. estudar as terminologias médicas, que correspondem ao escopo deste trabalho, destacando os problemas encontrados na sua utilização decorrentes da ausência de padronização, tanto a nível mundial quanto local;
3. estudar as metodologias empregadas em sistemas já desenvolvidos para a solução das diferenças entre as terminologias em uso no ambiente médico;
4. geração de uma rede semântica composta de conceitos médicos e relações entre estes conceitos;
5. geração de um banco de dados de termos médicos oriundos de diversos vocabulários e relações entre eles;
6. implementação de um protótipo de um Sistema para a Formulação de Consultas à MEDLINE onde serão postos em prática os estudos realizados.

1.2 Organização do texto

Este trabalho está organizado em quatro capítulos. O capítulo 1 consiste da introdução deste trabalho. O capítulo 2 apresenta um estudo sobre os componentes de um SRI, destacando o processo de indexação e construção de uma consulta em um sistema. São descritos componentes que auxiliam estes processos bem como os problemas que um sistema de recuperação enfrenta na interação com o usuário.

O capítulo 3 é referente a terminologias e trata de aspectos relativos aos problemas que atualmente decorrem da falta de uma padronização nas terminologia, da área médica. Nele também é apresentado um sistema desenvolvido para lidar com

este problema de terminologias na medicina. Este sistema visa possibilitar a integração entre vocabulários e classificações atualmente em uso.

O capítulo 4 descreve a aplicação desenvolvida. É apresentada a arquitetura do sistema e são descritos os dois componentes deste sistema e suas respectivas funcionalidades. Aspectos de interação com os diversos tipos de usuário identificados também são apresentados.

E, por fim, o capítulo 5 apresenta as conclusões e sugestões para trabalhos futuros.

2 A Recuperação de Informações

"A recuperação de informações não é um evento, mas sim um processo."

Este capítulo apresenta uma revisão sobre Sistemas de Recuperação de Informações, destacando os processos de indexação e formulação de consultas. Alguns aspectos considerados determinantes para a eficiência da recuperação, bem como maneiras de avaliá-la, são descritos.

2.1 Sistema de Recuperação de Informações

Um Sistema de Recuperação de Informações (SRI) pode ser visto como uma memória, onde itens são guardados e então identificados para posterior recuperação. A recuperação desses itens inicia quando consultas fazem requisições que especificam as propriedades que os itens relevantes e desejados têm, como pode ser visto na figura 2.1 abaixo. As propriedades especificadas na consulta devem casar com a forma com que os documentos foram rotulados (indexados) na memória (Banco de Dados) [PAR 89].

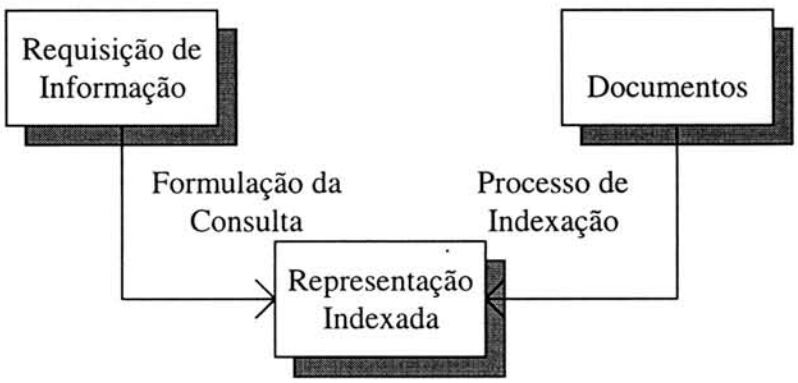

FIGURA 2.1 Relação entre os principais componentes de um SRI [PAR 89].

O maior desafio em qualquer processo de recuperação está em formular uma consulta que recupere os documentos considerados importantes. Isto depende da maneira como os documentos estão indexados e da forma como a relevância é atribuída a cada documento.

2.2 A Informação

Informação é um recurso necessário para desempenhar tarefas. Geralmente pode estar na forma de textos e dados expressos como fatos e regras, mas pode também estar expressa em outras formas como imagens e sons.

Informação é um dado que é útil a um cliente [WIE 96]. A palavra informação não representa um conceito passivo mas constitui um recurso que deve ser usado para aumentar a produtividade e auxiliar na tomada de decisões. A habilidade de definir a utilidade e relevância da informação é muito importante no momento de informar as pessoas o que elas necessitam saber, ao invés de oferecer apenas coisas gerais, muitas das quais inúteis.

Como qualquer outro recurso, a informação deve ser armazenada e tornada disponível quando necessária. A RI está relacionada com a estruturação, análise, organização, armazenamento, pesquisa e recuperação de informações. As informações armazenadas são respostas as consultas do usuário.

2.3 Desempenho da Recuperação: abrangência e precisão

Tradicionalmente o desempenho da recuperação tem sido avaliado usando duas medidas chamadas abrangência ("recall") da recuperação e precisão ("precision") [SAL 68]. Estas medidas são baseadas nos julgamentos subjetivos de relevância feita pelos usuários quando eles examinam citações de documentos e *abstracts* [SAL 94].

A abrangência da recuperação é a proporção de documentos importantes do banco de dados que foram recuperados e a precisão é a proporção de documentos que foram relevantes ao usuário dentro do conjunto que foi recuperado.

A abrangência da recuperação normalmente é expressa na forma de porcentagem. É difícil de ser estimada porque saber o número total de documentos relevantes a uma pesquisa contidos no banco de dados é uma tarefa árdua.

A precisão é importante pois o usuário não quer se distrair com coisas irrelevantes. Tanto quanto possível o sistema deve ser capaz de limitar informações e selecionar opções relevantes, considerando o interesse, experiência do usuário, bem como o tipo de tarefa que ele está desenvolvendo.

Pode também ser incluído um terceiro critério de avaliação, a utilização, que pode ser vista como o número de documentos obtidos na busca que foram efetivamente utilizados pelo usuário. A utilização está relacionada com o fato de que um documento é importante se o usuário faz uso dele e um documento, supostamente relevante, não é bom se o usuário não faz uso dele. A utilização pode ser definida de várias formas: número de documentos que foram lidos, número de documentos que foram citados em artigos, número de documentos que foram copiados para outros meios físicos e número de documentos que o usuário pode lembrar após um intervalo de tempo.

2.4 A Indexação

A RI depende da indexação [LEW 96] . A indexação é a base para recuperar documentos relevantes às necessidades dos usuários. Ela tem como meta aumentar a precisão e a abrangência da recuperação e, desta forma, acelerar o acesso à informação [WIE 96], valendo-se, para isso, de uma linguagem de indexação. A indexação requer uma linguagem com um vocabulário de termos e um método para construir requisições e descrições do documento.

O processo de indexação em um sistema de manipulação de documentos tem a mesma função que a indexação tem em uma biblioteca, que é a de fornecer para cada item armazenado um conjunto de indicadores que refletem o conteúdo do documento [SAL 75].

O processo de indexação cria uma estrutura de arquivo indexado, geralmente referenciado como arquivo invertido, sendo que ele inicia com um termo e identifica todos os documentos que são indexados por aquele termo [BRO 94]. Quando se pesquisa por alguma informação usando arquivo invertido, o índice é examinado para determinar quais itens satisfazem a requisição da pesquisa. Pesquisando através dos índices torna o serviço muito mais rápido do que através dos documentos originais [PAR 89].

As primeiras teorias de indexação usavam um critério baseado na freqüência da ocorrência de termos presentes em cada um dos documentos de uma coleção. Os termos mais freqüentes dentro de um determinado documento recebiam pesos mais altos e, desta forma, estavam identificados como índices [SAL 75].

Mais tarde se tornou claro que a utilidade de um dado termo para a representação do conteúdo não pode ser atribuída suficientemente bem olhando apenas para os termos que ocorrem em um documento por vez. Ao invés disso, pode ser necessário determinar o comportamento do termo dentro da coleção de documentos. Um bom termo para indexar e recuperar é aquele que identifica o conteúdo de documentos individualmente e que, ao mesmo tempo, consegue distinguir um documento do restante da coleção [SAL 87].

2.4.1 Linguagem de Indexação

A linguagem de indexação pode ter seus termos selecionados a partir dos textos a serem indexados ou ela pode estar limitada por uma vocabulário controlado. As linguagens de indexação variam de acordo com inúmeros fatores, tais como: a forma dos termos que a compõem, a ênfase dada aos termos, as relações entre eles, entre outros.

Cada item armazenado tem seu conteúdo identificado por um conjunto de termos, palavras chaves, descritores ou conceitos, pertencentes a linguagem de indexação. Estes termos identificam o conteúdo de cada item e controlam as operações de recuperação indicando aqueles itens cujos termos se mostram mais similares as formulações de consultas dadas. Um peso pode estar ligado aos termos refletindo a sua importância presumida.

O processo de indexação de documentos pode ser executado manual ou automaticamente.

2.4.2 Indexação Manual

A indexação manual é uma tarefa difícil e normalmente as pessoas que fazem isso, os indexadores, são auxiliados pela aplicação de regras que ajudam a manter a consistência da indexação. Os indexadores devem ler e compreender muitos documentos, lembrar o vocabulário controlado e as regras para atribuir termos do vocabulário controlado.

Métodos para auxiliar a indexação manual incluem dicionários, referências cruzadas, *thesaurus*, especificações do escopo para definir o significado ou interpretação de cada termo disponível, planilhas para especificar a ordem e o método a ser usado na atribuição de termos, relações entre termos e sumário estatístico descrevendo o desempenho da indexação já feita, listando, por exemplo, a freqüência existente de atribuição de cada termo do índice. Na prática, a indexação manual é muito trabalhosa e difícil e os resultados podem não ser aqueles esperados. Com grandes bancos de dados, torna-se praticamente impossível manter a consistência total quando trabalha-se com diferentes indexadores. Até mesmo o próprio indexador pode notar inconsistências na sua indexação no decorrer do seu trabalho.

Na prática operacional, o processo de indexação é quase invariavelmente executado manualmente por pessoal treinado. A pessoa que acompanha a indexação deve ser intimamente conhecedora do vocabulário de indexação e a sua prática, mas também deve conhecer as características da coleção. Um conjunto de itens eficientemente indexados deve refletir o tipo de consultas do usuário que o sistema espera processar no futuro. Pessoas que indexam e não apresentam alguma experiência requerida comumente atribuem termos inapropriados ou pesos incorretos ou relações entre termos sem importância. Diferentes indexadores podem indexar um mesmo item com termos diferentes e esta inconsistência afeta a performance da recuperação.

2.4.3 Indexação Automática

A indexação automática é freqüentemente baseada em análises estatísticas de palavras e frases usadas no texto dos documentos [PAR 89]. Ela ocorre ou sobre os textos originais dos documentos ou sobre certos trechos do texto, tais como títulos e *abstracts*. Normalmente, as palavras que ocorrem em cada documento são listadas e certas medidas estatísticas são feitas quase sempre baseadas na freqüência de palavras dentro dos documentos, na freqüência total da coleção, ou na distribuição de freqüência através dos documentos da coleção. As palavras de maior freqüência são rejeitadas e pesos são atribuídos as palavras restantes, de acordo com uma computação estatística previamente realizada.

Na maioria dos sistemas, o conjunto de termos resultante, com peso ou não, é usado como identificação do conteúdo do documento ou consulta. Alternativamente, os termos ponderados extraídos a partir dos textos originais podem ser expandidos

pela adição de novos termos relacionados, identificados estatisticamente pelo estudo do vocabulário da coleção ou obtido pela consulta de dicionários, lista de sinônimos e *thesaurus*.

No modelo tradicional, a indexação é vista como um processo de classificação, onde os documentos são rotulados de acordo com os assuntos abordados. Este tipo de índice é semelhante ao que se encontra no fim dos livros, exceto ao fato de que é construído fora do vocabulário controlado dos termos que são escolhidos para abranger os conceitos possíveis sem incluir termos com significado sobrecarregado e sinônimos.

Segundo [PAR 89], um método possível de indexação possui os seguintes passos :

a) calcular a freqüência de cada termo para cada documento no banco de dados. A freqüência do termo K no documento I é denotada por :

$$FREQ_{IK}$$

b) determinar a freqüência total de cada palavra em todos os documentos do banco de dados, que é :

$$TOTFREQ_{K}$$

c) ordenar as palavras em ordem descendente de acordo com a sua freqüência no banco de dados e remover aquelas de maior freqüência e os de menor freqüência;

d) usar as demais palavras de média freqüência como índices para o banco de dados.

O método básico de indexação automática pode ser estendido. Alguns métodos envolvem "limpar" palavras removendo plurais e outras palavras que possuem o mesmo radical e diferem apenas no sufixo. Outros ainda mais complexos examinam a ocorrência estatística das palavras e removem aquelas que aparecem com sinônimos ou com significado sobrecarregado ([LOP 96] e [LOP 96a]).

Embora os detalhes do processo de análise variem de um sistema para outro, todos os métodos de indexação automática são necessariamente "derivativos", no sentido que o documento ou consulta originais servem como a entrada principal para a operação de indexação [SAL 75]. Este método tem recebido algumas críticas:

a) a linguagem de entrada está restrita ao modo particular de expressão de autores de determinados documentos ou consultas;

b) linguagem dos autores pode ser sensível a um período de tempo específico ou a um ambiente em particular. Desta forma a terminologia pode se tornar desatualizada ou conduzir a conotações incorretas;

c) uma metodologia estatística baseada em técnicas de contagem de palavras, em alguns casos, é inadequada para a análise de textos.

Segundo [KRO 92] existem dois problemas associados a representação do conteúdo de documentos usando palavras. O primeiro está relacionado a ambigüidade das palavras utilizadas, o que pode causar a recuperação de documentos não

relevantes. Neste caso, o sistema pode fornecer frases que utilizam o termo permitindo que o usuário escolha aquela que emprega o termo no sentido desejado. O segundo problema esta relacionado ao fato de que um documento pode ser relevante à uma dada requisição embora ele não contenha exatamente as palavras utilizadas nela. O sistema, então, pode permitir a expansão da consulta utilizando palavras relacionadas a partir de um *thesaurus*. O problema central é que a semântica do texto não é bem representada por um conjunto de palavras consideradas individualmente [CRO 93]. Assim, a performance dos sistemas de recuperação que se baseiam no mapeamento exato de palavras do texto para palavras especificadas pelo pesquisador será, na melhor das hipóteses fraca.

O que se sabe é que o usuário não está preocupado em recuperar documentos que possuam exatamente determinada palavra, mas sim documentos que expressam o conceito que ela representa.

Ferramentas que auxiliam a normalização da linguagem podem ser utilizadas neste método automático, por exemplo, na forma de *thesaurus*, para mitigar os efeitos da variabilidade de linguagem entre autores e, se necessário, procedimentos estão disponíveis para atualizar o vocabulário armazenado a medida que mudanças ocorram na terminologia ou no ambiente de indexação.

2.5 A Pesquisa Intermediária

Conhecimento prévio é normalmente requerido para entender a política de indexação, empregada nos bancos de dados, e para formular consultas de pesquisa na linguagem do banco de dados. Uma vez que a maioria dos usuários não apresenta esta perícia, eles requerem ajuda de uma pesquisa intermediária para realizar a consulta. Bibliotecários são normalmente os intermediários entre os fornecedores de informação e pesquisadores da informação [WEB 96].

Intermediários humanos são especialistas que tem muito conhecimento para apoiar as suas ações [IIV 95].

O papel do intermediário é bastante óbvio em uma biblioteca. Nela o usuário utiliza a coleção da biblioteca e é auxiliado pelas ferramentas que fornecem as referências. Algumas dessas ferramentas passaram para a forma eletrônica, como os sistemas de informação on-line, bancos de dados on-line e CD-ROMs.

A maior tarefa do intermediário da pesquisa é escolher a ferramenta de pesquisa adequada, informação on-line por exemplo, e assim suprir as necessidades do usuário em termos de informações desejadas. Ele pode ser visto como uma terceira entidade que atua entre o usuário e o SRI.

Pode-se perceber dois canais de comunicação na operação durante uma pesquisa. O primeiro canal é usado pelo usuário e o intermediário na formulação de consultas de pesquisa. O segundo é usado pelo intermediário na interação com o banco de dados on-line, por exemplo. Falhas de comunicação em qualquer destes canais resultará numa recuperação fraca e pobre em termos de conteúdo esperado e performance.

Para que se possa conduzir uma busca bibliográfica eficiente, ao menos quatro classes de conhecimento são necessárias :

a) **Conhecimento procedural** no uso dos sistemas de computação, necessário a fim de iniciar e manter um padrão eficiente de comunicação com o sistema de recuperação;

b) **Conhecimento do domínio** que consiste do conhecimento sobre conceitos, relações entre conceitos e a terminologia do domínio do banco de dados;

c) **Estratégias** para formulação de consultas no banco de dados;

d) **Conhecimento específico** da construção do banco de dados e da política de indexação é também necessário e deve ter implicações na decisão sobre qual estratégia de pesquisa é mais promissora no contexto atual.

O intermediário deve monitorar o progresso da busca, enquanto atualiza seu modelo das informações necessárias ao usuário e lembra a linguagem de comandos e a terminologia do banco de dados. Naturalmente, há muita inconsistência na recuperação quando diferentes pesquisadores selecionam termos para descrever a mesma requisição de consulta. Há também muita inconsistência entre os pesquisadores no momento de selecionar os termos que compõem a consulta. É muito provável que alguns pesquisadores sejam mais consistentes que outros. Influências no processo de pesquisa podem vir da experiência do pesquisador em relação ao processo de pesquisa e ao estilo pessoal de selecionar os termos da consulta.

Os usuários poderão querer acessar um SRI diretamente da sua casa onde não contarão com o auxílio de um intermediário nem possuirão disponível o vocabulário controlado.

2.6 A Consulta

RI é um processo onde as informações que o usuário necessita são comparadas com as informações que estão disponíveis. Isto cria o paradigma da RI, onde a informação desejada é representada como uma consulta e a informação existente (documentos) é representada como uma coleção de índices, os quais podem ser mapeados contra a consulta. Aqueles documentos cujos índices mais proximamente casam com a consulta são assumidos como relevantes e suas citações são informadas ao usuário. O processo de casamento pode ser determinístico ou baseado em alguma métrica de distância ou algo similar ([PAR 89] e [BUS 95]).

No caso determinístico, uma string booleana é construída e os documentos recuperados são aqueles que casam precisamente com a consulta, com as restrições expressas na consulta.

Outra opção de casamento consiste em desenvolver representações de consulta e documentos que permitam que medidas de similaridade sejam calculadas entre consultas e documentos. Isto permite que documentos sejam ordenados de acordo com sua similaridade ou relevância para a consulta. Apesar das dificuldades na

formulação e validação de consultas baseadas em similaridade esta é uma idéia intuitivamente atrativa e interessante devido ao grande papel que os conceitos de similaridade tem na formação dos conceitos e na memória dos seres humanos.

Uma terceira opção de casamento é denominada recuperação conceitual [PAR 89]. Nela o usuário expressa a informação que ele necessita como um conceito, ao invés de uma coleção de palavras chaves combinadas com operadores lógicos.

A linguagem utilizada na requisição de uma pesquisa deve ser a mais próxima da natural possível. Porém, quando se analisa uma linguagem de consulta baseada em língua natural pode-se verificar inúmeros problemas, tais como:

a) é comum fazer-se uso de palavras que não contribuem à especificação do conteúdo da informação e que devem ser eliminadas;

b) muitas palavras podem ser usadas para significar o mesmo conceito;

c) o significado de algumas palavras depende do contexto no qual estão sendo usadas;

d) construções sintáticas diferentes podem representar a mesma idéia geral;

e) referências indiretas são muito usadas em linguagem natural onde pronomes, coletivos, etc. são usados para referenciar entidades conhecidas pelo contexto, mas cuja identificação do antecedente é complexa;

f) podem existir relações entre palavras, não contidas no texto, mas que podem ser deduzidas pelo contexto ou a partir de outros textos já analisados;

g) o significado de muitas palavras pode mudar com o tempo e novas palavras podem ser criadas para referenciar entidades já usadas e conhecidas.

2.6.1 Expansão da Consulta

Embora a eles não pareça, mesmo pesquisadores que se consideram experientes na utilização de SRI's poderiam melhorar o desempenho das suas pesquisas. Uma pesquisa que alcança resultados satisfatórios é conseqüência direta de uma consulta bem formulada aliada a um conjunto de informações armazenadas da maneira correta. Mas como pode-se determinar que um conjunto de informações foi indexado corretamente?

É natural esperar-se que especialistas na área discordem sobre os melhores termos para denominar um conceito, e isso será discutido na seção 2.9. Dentro da própria informática dicionários para a área são construídos com a finalidade de definir melhor alguns conceitos. Um exemplo de dicionário para informática pode ser encontrado em [WIE 96]. Como pode-se perceber, "estar correta" é uma qualidade bastante subjetiva para uma indexação e seu julgamento está intimamente ligado a análise dos resultados obtidos pela pessoa que formulou a expressão de consulta.

Se o desempenho de um SRI está relacionado a qualidade das consultas submetidas pelo usuário recursos que visem o aprimoramento de uma consulta devem ser oferecidos pelo sistema. As consultas normalmente podem ser expandidas por

termos adicionais, selecionados automaticamente ou com a cooperação do usuário [TER 96]. Algumas estratégias típicas de expansão de consultas incluem a seleção de termos julgados significativos para os documentos considerados relevantes pelo usuário após a verificação dos resultados obtidos através da recuperação executada pela consulta original [HIR 95].

A expansão de uma consulta pode ser feita de forma que itens do *thesaurus* são recuperados e são adicionados a ela [TER 96]. A adição de um termo sempre é mais recomendada do que a sua substituição porque ela representa um aumento comprovado no desempenho da consulta. Uma técnica de expansão de termos de uma consulta, descrita em [GAU 91], consiste dos seguintes passos:

1.inclusão dos sinônimos do termo em questão;

2.inclusão de termos genéricos ao termo extraídos do *thesaurus*;

3.inclusão dos termos que representam conceitos relacionados ao termo;

4.inclusão de termos específicos ao termos extraídos do *thesaurus*.

Outras técnicas de expansão sugerem a inclusão de termos indexadores de documentos já comprovados como relevantes, termos que correspondem a variações léxicas ou sugeridos pelo próprio usuário.

2.7 Dicionários

Dicionários são textos especiais cujo assunto principal é uma linguagem, ou um par de linguagens, no caso de um dicionário bilíngüe [GUT 96]. A função de um dicionário é fornecer uma ampla variedade de informação sobre palavras, tais como etimologia, pronúncia, entonação, morfologia e sintaxe, para dar definições de sentido das palavras e fornecer conhecimento não apenas sobre a linguagem mas sobre sua relação com o contexto em que pode estar inserida.

Os dicionários são usados para a normalização da linguagem. Eles não eliminam totalmente a ambigüidade, mas reduzem os efeitos de muitas das irregularidades pelo uso de algoritmos de mapeamento, os quais podem ser encontrados em [PAR 89].

Os dicionários variam muito na informação que contém e no número de sentidos que descrevem para as palavras. Mesmo grandes dicionários não contém uma listagem exaustiva de todos os sentidos de todas as palavras, mesmo porque novas palavras estão a cada momento sendo inseridas em nosso vocabulário.

2.7.1 Tipos de Dicionários

Basicamente um dicionário pode ser de diversos tipos [LOP 96] e [LOP 96a]), tais como:

a) **Dicionário negativo**: contém termos cuja utilização na indexação está descartada;

b) **Stop List**: utilizado na indexação textual, contém preposições, artigos e outras palavras que não devem ser utilizadas como índices pois não distinguem o documento que o possui do restante da coleção;

c) **Dicionário de frases**: contém expressões constituídas por palavras que quando utilizadas em conjunto assumem um sentido único;

d) **Vocabulário controlado**: um conjunto definido de termos, um dicionário, a partir do qual frases são tiradas para expressar conceitos. "Controlado" significa que há alguma uniformidade léxica reforçada para somente permitir expressões a partir do conjunto de termos pré-definidos. Escolher nomes para novas entradas é uma tarefa difícil, uma vez que deve-se escolher nomes simples e evitar ambigüidades. Manter o vocabulário controlado significa adicionar novos termos e resolver conflitos e ambigüidades. Esta não é uma tarefa trivial e requer esforço contínuo;

e) ***Thesaurus* e *Metathesaurus***: eles serão vistos com maior profundidade na seção 2.7.2 e 2.7.3 respectivamente.

2.7.2 *Thesaurus*

Um *thesaurus* é um agrupamento em classes de palavras ou raízes de palavras, contendo ou não um conjunto de relações entre estas palavras [JIN 93]. A dificuldade em seu uso esta em identificar as palavras que devem ser incluídas nele e em determinar que tipo de relações ele deve conter.

Tal classificação de termos traz como principais as seguintes vantagens ([SAL 75] e [CHE 96]):

a) fornecer um certo grau de normalização da linguagem oferecendo para cada termo de entrada um termo pertencente ao *thesaurus*;

b) ampliar o vocabulário de entrada adicionando termos extraídos do *thesaurus* aproveitando seus relacionamentos e, assim, aumentar abrangência da recuperação;

c) utilizar os relacionamentos de generalização e especialização disponíveis para expandir a consulta a fim de disponibilizar termos mais gerais ou mais específicos.

Um *thesaurus* é normalmente uma classificação de termos em classes de similaridade. O critério de classificação difere de sistema para sistema. Às vezes os termos devem ser estritamente sinônimos para se situarem na mesma classe, neste caso um dicionário de sinônimos é produzido. Alternativamente, os termos podem estar relacionados em algum sentido. Quando os *thesauri* são construídos manualmente o relacionamento entre os termos agrupados em uma dada classe é geralmente lingüístico, implicando em similaridade de significados. Quando a construção de um *thesaurus* é automática, o relacionamento é menos formal e pode estar limitado simplesmente as características de coocorrência entre termos nos documentos da coleção. Normalmente, nenhum relacionamento é especificado entre classes do *thesaurus*.

O *thesaurus* deve conter somente termos que são interessantes em uma determinada área para a qual ele se destina. Os termos ambíguos devem ser incluídos apenas para o sentido útil à coleção que se está considerando [SAL 75]. Quando isto não é possível deve-se ter cuidado no momento de expansão de uma consulta com termos do *thesaurus* para que seja utilizado o sentido correto da palavra [KRO 92].

2.7.2.1 Construção de *Thesaurus*

A construção de um *thesaurus* pode ser feita manual ou automaticamente. Manualmente eles são caros de construir e difíceis de atualizar. A correta determinação das relações entre itens é uma tarefa difícil mesmo que seja feita por especialistas humanos.[JIN 93]

Thesauri construídos de forma automática têm seus termos dependentes da coleção que os origina. São construídos baseados na coocorrência e julgamentos de relevância, usados para estimar a probabilidade de que termos do *thesaurus* são similares a termos da consulta. Apenas um tipo de relação de associação é estabelecido [JIN 93].

Para a construção de um *thesaurus* deve-se, inicialmente, gerar o vocabulário empiricamente indexando um conjunto representativo de documentos. Nesta primeira etapa pode-se selecionar alguns documentos e tentar extrair deles termos. Pode-se extrair o vocabulário de outro já existente, um *thesaurus* mais geral, ou desenvolver um *thesaurus* especializado dentro de uma estrutura geral de outro. Pode-se, também, colocar termos de diversas fontes incluindo glossários, outras publicações e de especialistas no assunto [CHE 96]. Após esta etapa deve ser feito um refinamento do vocabulário onde devem ser analisados por exemplo: variantes léxicas (verbos e adjetivos), plurais e singulares, e termos inúteis devem ser excluídos. Também deve-se criar uma estrutura hierárquica para a classificação [OGG 95].

Um *thesaurus* deve ser validado a fim de que sejam testados seus termos em relação a sua classificação e em relação a cobertura de conceitos abrangidos [JIN 93]. A validação de um *thesaurus* é feita a partir da expansão de consultas e medidas de eficiência da recuperação.

Informações mais detalhadas sobre classificação automática de termos pode ser encontrada em [JON 68] e [JON 70] e sobre construção automática de *thesaurus* pode ser encontrada em [SAL 68], [SAL 83], [SOE 74], [OGG 95] e [CHE 96].

2.7.3 *Metathesaurus*

Um *metathesaurus* pode ser definido como um dicionário mais completo, mais abrangente que um *thesaurus*.

O *metathesaurus* consiste em uma base de informações sobre conceitos que aparecem em um ou mais diferentes vocabulários controlados e classificações disponíveis em um determinado domínio. Em geral, o escopo do *metathesaurus* é determinado pelo escopo combinado dos vocabulários que o originaram. O

metathesaurus preserva o significado, conexões hierárquicas e outros relacionamentos entre termos presentes em um vocabulário enquanto adiciona certas informações básicas sobre cada um dos conceitos e estabelece novos relacionamentos entre conceitos e termos a partir de diferentes vocabulários [UML 93].

O *metathesaurus* está organizado por conceitos ou significados. Em essência, sua proposta é ligar nomes alternativos e visões do mesmo conceito e identificar relacionamentos úteis entre diferentes conceitos.

A operação de edição mais cara é a revisão manual dos termos unidos. Muito esforço tem sido feito para minimizar o número de uniões incorretas que acabam por ser excluídas pelo especialista. Os métodos de união de terminologias exploram casamento entre termos.

2.8 Feedback de Relevância

Com a finalidade de melhorar a performance da consulta pode-se aplicar o feedback de relevância [SAL 68]. O feedback de relevância é um método para melhorar a recuperação através da indicação dos documentos relevantes por parte do usuário de acordo com sua necessidade de informação [HAI 95]. A idéia básica do feedback de relevância consiste em refinar uma consulta baseando-se na relevância dos documentos obtidos com ela. Para isso são atribuídos pesos aos termos que compõe a consulta original e aqueles termos que aparecerem em documentos relevantes tem seus pesos aumentados, enquanto outros que aparecerem em documentos irrelevantes são removidos da consulta ou tem seus pesos reduzidos [ALL 95]. Durante o refinamento da consulta novos termos, presentes em documentos considerados relevantes, podem ser incluídos a ela.

Em suma, o processo consiste em executar uma pesquisa inicial e apresentar ao usuário uma certa quantidade de informação recuperada. O usuário examina os documentos recuperados e identifica os relevantes e os irrelevantes. Aqueles julgados relevantes são devolvidos ao sistema e os pesos dos termos da consulta que indexam esses documentos são atualizados. Já os termos que ocorrem em documentos considerados irrelevantes têm os seus pesos rebaixados.

As consultas alteradas são novamente submetidas ao sistema. Se o desempenho do sistema estiver de acordo com o esperado documentos adicionais serão recuperados. Este ciclo pode se repetir quantas vezes o usuário solicitar, porém, os documentos recuperados tendem a convergir [SAL 68].

A figura 2.2 apresenta o processo básico que um SRI executa e nela pode-se visualizar a etapa de feedback integrada com as demais. De um lado existe a necessidade de informação do usuário e, de outro lado, existem os documentos, que constituem respostas a esta necessidade. Essa necessidade de informação deve ser formalizada, utilizando termos adequados. Da mesma forma, os documentos textuais do sistema estão representados a partir desses termos. No próximo passo, o sistema, pelo método de comparação, recupera os documentos cujos termos indexadores são os mesmos que o usuário especificou em sua consulta. Esses documentos são

apresentados ao usuário que fez a requisição. Neste momento, o usuário pode avaliar os documentos recuperados e, a partir deles, reescrever a consulta inicial. Todo esse processo descrito pode se repetir até que o usuário esteja plenamente satisfeito com os resultados alcançados.

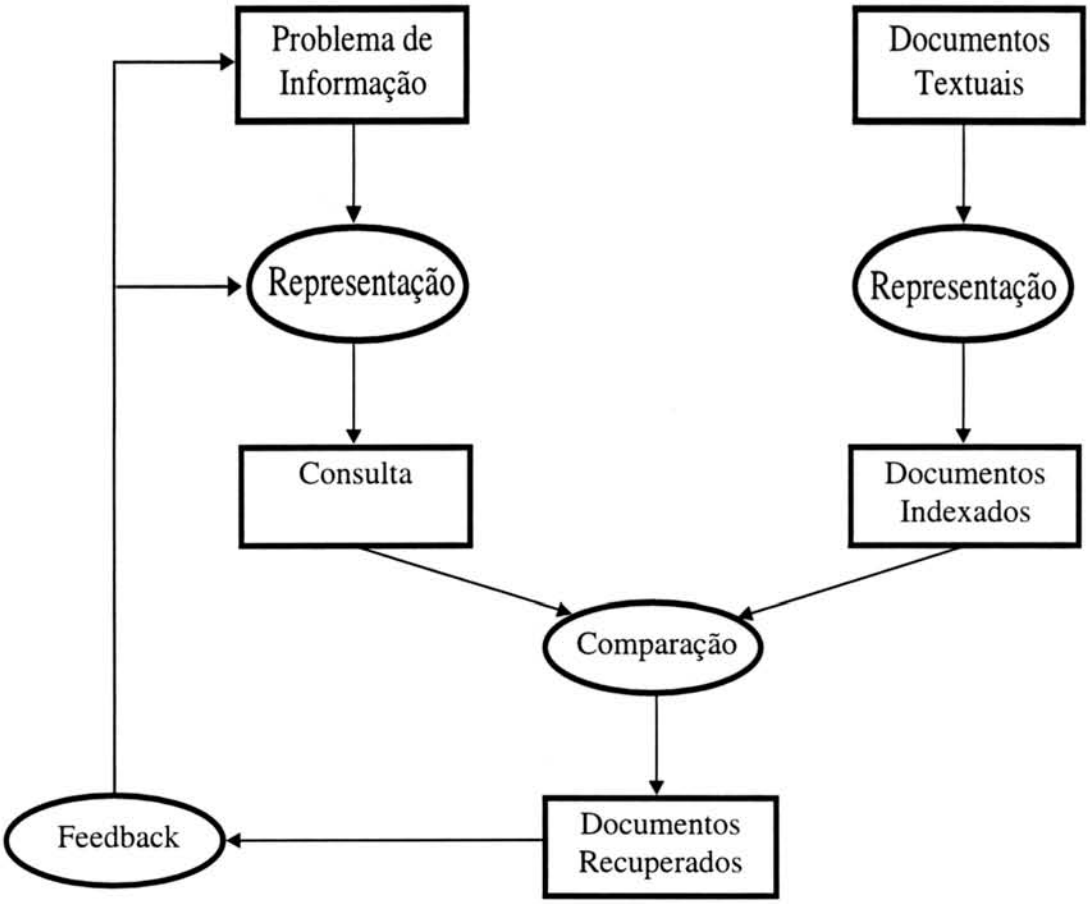

FIGURA 2.2 Processo básico de recuperação de informações [CRO 93]

Quanto a eficácia do feedback, testes realizados, e descritos em [DAN 95], comprovaram que o feedback aumenta a precisão entre 40% e 60%. Maiores informações técnicas sobre o feedback de relevância podem ser encontradas em [SAL 68], [SAL 75], [SAL 83] e [SOE 74]. As equações de ajuste dos pesos podem ser encontradas em [ALL 95] e [PAR 89].

2.9 Problemas encontrados em Sistemas de Recuperação de Informações

Algumas falhas detectadas em SRI's são decorrentes do processo de indexação [RAJ 95], da linguagem de indexação, da formulação da consulta e/ou da interação do usuário com o sistema [SAL 75]. As falhas no processo de indexação estão relacionadas a quantidade de termos utilizada como índice. Quando se utiliza

poucos termos tem-se como resultado uma baixa abrangência da recuperação e, quando se utiliza muitos termos, tem-se uma baixa precisão.

A linguagem de indexação, no caso de estar composta por um vocabulário controlado, pode ser deficiente e não possuir todos os termos necessários. O usuário também pode tornar a recuperação inadequada pela sua má formulação da consulta ou desconhecimento do vocabulário correto para expressar sua necessidade de informação. Muitos usuários finais tem pouca habilidade ou possuem experiência limitada na formulação da consulta inicial ou em modificá-la após observar suas deficiências.

Em uma grande variedade de aplicações, dentre as quais se incluem os SRI's, os usuários devem utilizar as palavras corretas para acessar objetos ou executar ações desejadas. Os usuários novos normalmente utilizam palavras incorretas e, conseqüentemente, falham na sua tentativa. Este problema do vocabulário constitui um empedimento aos usuários tanto nas interações simples, como entrada de comandos, quanto complexas, como uma consulta a um banco de dados. Sendo assim, para que o usuário atinja o seu objetivo o sistema deve reconhecer termos que serão escolhidos espontaneamente [FUR 87].

Se todas as pessoas concordassem sobre como chamar as coisas, a palavra do usuário seria a palavra do projetista, que seria a palavra do sistema, e o que o usuário informasse seria mutuamente entendido. Infelizmente, as pessoas discordam sobre as palavras usadas para denominar coisas [FUR 87].

Os sistemas, em geral, têm um vocabulário limitado e as pessoas normais tem uma grande variação no vocabulário usado, o que se constitui o centro do problema [CHE 94]. Se o que se quer é que o sistema tenha maior chance de "entender" o usuário, uma alternativa seria permitir que o usuário iniciasse o diálogo com suas próprias palavras e o sistema, a partir delas, encontrasse interpretações relacionadas aos termos que está preparado para operar.

Sempre que uma palavra é aplicada a mais de um objeto é gerada uma situação ambígua. Termos menos freqüentes são mais precisos. Freqüentemente, palavras mais populares, mesmo sendo mais prováveis para um dado objeto, são também as mais prováveis para outros objetos. Existe uma variedade muito grande nas formas em que um conceito pode ser expresso [LEW 96]. O fato de se ter mais palavras de acesso a um objeto não implica logicamente em mais objetos por palavra e empiricamente a tendência média aponta para o inverso [FUR 87].

Obviamente os sistemas atuais não oferecem garantias para recuperação inteligente de informações. O próprio usuário do sistema deve conhecer como manipulá-lo a fim de conseguir informação relevante como resultado. Os usuários não estão cientes da dificuldade de se utilizar tais sistemas devido ao fato de acharem que tem obtido sucesso em suas buscas. Porém, estudos mostraram que eles estão enganados neste aspecto. Conforme [GAU 91], os usuários acabam se satisfazendo quando atingem 51% de abrangência na recuperação, o que indica que quase a metade da informação que interessava não foi recuperada. Já [DAN 95], demonstrou que embora os usuários sintam que recuperaram a maioria dos textos corretos, ou seja, que a abrangência da recuperação foi alta, de fato eles apenas recuperaram em torno de

25% dos textos relevantes. Qualquer uma das duas medidas aponta para resultados insatisfatórios.

Mas então, qual o motivo que leva usuários a não conseguirem uma recuperação satisfatória? As respostas encontradas para esta questão estão associadas diretamente ao usuário. A recuperação insatisfatória pode ser devido a:

a) falta de refinamentos sucessivos da consulta, necessidade de maior iteração, ou parada muito cedo da iteração;

b) dificuldades na construção da consulta;

c) problemas no vocabulário da consulta.

Quando o recurso de feedback de relevância está disponível, o que permite um maior refinamento da consulta original, os usuários encontram dificuldade em identificar o que deve ser mudado nela ou qual o ponto de partida a ser considerado para modificá-la.

Uma maneira de melhorar o acesso a informação seria fornecer acesso a um *thesaurus* e permitir que seus termos fossem utilizados para incrementar a consulta. Isto conduziria a uma melhora nos resultados alcançados [LEW 96]. Esta abordagem de utilizar termos de um *thesaurus* visando melhorar os resultados obtidos pela consulta foi aplicada no desenvolvimento do protótipo que será descrito no capítulo 4.

Na prática operacional, as melhores técnicas não aparecem integradas. Quando a forma de consulta usando linguagem natural está disponível métodos de pesagem podem não estar e o feedback de relevância é raramente fornecido.

2.10 MEDLINE

A MEDLINE (Medical Literature Analysis and Retrieval System On Line) é um sistema de recuperação de referências bibliográficas para a área da saúde. Sistemas como a MEDLINE tem como finalidade fornecer recursos para que os profissionais da área da saúde possam buscar soluções para os problemas, relacionados ao cuidado à saúde, com os quais se deparam na prática da medicina diária [BLE 93].

Cada vez mais tem se tornado necessário aos profissionais que atuam na área da saúde buscar, através da literatura, soluções aos problemas emergentes [FLO 95]. Os SRI's destinados à área da saúde surgiram devido a extrema necessidade de métodos adequados para o processamento e disponibilização de uma quantidade muito volumosa e complexa de literatura produzida por pesquisadores em toda a área da biologia.

A base de dados MEDLINE é produzida pela National Library of Medicine. Ela contém mais de 7 milhões de registros [WIL 96] de referências bibliográficas e resumos de artigos de cerca de 3.700 títulos de periódicos internacionais da área da saúde [BAS 95].

No Brasil, a MEDLINE em sua versão para a língua portuguesa pode ser acessada através da BIREME. A base de dados MEDLINE disponível na BIREME corresponde aos registros ingressados nos últimos 7 anos. A BIREME efetua atualizações mensais nessas bases de dados. As referências bibliográficas contidas na MEDLINE estão representadas segundo padrões do MEDLARS Indexing Manual da NLM. O vocabulário adotado pela NLM é o MeSH (Medical Subject Headings).

O MeSH é um vocabulário controlado de termos e frases na área da medicina e biologia e é utilizado para indexar o banco de dados da MEDLINE. Embora ele tenha sido construído de forma a atender de maneira geral a todas as áreas necessárias, o MeSH fornece somente cobertura limitada em sub-áreas altamente especializadas. O vocabulário MeSH está organizado em árvores de tal forma que a especificidade dos termos aumenta à medida que se move em direção as folhas [WIL 96]. Outro vocabulário adotado pela BIREME é o DeCS, o qual é uma tradução para a língua portuguesa do MeSH. Tanto o DeCS quanto o MeSH podem ser utilizados para consultas na MEDLINE [BAS 95]. A figura 2.3 apresenta o acesso remoto de um usuário à BIREME através da Internet.

A MEDLINE oferece recursos que permitem que o usuário selecione as informações que deseja aplicando filtros para reduzir os resultados da recuperação. Alguns filtros podem ser: nome do autor, período de publicação, tipo de publicação, entre outros.

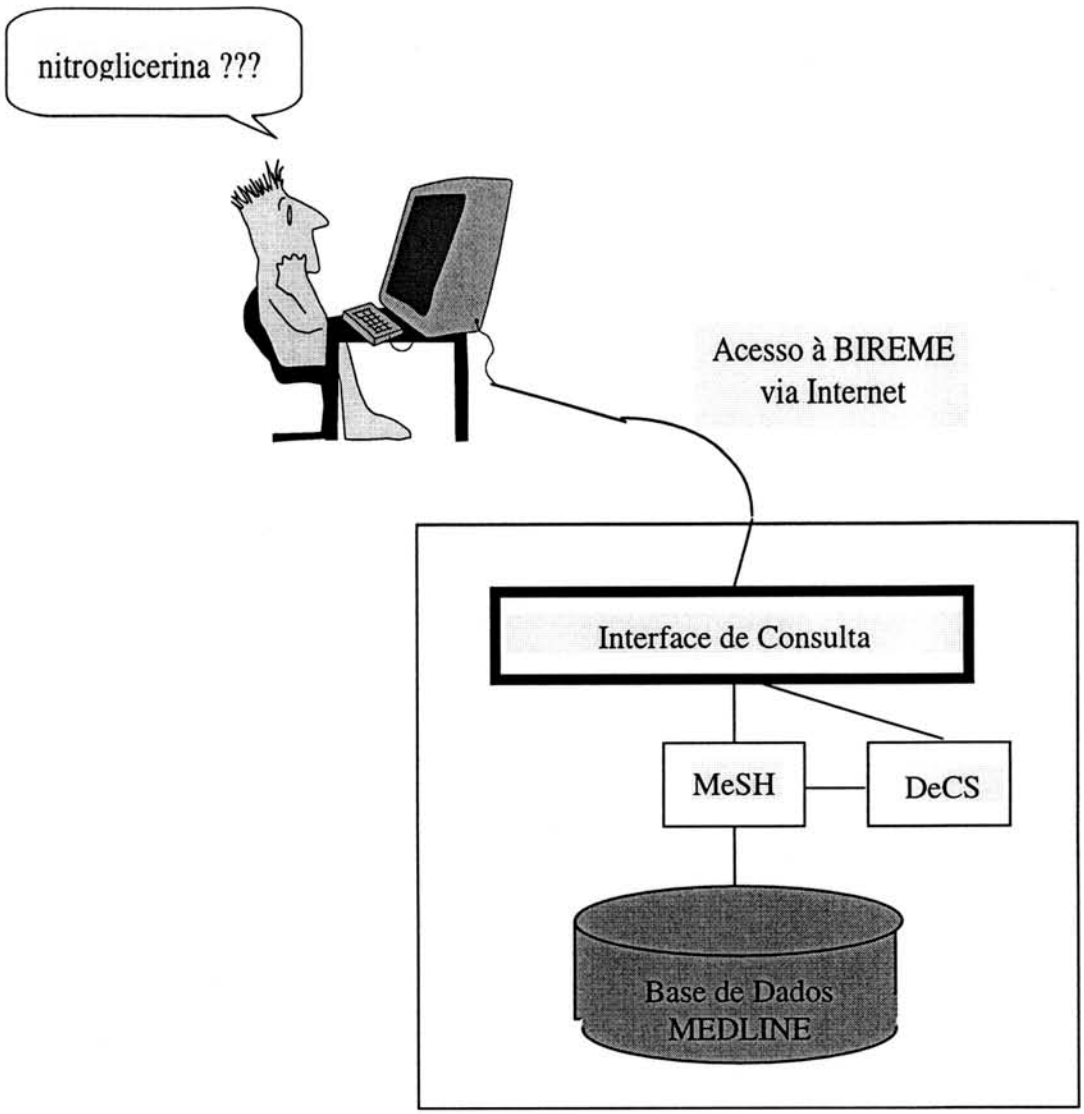

FIGURA 2.3 Esquema de conexão com a BIREME para o acesso à MEDLINE

2.10.1 Interação do usuário com a MEDLINE

Usuários, principalmente iniciantes ou inexperientes, que pretendam obter bons resultados a partir de consultas à MEDLINE devem contar com pelo menos um dos seguintes recursos:

a) intermediário humano que traduza a sua necessidade de informação para termos do vocabulário DeCS;

b) o próprio vocabulário DeCS.

Podem ocorrer situações em que um determinado usuário esteja acessando a MEDLINE via Internet, por exemplo, da sua própria casa. Nesta situação, é pouco provável que ele possa contar com a ajuda de uma pessoa intermediária entre ele o sistema que, por sua vez, possa orientá-lo quanto aos termos mais apropriados para a recuperação das informações que ele necessita. Além disso, mesmo que ele tenha a disposição o vocabulário utilizado na indexação dos dados, como ele é muito

volumoso, a tarefa de busca aos termos pode ser cansativa e dispersá-lo da sua atividade principal, que é a consulta.

O que normalmente ocorre é que os usuários acabam utilizando o método de tentativa e erro, o que ocasiona uma perda muito grande de tempo até que a informação procurada seja encontrada. A MEDLINE não oferece o feedback de relevância ou qualquer outro recurso de apoio a recuperação de referências.

A dificuldade em definir a consulta esta ligada também as diferentes terminologias que podem ser utilizadas em um domínio. O trabalho desenvolvido visa propor soluções aplicáveis a área biomédica na tentativa de permitir que o usuário utilize a MEDLINE contando com recursos que irão apoiar a sua definição da consulta. Para tanto um estudo detalhado sobre as terminologias biomédicas se fez necessário. Esse estudo é apresentado a seguir no capítulo 3.

3 Terminologias Biomédicas

Neste capítulo são discutidos aspectos relacionados a terminologias empregadas na comunicação dentro de uma área específica. Dado o escopo deste trabalho será dada ênfase as terminologias empregadas na área biomédica. Uma breve revisão sobre a origem das terminologias, além dos fatores relativos a padronização de terminologias, são abordados. Aqui também são apresentadas algumas vantagens, bem como desvantagens, que a utilização de terminologias acarreta a um ambiente de trabalho.

3.1 Histórico Evolutivo

O primeiro trabalho formal realizado sobre terminologias da informática médica foi feito pela National Library of Medicine (NLM) [WOL 77]. Inicialmente, por volta de 1960, tentou-se cobrir a informática médica de uma maneira mínima. Em vários anos subseqüentes foram feitas revisões com a finalidade de eliminar termos, adicionar outros e alterar alguns relacionamentos hierárquicos existentes entre os termos.

A primeira realização com a meta de englobar todos os campos da medicina no sentido de classificar os termos médicos foi a Classificação Internacional de Doenças (CID). Nela os termos não estão apenas listados mas também classificados. Após esta iniciativa os vocabulários e classificações biomédicas se multiplicaram, e cada profissão, e suas respectivas especialidades, definiram os seus próprios. Um histórico mais aprofundado sobre o crescimento das terminologias médicas pode ser encontrado em [FLI 95] e [GNA 93].

3.2 Terminologias como Forma de Expressão e o Papel das Terminologias na Informática Médica

A linguagem natural é o mais expressivo e poderoso método de representação do conhecimento disponível atualmente [ROT 93]. Uma terminologia, como qualquer vocabulário técnico, constitui um subconjunto desta linguagem natural. Uma terminologia é uma coleção de nomes de conceitos, ou termos, em algum domínio [TUT 95]. A utilização de uma terminologia visa permitir que idéias sejam expressas de forma mais precisa e clara. Na biomedicina a construção de uma terminologia completa que não apresente qualquer ambigüidade e que descreva todos os fenômenos médicos, e expresse relacionamentos entre termos e conceitos, constitui um grande desafio.

Para a informática, a utilização de uma terminologia possibilita, além da descrição de todos aspectos relativos ao domínio com maior precisão, que sejam desenvolvidas aplicações computacionais que façam uso de um vocabulário que esteja de acordo com o ambiente e público a que se destina.

Atualmente, inúmeras aplicações em todo o mundo voltadas à área médica utilizam alguma terminologia. Como exemplo pode-se citar aplicações de registros médicos, bancos de dados de referências bibliográficas da área médica, bancos de dados médicos factuais, entre outros.

Os problemas ocorrem, e se tornam complexos, quando se tenta desenvolver aplicações que cooperem entre si, pois, comumente, elas são desenvolvidas a partir de terminologias diferentes. É normal que a classificação e o vocabulário reflitam um ponto de vista ou dêem um enfoque maior a uma área específica mas, como nenhuma regra de compatibilidade tem sido mantida entre eles, o processo de cooperação enfrenta alguns problemas.

3.3 As Terminologias Médicas

A terminologia médica é o meio pelo qual pode-se descrever o estado em que um paciente se encontra aos demais membros de uma equipe pela qual este paciente está sendo atendido [MIL 95]. Esta definição constitui apenas uma das inúmeras aplicações onde uma terminologia é atualmente utilizada. O fato é que, dentro da área da saúde, todo o material produzido está descrito através de uma terminologia, um vocabulário controlado. Esta terminologia permite a representação de elementos apropriados, ações e relações entre elementos dentro deste domínio específico [VAN 95].

3.3.1 Desenvolvimento de uma Terminologia Médica

Os procedimentos essenciais para o desenvolvimento de uma terminologia médica, segundo [MIL 95], são três. O primeiro consiste em identificar uma terminologia preliminar, a qual servirá de base para a nova. O passo seguinte consiste em, a partir da colaboração de especialistas, adequar esta terminologia às necessidades do ambiente ao qual ela se destina. Por último, esta terminologia deve ser utilizada durante um certo período de tempo a fim de que os usuários possam identificar falhas ou dar sugestões de modo a adequá-la ainda mais ao ambiente de trabalho.

O objetivo aqui não é apresentar um estudo sobre o desenvolvimento de terminologias ou sequer propor algo neste sentido. Porém, as etapas de desenvolvimento de uma terminologia tornam possível a identificação de falhas que tem seus reflexos sentidos numa etapa posterior, quando ocorre a sua utilização.

Pelos procedimentos para o desenvolvimento de uma terminologia, acima citados, pode-se perceber que em nenhum momento existe a preocupação de manter qualquer tipo de consistência com algum padrão que não aqueles definidos pelo próprio ambiente onde a terminologia será empregada.

3.3.2 Problemas Encontrados nas Terminologias Médicas

A linguagem médica é mais imprecisa e ambígua do que outras linguagens científicas ou técnicas [WOL 77]. Já a alguns anos tem havido esforços no sentido de evitar que este problema aumente. Porém, é fácil perceber que este fenômeno se acelerou nas últimas décadas tornando os problemas de terminologia cada vez mais difíceis de se resolver .

Durante muitos séculos o vocabulário médico vem crescendo lentamente. Muitas áreas da medicina tiveram um avanço estrondoso nos últimos tempos, o que causou um enriquecimento da terminologia médica. Atualmente, um vocabulário médico de consenso geral é de grande importância [FOW 95].

Já em 1977, discutia-se as possíveis dificuldades que as diferentes terminologias, empregadas na medicina, trariam para o desenvolvimento dos sistemas de computação [WOL 77]. Já sabia-se que a comunicação na medicina se tornaria mais e mais inconsistente devido aos fato de que vários profissionais, como autores e professores, utilizavam vários nomes para denominar o mesmo conceito, além do que, alguns destes nomes podiam ser considerados incorretos, confusos ou obsoletos. Foram atribuídos a quatro fatores principais a responsabilidade da existência de diferentes terminologias:

a) especialização: no início do século XX todos os termos do vocabulário médico eram praticamente entendíveis por todos os médicos. Atualmente, devido a hiperespecialização houve uma mudança completa na situação a ponto de que a linguagem de um radioterapeuta não é entendida por um psiquiatra e vice-versa. E até mesmo dentro de uma especialidade alguns termos não tem a mesma aceitação;

b) multidisciplinaridade: a fim de fazer seu diagnóstico ou fazer indicações terapêuticas, o médico deve utilizar um conjunto de disciplinas que não a sua própria, tais como: genética, imunologia e bioquímica, o que pode fazer com que ele venha a utilizar termos muito específicos sem conhecer exatamente o significado real destes termos;

c) o papel das diferentes escolas de pensamento da medicina;

d) a velocidade da evolução da medicina gera um vocabulário de grandes dimensões. A terminologia médica não é precisa e o significado de vários termos está sobrecarregado.

Dados referentes a pacientes podem ser de muitos tipos, tais como informação radiológica, relatórios médicos, sumários de alta, entre outros. Estes dados podem ser originários de locais distintos, tais como laboratórios, consultórios, hospitais, departamentos de saúde pública, entre outros. Normalmente, o que se tem verificado é que para cada tipo de informação e local de origem uma denominação diferenciada é utilizada com a mesma finalidade. Fica bastante claro se imaginar o tipo de problemas a que este fato pode induzir. O mesmo exame realizado em diferentes laboratórios ou o mesmo diagnóstico feito por profissionais com formação diferente pode receber denominações diferentes as quais querem expressar a mesma informação. É natural, então, pensar-se que padrões para estas denominações são

essenciais para todos os tipos de informações utilizadas, tais como nome de drogas, diagnósticos, sintomas, locais anatômicos, procedimentos e muitos outros.[AME 94]. Porém, até o momento, não existe nenhum tipo de regra de padronização a ser seguida.

3.4 Diferenças entre Vocabulários Médicos

Os vocabulário médicos disponíveis hoje se diferenciam basicamente na terminologia empregada, no modo de classificação dos termos (estrutura hierárquica), no escopo e no nível de especificidade ou detalhamento dos conceitos descritos.

Por exemplo, considerando-se vocabulários de drogas em uso, sabe-se que não existe uma terminologia padronizada que descreva todos os nomes de drogas classificadas atendendo todos níveis de granularidade importantes. O que existe são vários vocabulários controlados de drogas, o que cria um problema de sobrecarga de domínio. Eles têm escopos variados e seus termos descrevem drogas em diferentes níveis de especificidade. Por isso, não há mapeamento um-para-um entre os termos, o que complica a identificação dos mesmos conceitos entre vocabulários ([GNA 93] e [ING 95]).

Como comentado na seção 3.2, um vocabulário expressa um ponto de vista, dá enfoque a uma área específica e está voltado para a utilidade a que se destina. Um vocabulário de drogas, por exemplo, pode estar organizado por classe química, ação clínica ou pelo mecanismo de ação.

A nível internacional vários grandes projetos tem tentado formalizar os vocabulários de drogas. Em 1953 a Organização Mundial da Saúde (OMS) começou a desenvolver um vocabulário global de drogas com a intenção de facilitar a troca de informação no amplo mundo das drogas. Seu projeto se estende até os dias de hoje. Para cada nova droga desenvolvida a OMS determina um Nome Internacional Não Proprietário (INN) e o recomenda a todos os seus membros, sendo que eles estão livres para aceitar ou rejeitar cada nome individualmente. O vocabulário dos nomes propostos pela OMS atualmente contém mais de 10.000 termos a nível conceitual de drogas químicas.

Um comparativo feito por [GNA 93], que pode ser visto na tabela 3.1, demonstra as diferenças entre os vocabulários de drogas dentro dos Estados Unidos. Foram selecionados sete vocabulários ou classificações de drogas e procurou-se pelos termos equivalentes em cada um deles. A droga escolhida foi ACETAMINOPHEN e estava presente em todos eles, porém variando o nome e o nível de detalhamento em alguns deles. Embora aqui no Brasil não tenha sido encontrado na literatura um estudo comparativo semelhante sabe-se que este problema também pode ser identificado.

TABELA 3.1-Nomes para a mesma droga em diferentes vocabulários [GNA 93]

Vocabulário	Nomes
COSTAR	ACETAMINOPHEN
INN	PARACETAMOL
USAN	ACETAMINOPHEN
USP	ACETAMINOPHEN TABLETS
	ACETAMINOPHEN CAPSULES
	ACETAMINOPHEN ORAL SOLUTION
	ACETAMINOPHEN ORAL SUSPENSION
	ACETAMINOPHEN TABLETS (CHEWABLE)
	ACETAMINOPHEN WAFERS
	ACETAMINOPHEN SUPPOSITORIES
	ACETAMINOPHEN FOR EFERVESCENT ORAL SOLUTION
SGN	ACETAMINOPHEN
MLDIP	ACETAMINOPHEN
PDR	TYLENOL ACETAMINOPHEN CHILDREN'S CHEWABLE TABLETS & ELIXIR
	TYLENOL ACETAMINOPHEN CHILDREN'S SUSPENSION LIQUID
	TYLENOL, EXTRA STRENGTH, ACETAMINOPHEN ADULT LIQUID PAIN RELIEVER
	TYLENOL, EXTRA STRENGTH, ACETAMINOPHEN GELCAPS, CAPLETS, TABLETS
	TYLENOL, INFANT'S DROPS AND INFANT'S SUSPENSION DROPS
	TYLENOL, JUNIOR STRENGTH, ACETAMINOPHEN COATED CAPLETS, GRAPE AND FRUIT CHEWABLE TABLETS
	TYLENOL, REGULAR STRENGTH, ACETAMINOPHEN TABLETS AND CAPLETS
	APAP-ELIXIR

3.5 Sistema de Linguagem Médica Unificada (UMLS)

Nenhuma terminologia nomeia todos os conceitos importantes na biomedicina [TUT 95]. Uma abordagem para criar uma terminologia biomédica mais completa é unir terminologias biomédicas existentes [CIM 94].

Devido ao fato de que as terminologias existentes podem estar sobrecarregadas, ou seja, uma terminologia pode nomear um conceito também nomeado por outra, as terminologias devem ser unidas. Algumas terminologias sugerem uniões através da sua própria estrutura. Todas as uniões devem ser aprovadas

por uma pessoa com conhecimento apropriado do domínio, um especialista [GNA 93].

Esta abordagem de criar uma terminologia biomédica mais completa através da união de terminologias biomédicas existentes deu origem a diversos sistemas computacionais, dentre os quais se destaca o UMLS (Unified Medical Language System).

A proposta do UMLS é auxiliar os profissionais da saúde e pesquisadores nas atividades de recuperação e integração de informação biomédica eletrônica a partir de uma variedade de fontes, independente da forma como conceitos similares estão expressos e apesar da dispersão da informação útil entre diferentes sistemas de computação [UML 93].

A abordagem UMLS envolve o desenvolvimento de uma máquina *Fonte de Conhecimento*, a qual pode ser usada por uma ampla variedade de programas de aplicação de maneira a compensar as diferenças nas formas em que os conceitos estão expressos. Além disso, ela é útil para identificar as fontes de informação mais relevantes para que usuário dirija uma consulta e para executar os procedimentos de pesquisa necessários para recuperar informação a partir destas fontes. A meta é facilitar ao usuário ligar informações a partir de registros do paciente, bancos de dados bibliográficos, sistemas especialistas, etc. O UMLS está projetado para utilização em uma variedade de ambientes biomédicos e sistemas relacionados a saúde.

O UMLS [UML 96] possui quatro componentes [UML 96d]:

1.*Metathesaurus*: contém informação sobre conceitos biomédicos, sua representação em diferentes vocabulários e *thesaurus*, seu uso e coocorrência em mais de 30 vocabulários biomédicos e classificações. Representa uma variedade de relacionamentos entre termos e suporta mapeamento a partir de termos do usuário para o vocabulário controlado apropriado [UML 96b].

2.Especialista Léxico: contém informação sintática para muitos termos do *metathesaurus*, incluindo verbos, os quais não aparecem no *metathesaurus* [SPE 96].

3.Rede Semântica: contém informação sobre tipos ou categorias de termos do *metathesaurus* ("Doença", "Vírus") e as relações permitidas entre estes tipos ("Vírus" causa "Doença") [UML 96a].

4.Mapa das Fontes de Informação: ou diretório, contém informações sobre o escopo, localização, vocabulário, regras sintáticas e condições de acesso a bancos de dados biomédicos de todos os tipos [UML 96c].

3.5.1 Rede Semântica

O objetivo da Rede Semântica (RS) é fornecer uma categorização de todos os conceitos representados no *metathesaurus* e fornecer um conjunto de relacionamentos úteis entre estes conceitos. Toda informação sobre estes conceitos específicos é encontrada no *metathesaurus*. A rede fornece informação sobre o conjunto de tipos semânticos básicos, ou categorias, os quais podem ser atribuídos a estes conceitos, e eles definem o conjunto de relacionamentos que pode ser mantidos entre tipos

semânticos de alto nível. A versão atual da rede semântica contém 135 tipos semânticos e 51 relacionamentos [UML 96a].

Os tipos semânticos são os nodos da rede e os relacionamentos entre eles são as ligações. Há grupos maiores de tipos semânticos para organismos, estruturas anatômicas, funções biológicas, químicas, eventos, objetos físicos, e conceitos ou idéias. Um aspecto importante é que o nível de granularidade varia através da rede. A intenção é estabelecer um conjunto de tipos semânticos que serão úteis para uma variedade de tarefas. O escopo atual dos tipos semânticos do UMLS é bastante amplo, permitindo a categorização semântica de uma grande proporção da terminologia em múltiplos domínios.

A ligação primária é a "é_um". Ela estabelece a hierarquia de tipos dentro da rede e é usada para decidir sobre o tipo semântico mais específico disponível para atribuição de um conceito do *metathesaurus*. Além disso, um conjunto de relações não hierárquicas potencialmente úteis entre os tipos foram identificadas. Estas estão agrupadas em cinco maiores categorias, as quais são também relações: "fisicamente relacionado a", "espacialmente relacionado a", "temporariamente relacionado a", funcionalmente relacionado a", "conceitualmente relacionado a".

Estas relações são estabelecidas entre nodos de alto nível da rede sempre que possível e são, geralmente, herdadas pela ligação "é_um" por todos os filhos destes nodos.

3.5.2 *Metathesaurus*

O *Meta* 1.4, quinta versão experimental do *metathesaurus* do UMLS contém 26 terminologias diferentes, usa 371.742 ocorrências de termos formando um total de 190.863 conceitos únicos. Um total de 180.879 de ocorrências de termos constituíram as uniões de terminologias. Esses dados são de 1995, a NLM lançará a sexta e a sétima versão em 1996 [UML 96].

O *metathesaurus* mais que dobrou de tamanho entre a primeira e a quinta versão. A sexta e a sétima podem juntas produzir outra dobra. Ele continua sendo ampliado em seu conteúdo e é o componente central do vocabulário da UMLS.

O *metathesaurus* está organizado por conceitos ou significados. Em essência, sua proposta é ligar nomes alternativos e visões do mesmo conceito juntas e identificar relacionamentos úteis entre diferentes conceitos [UML 96b]

3.5.3 Críticas ao UMLS

Segundo [YAN 93] o UMLS tem apresentado uma baixa performance devido aos seguintes fatores:

a) os conceitos principais do UMLS, ou seja, o vocabulário MeSH, não é rico ou preciso o suficiente para identificar o conteúdo dos documentos;

b) os sinônimos e variantes léxicas do UMLS não são ricos os suficiente para cobrir o vocabulário de consultas e documentos.

Descrições de sistemas que utilizam o metathesaurus do UMLS podem ser encontradas em ([MIL 92] e [PEN 93]).

3.6 Vocabulários Médicos no Brasil

Da mesma forma que no resto do mundo, no Brasil também existem vários vocabulários médicos, os quais empregam terminologias diferentes. Alguns deles correspondem a traduções de vocabulários originalmente definidos em outros idiomas, como o CID e o DeCS. Nenhum deles é integralmente aceito ou nomeia todos os conceitos da biomedicina em todos os seus níveis de especificidade.

Como exemplo pode-se citar o vocabulário que o Sistema Único de Saúde (SUS) utiliza para descrever os procedimentos médicos que definem o pagamento que deve ser repassado aos hospitais conveniados. Além deste, existe o vocabulário da Associação dos Médicos do Brasil (AMB) que descreve os convênios dentro das mais diversas áreas, o do Conselho Federal de Farmácia (CFF) que descreve nomes para inúmeras drogas, considerando várias classificações, o da Bireme, utilizado para consultar os seus bancos de dados, entre outros.

O fato é que, no Brasil, como no resto do mundo, cada entidade utiliza os seus próprios termos. A diferença é que em outros países, especialmente nos Estados Unidos, isto constitui uma grande preocupação e grandes esforços tem sido feitos no sentido de desenvolver mecanismos que possibilitem a criação de um vocabulário amplamente aceito e, da mesma forma, completo. A metodologia empregada geralmente consiste em unir os vocabulários em uso, já que não existe um único exemplar que seja totalmente aceito, não existem regras de padronização e nenhuma regra de compatibilidade é obedecida na criação. Já no Brasil, não se tem conhecimento de alguma iniciativa com a finalidade de buscar uma solução semelhante para este problema.

O grande prejuízo que existe devido a inexistência de um vocabulário padronizado ou de mecanismos que possibilitem a integração entre diferentes terminologias no Brasil é a impossibilidade de que aplicações biomédicas desenvolvidas compartilhem e troquem informações. Para que seja possível a integração de sistemas clínicos quase sempre se requer uma fase de tradução, onde vocabulários são comparados e os conceitos similares são casados [ROC 93].

A integração é uma necessidade natural quando se considera as aplicações destinadas a área médica e, principalmente, neste momento em que estuda-se possibilidades para a integração dos sistemas utilizados em diversos hospitais e outras instituições, os quais poderão ter livre acesso à informações contidas em diversos locais. Além disso, na atual sociedade em que vivemos, onde cada vez mais as informações estão integradas e disponíveis é urgente que se busque soluções para estes problemas.

Neste capítulo e no anterior foram apontados uma série de impecilhos associados à utilização de aplicações, tais como: diferenças nos vocabulários de comunicação, falta de funcionalidades necessárias para um bom desempenho do

sistema de pesquisa, falta de padronização nas terminologias, entre outros. No próximo capítulo é apresentado o protótipo de um sistema que propõe soluções para algumas dessas dificuldades. Esse protótipo visa apoiar o usuário no momento da formulação de uma consulta à MEDLINE. Para tanto ele conta com recursos de integração de diferentes vocabulários, os quais permitem que o usuário use qualquer termo na consulta, e funcionalidades que permitem a expansão de uma consulta. Com isso, almeja-se oferecer ao usuário uma ferramenta que o auxilie na formulação de uma consulta à MEDLINE e assim agilize a sua pesquisa por informações.

4 Protótipo de um Sistema para Formulação de Consultas à MEDLINE

Neste capítulo é apresentado o Protótipo de um Sistema para Formulação de Consultas à MEDLINE. Nele são discutidas as principais características do sistema. Inicialmente é dada uma visão geral da aplicação e um exemplo da sua utilização e, em seguida, são descritos seus componentes e aspectos de interação como os diferentes tipos de usuário.

4.1 Visão Geral

O protótipo do Sistema para Formulação de Consultas à MEDLINE foi desenvolvido com o intuito de oferecer auxílio ao usuário no momento da definição de uma consulta à MEDLINE. Mais propriamente ele fornece mecanismos que auxiliam o usuário na definição dos termos que representam a sua necessidade de informação, e que irão compor a sua consulta.

O Sistema para Formulação de Consultas à MEDLINE possibilita a integração de diferentes terminologias médicas, originárias de vocabulários e classificações disponíveis em língua portuguesa e atualmente em uso e, desta forma, permite que um usuário possa utilizar qualquer termo na formulação da sua consulta, não apenas aqueles que a MEDLINE reconhece.

Partindo-se de um vocabulário preferido é feito um mapeamento de termos de outros vocabulários para seus correspondentes no vocabulário preferido. Naturalmente, esta é uma explicação bastante simplista, uma vez que os vocabulários têm uma série distinções se comparados, como visto na seção 3.4.

O vocabulário escolhido como preferido foi o DeCS (Descritores Médicos da Saúde), o qual indexa a MEDLINE [DES 87]. Isto permite que a aplicação desenvolvida se constitua em um auxílio ao usuário no momento da definição de uma consulta à MEDLINE. Mais propriamente ele fornece mecanismos que auxiliam o usuário na definição dos termos que representam a sua necessidade de informação, e que irão compor a sua consulta.

Esta aplicação é composta dos módulos de **Interface do Usuário**, **Mecanismo de Consulta**, **Módulo de Controle**, ***Metathesaurus*** e a **Rede Semântica de Conceitos Médicos**. As suas funcionalidades, bem como a interação que ocorre entre eles, serão apresentadas na seção 4.4.

O protótipo do sistema foi desenvolvido utilizando a ferramenta para construção de aplicações Delphi ([BOR 95, BOR 95a, BOR 95b, MAT 96]). O gerenciador de banco de dados escolhido para armazenar a Rede Semântica e o

Metathesaurus foi o Watcom. A comunicação entre as ferramentas da aplicação e o gerenciador de banco de dados é feita via ODBC (Open Database Connectivity).

4.2 Terminologias Utilizadas

Para o desenvolvimento deste protótipo foram utilizados os vocabulários DeCS e CID9. O DeCS foi selecionado porque indexa a MEDLINE e, portanto, possui os termos que devem compor a consulta. O CID9 foi selecionado por já estar disponível eletronicamente, fator que facilitou a sua inclusão no Metathesaurus.

Do vocabulário DeCS foram selecionados termos, totalizando 2.045 termos, presentes nas seguintes classes:

a) A7 Sistema Cardiovascular;

b) C14 Doenças Cardiovasculares;

c) D18 Agentes do Sistema Cardiovascular;

d) D26 Drogas e Agentes Variados;

e) E3 Anestesia e Analgesia;

f) E4 Técnicas Cirúrgicas;

g) E5 Técnicas Variadas;

h) G2 Ocupações em Saúde;

i) G3 Ambiente e Saúde Pública;

j) G9 Fisiologia Respiratória e Circulatória.

Estas classes foram selecionadas no momento do desenvolvimento do protótipo a fim de restringir a quantidade enorme de termos que o DeCS possui. Do vocabulário CID-9 totalizou-se 5.951 termos.

4.3 Exemplo de Utilização

Antes de descrever individualmente os componentes da aplicação, um cenário exemplo será apresentado para ilustrar como cada componente opera em conjunto buscando fornecer assistência durante o processo de formulação da consulta.

A tela de consulta da aplicação pode ser visualizada na figura 4.1. Nesta tela o usuário tem a sua disposição um conjunto de opções que permitem que ele defina sua consulta através de termos digitados manualmente ou selecionados no *metathesaurus.*

Primeiramente, então, o usuário deve definir sua consulta booleana utilizando a caixa de edição "TERMO", ou procurando termos diretamente no *metathesaurus,* e relacionando-os por meio dos operadores booleanos disponíveis ("E" ou "OU").

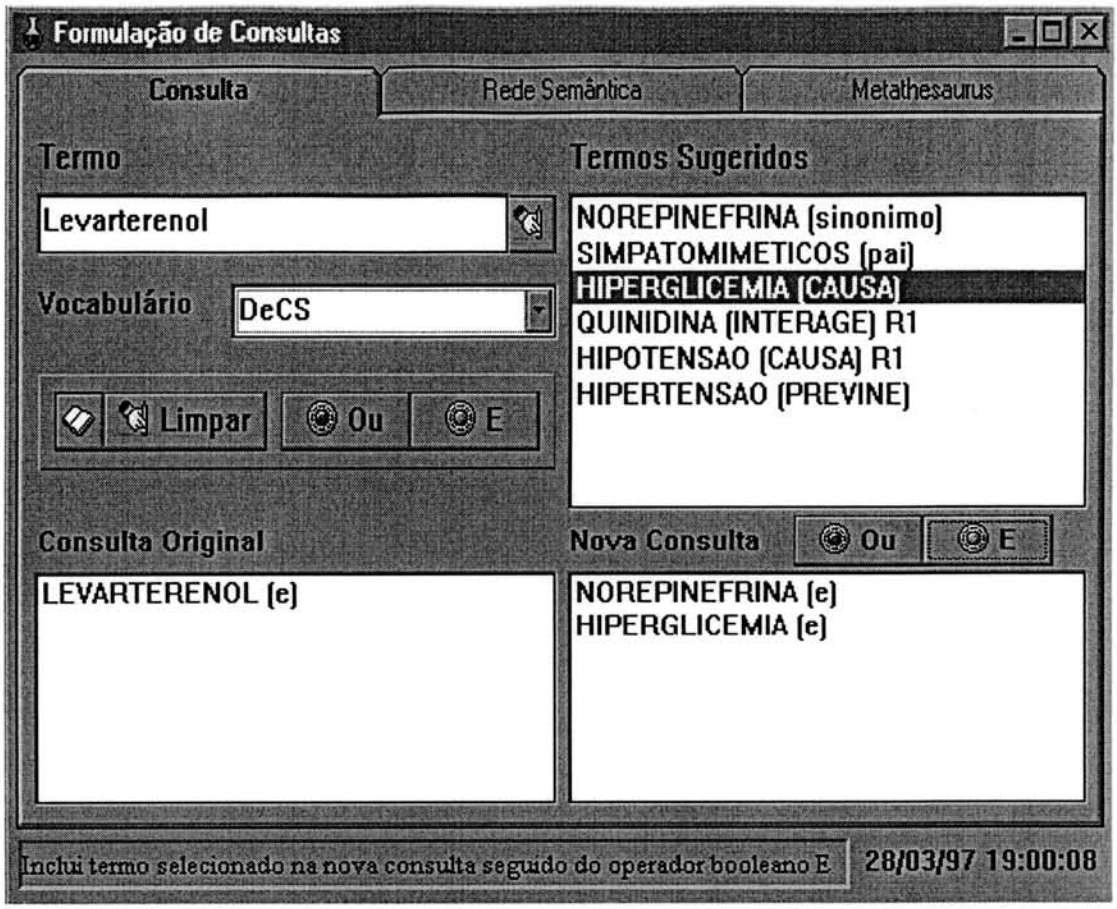

FIGURA 4.1 Tela de consulta da aplicação

A cada termo informado em conjunto com um operador booleano é executado um processamento. Esse processamento é responsável por analisar os termos da consulta e sugerir a substituição ou a inclusão de novos termos permitindo, assim, que a consulta possa ser diretamente aplicada a MEDLINE.

O processamento de cada termo possibilita que conceitos expressos na consulta possam ser generalizados, especializados ou substituídos por outros pertencentes ao vocabulário do sistema a que se destina a consulta.

Neste exemplo, suponha que o primeiro termo da consulta do usuário seja "LEVARTERENOL". O primeiro passo que o sistema realiza consiste em apresentar na caixa "TERMOS SUGERIDOS" uma lista de termos indicados para substituição ou adição a consulta. O primeiro termo da lista indica, neste caso, um termo sinônimo pertencente ao vocabulário DeCS, "NOREPINEFRINA". Esse termo encontrado no *metathesaurus* será utilizado para a localização de outros termos que se relacionem ao conceito que o usuário está pesquisando e que serão sugeridos pelo sistema. Estas sugestões serão posteriormente filtradas pelo usuário que fez a consulta original.

Os termos sugeridos pelo sistema ao usuário são extraídos do metathesaurus e são os seguintes:

a) primeiramente, termos mais genéricos na hierarquia: "SIMPATOMIMÉTICOS";

b) em seguida aqueles que se relacionam ao termo "NOREPINEFRINA" ("HIPERGLICEMIA", "QUINIDINA", "HIPOTENSÃO" e "HIPERTENSÃO");

c) finalmente, os termos mais específicos na hierarquia, sendo que o termo utilizado como exemplo não possui nenhum;

A partir dos termos sugeridos pelo sistema o usuário pode fazer uma filtragem para deixar na "NOVA CONSULTA" apenas aqueles que ele considera que contribuirão com o desempenho da sua pesquisa.

4.4 Arquitetura do Sistema

A figura 4.2 apresenta a arquitetura da aplicação desenvolvida. Esta aplicação contém os seguintes módulos:

a) **Interface do Usuário**: permite a interação dos usuários com a aplicação.

b) **Mecanismo de Consulta**: responsável pelo processamento da consulta do usuário. Este mecanismo recebe a consulta original, envia os termos que a compõem ao módulo de controle e após recebe os termos corretamente mapeados para o vocabulário destino retornando-os ao usuário através da sua interface.

c) **Módulo de Controle**: recebe termos do mecanismo de consulta e procura outros equivalentes ou relacionados a eles no *Metathesaurus*. Controla qualquer acesso ao *Metathesaurus* e a Rede Semântica.

d) ***Metathesaurus***: contém todos os termos conhecidos pela aplicação pertencentes aos mais diversos vocabulários. Ele guarda, para cada termo, a categoria semântica a qual o termo pertence, as relações estabelecidas com outros termos do *metathesaurus*, o vocabulário ao qual ele pertence, entre outras informações de controle.

e) **Rede Semântica de Conceitos**: fornece uma categorização consistente de todos os conceitos representados no *metathesaurus* e um conjunto de relacionamentos úteis entre estes conceitos.

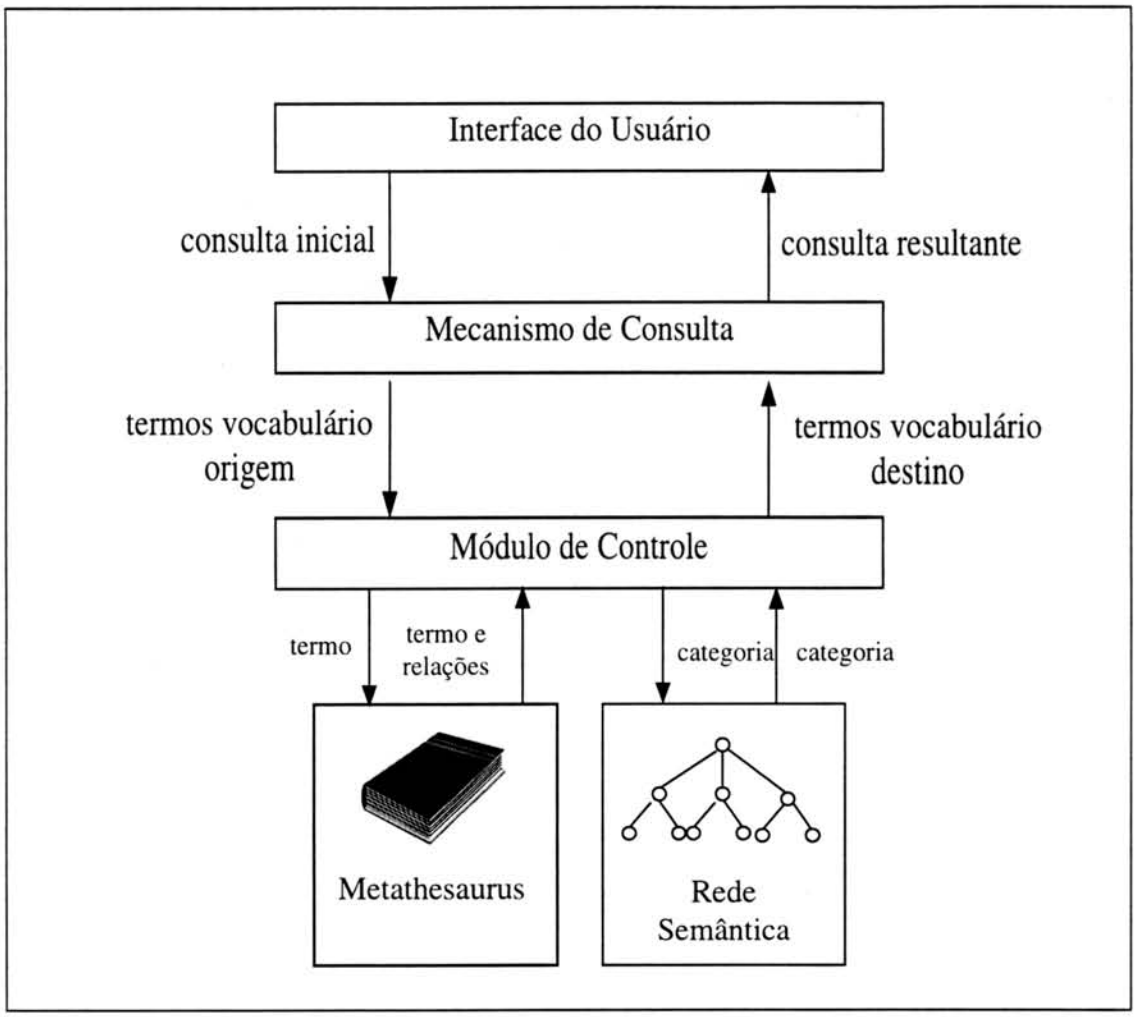

FIGURA 4.2 Arquitetura do Sistema

4.5 Categorias Semânticas

O objetivo da representação do conhecimento, que neste sistema é feita através de uma rede semântica, é nomear, descrever e organizar objetos relevantes ao domínio do conhecimento, além de expressar relacionamentos entre estes objetos, tudo isso de maneira eficiente [ROT 93]. Da mesma forma, uma terminologia médica controlada, adequada para processamento computacional, deve possuir estas qualidades na sua representação [PAT 93a].

Todos os termos médicos, os quais descrevem e indexam eventos médicos, podem ser atribuídos às seguintes categorias semânticas [ROT 93] :

a) topografia : termos anatômicos detalhados;

b) morfologia : termos usados para descrever mudanças estruturais no corpo e, além disso a nomenclatura de tumor encontrado na seção de morfologia do CID-9;

c) função : terminologia química relacionada a sinais e sintomas relacionados com os termos usados para descrever processos bioquímicos e psicológicos;

d) organismos vivos : uma classificação integral do reino animal incluindo bactérias e vírus abrangendo todos os patogens e animais causadores de doenças;

e) drogas químicas : uma compilação dos termos de drogas com referências cruzadas a nomes genéricos e trocados, cada termo está atribuído a sua classe e para drogas compostas os seus constituintes;

f) diagnósticos/doença : arquivos de terminologia clínica do nome de doenças encontradas na medicina incorporando cada entidade encontrada no CID-9;

g) ocupação : lista de ocupações;

h) procedimentos : lista ampla de procedimentos administrativos, terapêuticos e de diagnóstico usados por todo pessoal de tratamento médico em todas as especialidade de cada disciplina médica;

i) geral : ligações sintáticas e termos qualificadores;

j) agentes físicos, forças e atividades : lista de recursos, forças e atividades comumente relacionadas a alguma doença;

k) social : uma lista embrionária de condições sociais relacionadas com herança étnica ou religiosa, status familiar e condições econômicas.

A base de conhecimento da aplicação contém as categorias semânticas acima descritas bem como relações entre estas categorias. A forma de representação do conhecimento selecionada para a representação destas categorias e relações foi a rede semântica. Isto se deve ao fato de que ela se apresenta como uma forma de representação natural para o conhecimento envolvido na aplicação. Além disso, ela permite que pessoas com poucos conhecimentos em informática compreendam facilmente a maneira como o conhecimento está organizado.

4.6 Rede Semântica

A rede semântica representa informações como um conjunto de nodos conectados entre si através de arcos rotulados, que são ligações que representam relações entre estes nodos [RIC 93].

Para este protótipo foi definido que associado a cada conceito presente no *metathesaurus* há uma categorização que especifica os relacionamentos que ele possui com os demais conceitos. A categorização, então, tem a finalidade de atribuir propriedades semânticas a um conceito.

As definições que seguem respeitam a estrutura já conhecida de uma rede semântica e foram definidas especialmente para este protótipo com a preocupação de

formar uma base de conhecimento a ser utilizada para o propósito final a que se destina esta aplicação.

4.6.1 Nodos

Os nodos em uma rede representam conceitos. Um conceito é uma classe abstrata, ou conjunto, cujos membros são objetos do mundo real, instâncias, que estão agrupados pois compartilham propriedades ou aspectos em comum. Cada conceito representado pelos nodos da rede semântica será denominado categoria semântica. O exemplo 4.1 apresenta algumas categorias semânticas.

Exemplo 4.1: As categorias semânticas:

Lesão ou Envenenamento

Doenças ou Síndromes

Material Dental ou Biomédico

Procedimento de Laboratório

representam conceitos e estão presentes na rede semântica.

As categorias semânticas podem possuir especificações ou subcategorias. Uma vez que uma categoria representa um conjunto de instâncias, uma subcategoria representa um subconjunto destas instâncias. O exemplo 4.2 apresenta uma categoria e suas respectivas subcategorias.

Exemplo 4.2:

Estrutura Anatômica

- Estrutura Embrionária
- Anormalidade Congênita
- Anormalidade Adquirida
- Estrutura Anatômica Completamente Formada.

As propriedades de uma categoria correspondem as relações que esta categoria possui com outras categorias. As subcategorias, além das suas propriedades, herdam propriedades da categoria a qual estão associadas. Relações entre categorias e subcategorias são denotadas pela relação de generalização/especialização que será apresentada na seção 4.6.2.

Cada categoria representada na rede semântica tem uma definição associada, útil para determinar mais claramente seu significado. O exemplo 4.3 apresenta a definição da categoria Modelo Experimental de Doença. As definições completas de todas as categorias que formam a rede semântica podem ser vistas no Anexo 2.

Exemplo 4.3: Definição de um conceito presente na rede semântica:

Conceito: Modelo Experimental de Doença

Definição: Uma representação em um organismo não humano de uma doença humana com a proposta de pesquisa nos seus mecanismos ou tratamento.

4.6.2 Ligações

Ligações na rede representam relações entre categorias. Elas são rotuladas para indicar os diferentes tipos de relações.

As relações disponíveis podem pertencer a uma das seguintes classes:

a) Relação de Generalização/Especialização

Esta relação é utilizada para relacionar subcategorias às suas respectivas categorias. É representada pelo rótulo "é um".

b) Relações Não Hierárquicas

Quando duas categorias estão ligadas por uma relação não hierárquica pode-se considerar que para a maioria, se não para todas, das subcategorias da primeira categoria, existe no mínimo uma subcategoria da segunda categoria para a qual o relacionamento é verdadeiro [PAT 93].

Estas relações são fundamentalmente diferentes da relação de generalização/especialização "é um" pelos seguintes motivos:

I. elas especificam que instâncias de uma mesma categoria estão relacionadas a instâncias de outra categoria em oposição a especificar que instâncias de uma categoria estão relacionadas a outra categoria;

II. elas podem não ser verdadeiras para toda instância da categoria.

O exemplo 4.4 apresenta alguns exemplos de relações não hierárquicas.

Exemplo 4.4: Relações não hierárquicas:

afeta

avalia efeito de

causa

A mesma relação pode ocorrer entre múltiplos pares de categorias. Isto é, ligações representando a mesma relação podem ocorrer em muitos lugares na rede, como pode ser visto na figura 4.3. Nela também pode-se verificar que entre o mesmo par de categorias pode ocorrer mais de um tipo de relacionamento.

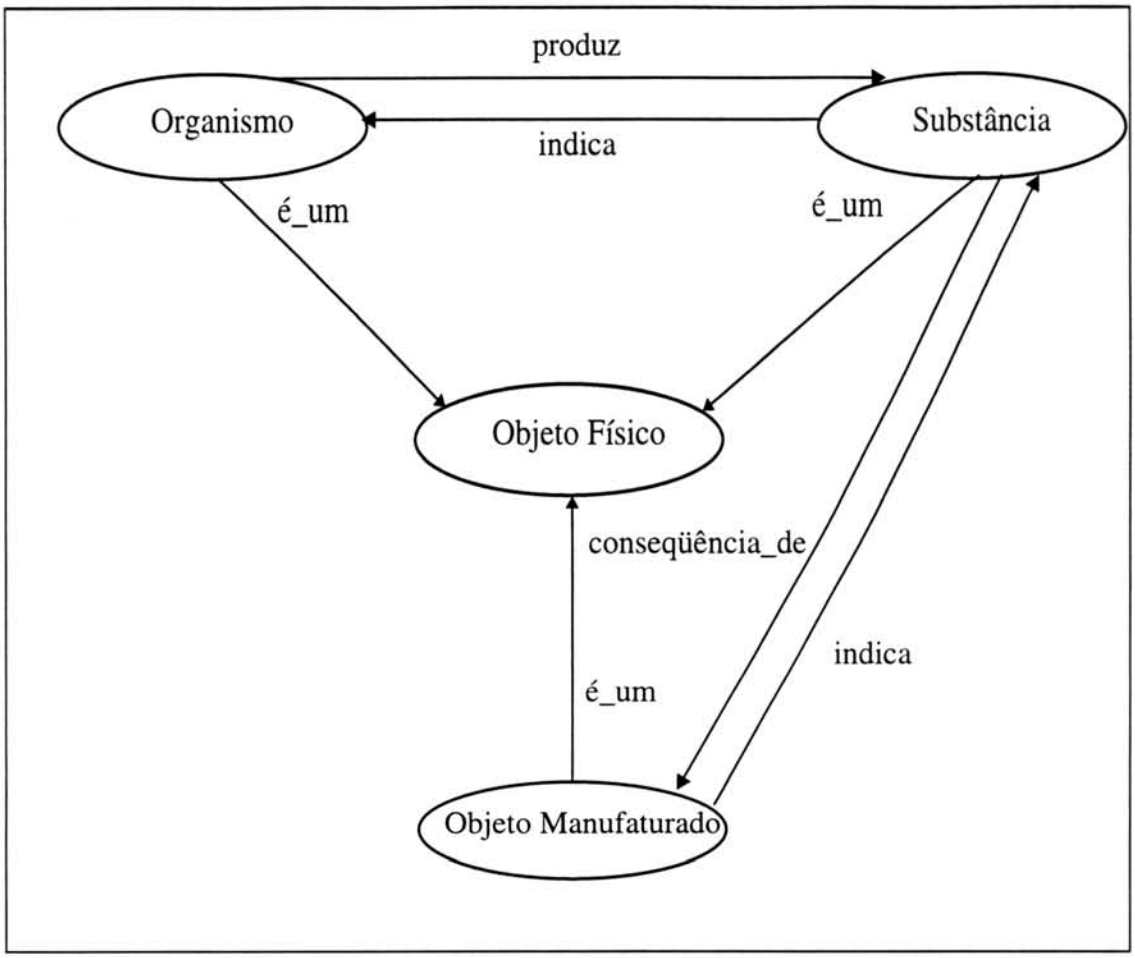

FIGURA 4.3 Exemplo de relacionamentos na Rede Semântica

4.6.3 Relações Inversas

As ligações entre nodos na rede são direcionadas, tendo um único sentido. Assim as relações que elas representam tem também um único sentido. Uma ligação do nodo A para o nodo B representa uma relação da categoria A para a categoria B, mas não representa uma relação da categoria B de volta à categoria A. Na prática, entretanto, as relações vem aos pares. Uma relação de A para B implica na existência de uma relação de B para A, denominada relação inversa [PAT 93]. Uma relação e sua inversa podem compartilhar o mesmo nome ou podem ter nomes distintos.

4.6.4 Propriedades de uma Relação

Cada uma das relações tem como propriedades associadas uma definição e sua respectiva relação inversa. O exemplo 4.5 mostra uma relação e suas propriedades. As propriedades das demais relações disponíveis na rede semântica podem ser vistas no Anexo 1.

Exemplo 4.5: Uma relação e sua propriedades:

Relação: Trata

Definição: aplica um remédio com o objetivo de efetivar a cura ou controlar uma situação.

Relação inversa: é tratado por

4.6.5 Herança de relações

As relações não hierárquicas podem ser herdadas. Uma vez que uma relação pode se aplicar a qualquer instância de uma categoria ela também se aplicará àquelas instâncias pertencentes a qualquer uma das suas subcategorias. Conseqüentemente, uma subcategoria terá todas as relações que a categoria tem. As relações na rede são herdadas pela ligação "é_um" e a condição de herança é definida no momento em que o relacionamento é estabelecido.

Os relacionamentos semânticos são estabelecidos entre categorias e não necessariamente se aplicam a todas as instâncias de conceitos atribuídos às estas categorias. Uma relação que ocorre entre um par de categorias pode não ser verdadeira se ocorrer entre determinados conceitos pertencentes estas categorias. Neste caso, então o relacionamento é dito bloqueado.

A figura 4.4 exibe uma porção da rede semântica onde pode-se verificar as seguintes representações: "Sinal ou Sintoma" é o resultado da ação de uma "Doença ou Síndrome" e "Substância Farmacológica" trata uma "Doença ou Síndrome". Pelo mecanismo de herança a subcategoria "Doença Comportamental ou Mental" herda as relações da categoria a qual esta ligada por uma relação de generalização na rede semântica. Por conseguinte tem-se que "Sinal ou Sintoma" é o resultado da ação de uma "Doença Comportamental ou Mental" e "Substância Farmacológica" trata uma "Doença Comportamental ou Mental".

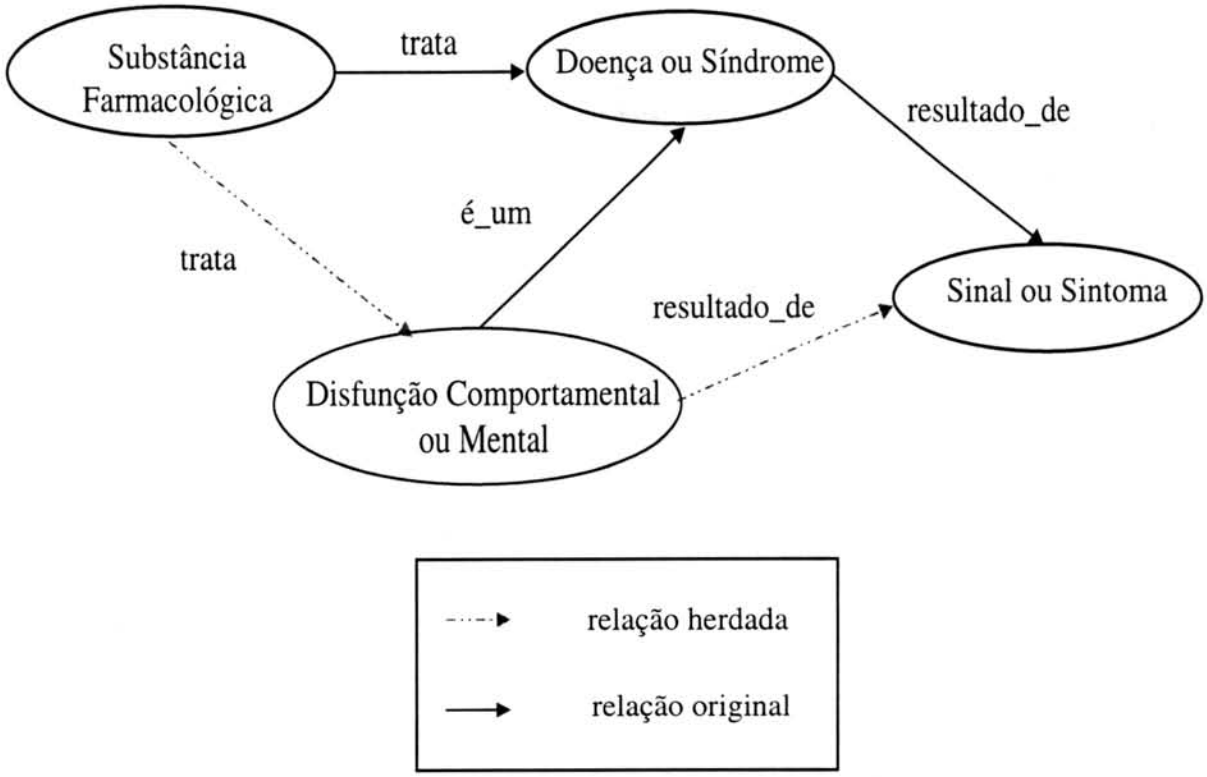

FIGURA 4.4 Herança de relações na Rede Semântica

4.6.6 Autorelacionamento

Um categoria pode manter um autorelacionamento. A figura 4.5 apresenta um exemplo onde um categoria, "Substância Farmacológica", que possui um autorelacionamento, expresso pela relação "interage_com".

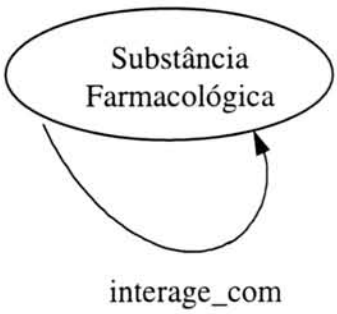

FIGURA 4.5 Exemplo de Autorelacionamento

4.6.7 Relações Ternárias

Uma relação ternária se faz necessária quando é importante representar uma informação que envolva dois ou três categorias diferentes e duas relações, também diferentes.

Suponha que se queira representar a seguinte informação: "duas substâncias farmacológicas são combinadas para prevenir uma doença ou síndrome". A relação "combinação que previne" não pode ser expressa. Como solução, duas relações "interage_com" e "previne" deve ser empregadas. Conforme pode ser observado na figura 4.6 por exemplo, seria necessário criar um relacionamento interage_com entre substância biologicamente ativa e substância farmacológica e, a partir dele, uma relação previne com doença ou síndrome.

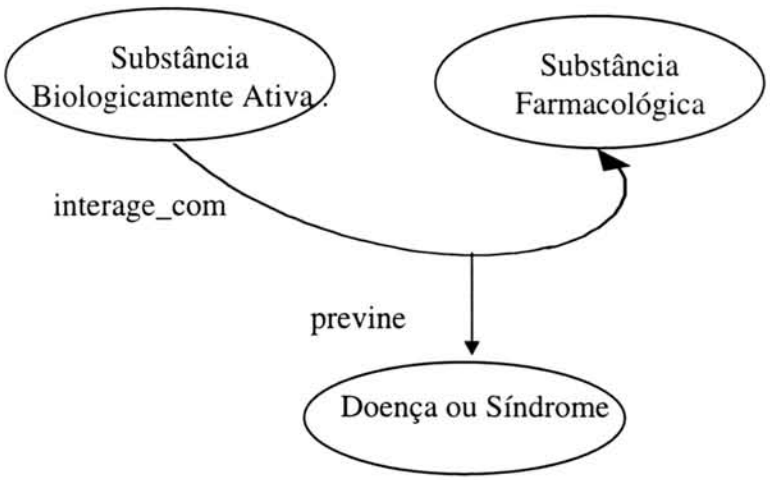

FIGURA 4.6 Relação Ternária "combinação que previne"

4.6.8 Considerações Finais sobre a Rede Semântica

Relações e categorias se definem entre si. Relações não existiriam sem categorias para relacionar, mas um categoria é logicamente indistingüível de outra a menos que ou ela tenha um conjunto diferente de relações ou se o conjunto de relações for o mesmo (o que poderia ser o caso de uma categoria e sua subcategoria) ao menos um categoria (pode ser a subcategoria, por exemplo) tem um valor para uma de suas relações que difere de outro valor da categoria para a relação. Assim, adicionar uma categoria a rede é logicamente justificável se o conjunto de relações for único ou se seu valor para ao menos uma relação diferir de qualquer outra categoria existente tendo o mesmo conjunto de relações. No Anexo III pode-se verificar alguns dos relacionamentos definidos entre categorias da rede semântica.

Na rede da aplicação não são fornecidos nodos para representar instâncias de um categoria, somente nodos para representar as próprias categorias.

Os relacionamentos fornecidos na rede não representam todas as relações possíveis entre as categorias. O que se pretendeu foi gerar um conjunto mínimo de relacionamentos e possibilitar que novos possam ser, da mesma forma, estabelecidos.

4.7 Metathesaurus

O m*etathesaurus* contém os termos e respectivos atributos de um determinado vocabulário. Cada termo é identificado como representando um conceito e pertencendo a, no mínimo, uma categoria semântica da rede. Os termos, oriundos de diferentes vocabulários, são considerados uma forma de expressão de um conceito.

O significado atribuído aos termos do *Metathesausus* é dependente do escopo e da própria granularidade da Rede Semântica. Na maioria dos casos a categoria semântica mais específica disponível na hierarquia é atribuída ao termo. Por exemplo, na rede semântica tem-se a categoria "Mamíferos". O termo "macaco" deve ser associado a "Mamíferos", uma vez que não há outra categoria mais específica associada a "Mamíferos".

A granularidade variável das categorias semânticas tem implicações na interpretação do significado atribuído a um termo. Por exemplo, a categoria "Objeto Manufaturado" tem como subcategorias: "Dispositivo Médico" e "Dispositivo de Pesquisa" , conforme pode ser verificado no Anexo 2. Naturalmente, existem outros objetos manufaturados além dos dispositivos médicos e de pesquisa. Ao invés de se criar subcategorias para os objetos que não são nem dispositivos médicos nem de pesquisa atribui-se diretamente a categoria mais geral "Objeto Manufaturado".

Assim, o significado de cada termo é definido pelo seu vocabulário fonte, pelo contexto em que está inserido dentro da rede semântica, pelos sinônimos e demais relações estabelecidas com outros termos.

4.7.1 Estrutura Lógica

O *Metathesausus* contém dois tipos de termos: termos preferidos, ou primários, e termos secundários. Um termo preferido é aquele que o sistema considera como sendo a melhor palavra ou frase para representar um conceito. Cada termo secundário está associado por uma relação de equivalência a um termo preferido. Um termo secundário corresponde a uma variante léxica do termo preferido correspondente. O termo preferido e seus termos secundários associados tem o mesmo significado semântico. Diferentes termos, com o mesmo sentido, estão ligados ao mesmo termo preferido.

Na tabela 4.1 pode ser verificada a estrutura lógica dos termos no *Metathesausus*. Nela pode-se perceber o termo preferido e os secundários equivalentes a ele. Cada termo contém um conjunto de atributos que identificam seu vocabulário de origem, a sua identificação dentro deste vocabulário e o seu status. Informações comuns a estes termos, como a categoria a qual pertencem e as suas relações com os demais termos, também estão armazenadas no *Metathesausus.*

TABELA 4.1-Estrutura Lógica do Metathesaurus

Termo	Vocabulário	Identificador	Status
norepinefrina	DeCS	D18.222.863.630	preferido
noradrenalina	A	A00	secundário
levarterenol	B	B00	secundário
l-noradrenalina	C	C00	secundário
l-beta-[3,4-diidroxifenil]-alfa-aminoetanol	D	D00	secundário
Categoria	**Relações**		
Substâncias Químicas	causa(hiperglicemia) previne(hipertensão arterial)		

Os termos preferidos pertencem ao vocabulário DeCS, enquanto os termos secundários podem pertencer a qualquer outro vocabulário existente ou ao próprio vocabulário do usuário.

4.7.2 Repetição de Termos

Uma das situações especiais que devem ser previstas na construção do *Metathesausus* diz respeito a repetição de palavras ou frases. Um determinada palavra ou frase pode nomear mais de um conceito. Esta situação ocorre em um número relativamente pequeno de casos, normalmente em diferentes vocabulários do Metathesaurus. Nestes casos devem ser geradas tantas entradas no *Metathesausus* quantos forem os sentidos que estas palavras ou frases podem assumir, ou quantas forem as classificações em que ela pode ser enquadrada em um mesmo vocabulário. Assim, um termo pode pertencer a mais de uma categoria semântica, pois, mesmo que ele tenha apenas um significado semântico ele pode ser classificado em um vocabulário em diversas categorias, dependendo do enfoque desejado.

Na tabela 4.2 pode ser visto um exemplo com termos extraídos do DeCS. Nela tem-se um termo, "Microcirculação", que se repete dentro do próprio vocabulário e aparece classificado em dois contextos. Na primeira classificação o termo "Microcirculação" aparece como um componente do Sistema Cardiovascular e, por este motivo, pertence a categoria "Parte do Corpo, Órgão ou Componente de Órgão". Na segunda classificação este termo aparece ligado a Fisiologia Respiratória e Circulatória, o que indica que ele pertence a categoria "Função de Órgão ou Tecido" [GIL 87]. Isto gera duas entradas no *Metathesaurus*, uma para cada categoria diferente a qual o mesmo termo pode pertencer, de acordo com o contexto considerado.

TABELA 4.2-Exemplo de termo com dupla entrada no Metathesaurus

Sistema Cardiovascular	**A7**
Vasos Sangüíneos	A7.231
Microcirculação	A7.231.432

Fisiologia Respiratória e Circulatória	**G9**
Fisiologia do Sistema Cardiovascular	G9.330
Circulação Sangüínea	G9.330.163
Microcirculação	G9.330.163.645

4.8 Algoritmo da Aplicação

A aplicação utiliza um algoritmo que tem finalidade de acrescentar ou sugerir termos a uma consulta a fim de melhorar os resultados a serem obtidos através dela. O algoritmo é responsável por, a partir da consulta original e da especificação do vocabulário em que se deseja a consulta resultante, gerar uma nova consulta formada apenas por termos do vocabulário solicitado que sejam similares ou mantenham relacionamentos, hierárquicos ou não, com os que estão na consulta original.

Para expandir um conceito seguiu-se a metodologia apresentada na seção 2.6.1. Esta metodologia é composta de técnicas largamente aplicadas em sistemas de recuperação de referências bibliográficas, as quais foram adaptadas a esta aplicação. Os passos considerados para a expansão de um conceito são os seguintes:

1.substituir o termo expresso na pesquisa original pelo seu sinônimo pertencente ao vocabulário do sistema a que se destina a pesquisa;

2.incluir sinônimos deste termo substituto pertencentes ao vocabulário do sistema a que se destina a pesquisa;

3.incluir os termos mais genéricos deste termo encontrados no *metathesaurus*;

4.incluir termos que possuam relações, representadas no *metathesaurus*, com este termo;

5.incluir os termos mais específicos deste termos encontrados no *metathesaurus*.

4.8.1 Algoritmo

Definições para melhor compreensão do algoritmo:

a)Sinônimos(termo, lista_sinônimos, vocabulário): função que recebe como parâmetro um termo e o nome de um vocabulário e retorna uma lista de sinônimos que pertençam ao vocabulário indicado;

b)Pais(termo, lista_pais, vocabulário): função que recebe como parâmetro um termo e o nome de um vocabulário e retorna uma lista com todos os termos mais genéricos do termo que pertençam ao vocabulário indicado;

c)Filhos(termo, lista_filhos, vocabulário): função que recebe como parâmetro um termo e o nome de um vocabulário e retorna uma lista com todos os termos mais específicos do termo que pertençam ao vocabulário indicado;

d)Relações(termo, lista_relações, vocabulário): função que recebe como parâmetro um termo e o nome de um vocabulário e retorna uma lista com todos os demais termos que se relacionam ao primeiro termo e que pertençam ao vocabulário indicado;

e)consulta_nova: lista de termos que comporão a nova consulta;

f)lista_substituição: lista de termos que podem substituir na nova consulta termos da consulta original;

g)lista_inclusão: lista de termos que podem ser incluídos na nova consulta com fins de expansão;

Passos do algoritmo:

1. Separe os termos da consulta dos operadores booleanos;
2. Inicialize lista_substituição, lista_inclusão e consulta_nova;
3. Para cada termo faça:

    3.1 Se o termo está presente no *metathesaurus* então

        3.1.1 Se vocabulário de origem do termo for igual ao vocabulário destino então inclua-o na consulta_nova;

        3.1.2 Se vocabulário de origem do termo não for igual ao vocabulário destino então

            3.1.2.1 Sinônimos(termo, lista_sinônimos, vocabulário);

            3.1.2.2 Insira lista_sinônimos na lista_substituição;

            3.1.2.3 Pais(termo, lista_pais, vocabulário);

            3.1.2.4 Insira lista_pais na lista_inclusão;

            3.1.2.5 Relações(termo, lista_relações, vocabulário)

            3.1.2.6 Insira lista_relações na lista_inclusão;

            3.1.2.7 Filhos(termo, lista_filhos, vocabulário);

            3.1.2.8 Insira lista_filhos na lista_inclusão;

            3.1.2.9 Apresente lista_substituição do termo para que o usuário escolha novo(s) termo(s) e inclua-o(s) em consulta_nova;

3.1.2.10 Apresente lista_inclusão do termo para que o usuário escolha novo(s) termo(s) para expandir a consulta e inclua-o(s) em consulta_nova;

3.2 Se o termo não esta presente no *metathesaurus*

3.2.1 Solicite ao usuário o vocabulário origem do termo;

3.2.2 Solicite ao usuário a categorização do termo na Rede Semântica;

3.2.3 Solicite ao usuário a identificação de termos equivalentes na categoria selecionada, caso haja algum;

3.2.4 Percorra a rede semântica e recupere todos os relacionamentos em que esta categoria está envolvida e também os relacionamentos herdados;

3.2.5 Solicite ao usuário a atribuição de valores completando estes relacionamentos;

3.2.6 Vá para o passo 3.1.1.

3.3 Apresente consulta_nova ao usuário.

4.9 Interação entre o Usuário e a Aplicação

Nesta aplicação foram diferenciados dois tipos de usuários: comum e privilegiado. Isso foi necessário porque algumas atividades que o sistema desempenha, por envolverem conhecimento específico da área médica e biológica, devem ser realizadas por um especialista capacitado a fim de que seja mantida a consistência tanto da Rede Semântica quanto do *Metathesaurus.*

4.9.1 Permissões para o Usuário Comum

O usuário comum tem acesso a informação básica sobre cada categoria presente no *Metathesaurus.* Esta informação básica inclui:

a)nome preferido da categoria;

b)nomes secundários e respectiva origem;

c)categoria semântica a qual pertence o categoria;

d)relações que esta categoria possui com outras categorias;

e)definição desta categoria.

A figura 4.7 apresenta a tela de manipulação do metathesaurus.

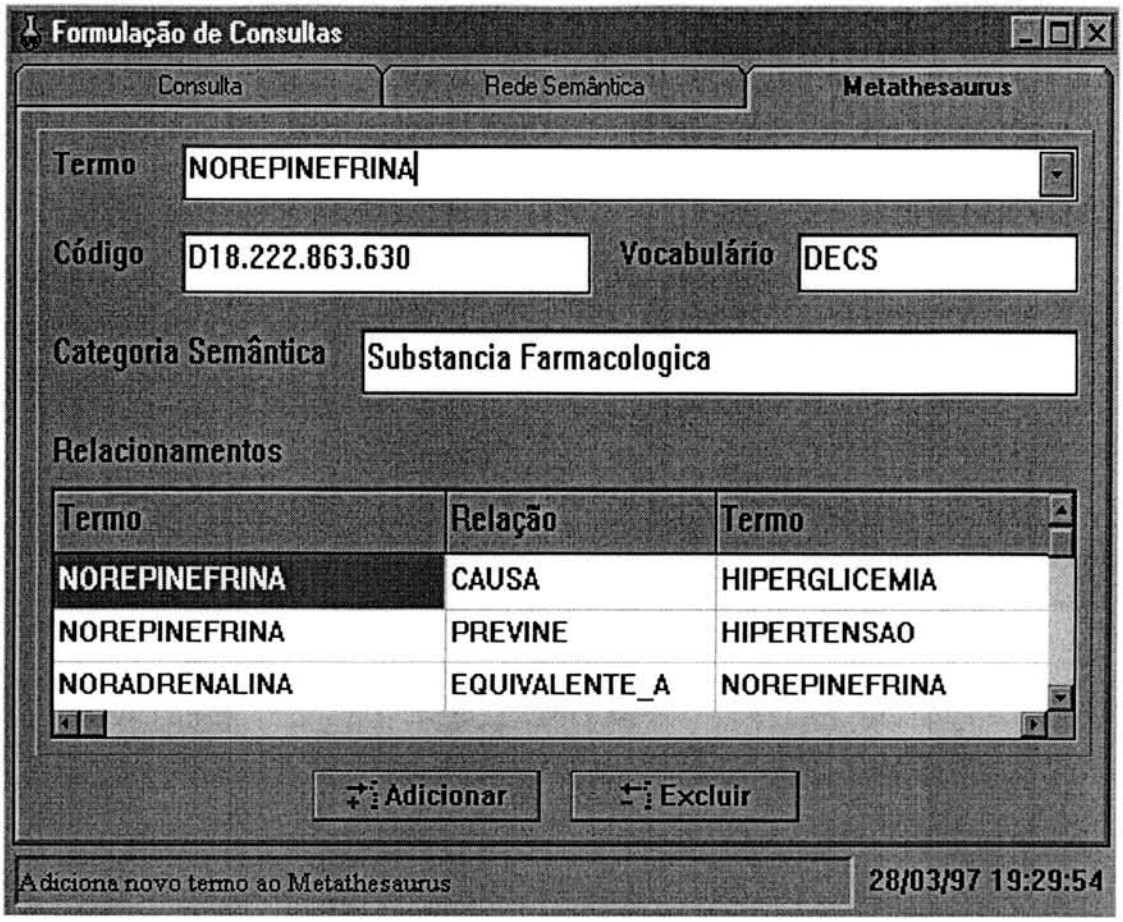

FIGURA 4.7 Tela de manipulação do Metathesaurus

O usuário também tem permissão de fazer consultas sobre a Rede Semântica, tanto a nível de categorias semânticas e relações quanto a nível de relacionamentos. A figura 4.8 apresenta a tela de manipulação da rede semântica.

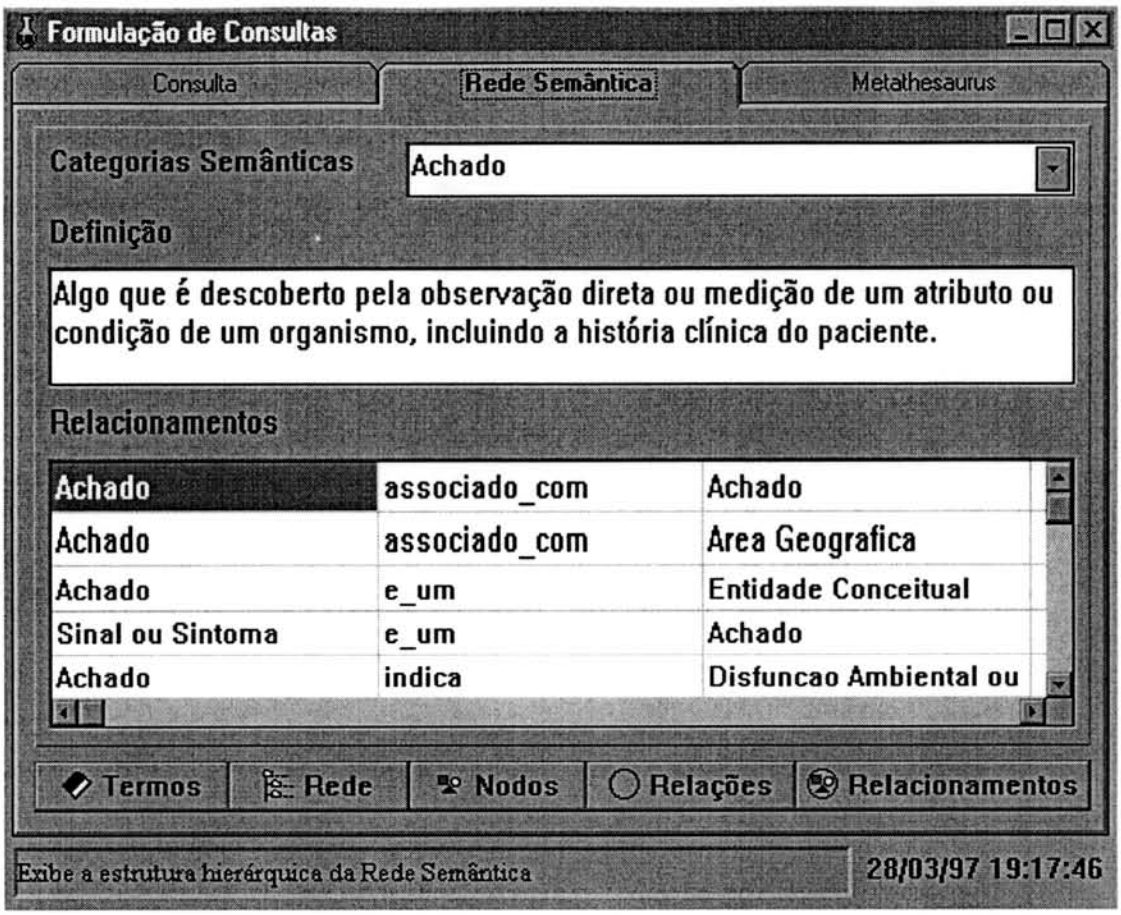

FIGURA 4.8 Tela de manipulação da Rede Semântica

A nível de categorias semânticas e relações ele pode solicitar:

a)a definição de uma categoria semântica;

b)os ancestrais de uma categoria;

c)os descendentes de uma categoria;

d)a definição de uma relação;

e)a inversa de uma relação.

A nível de relacionamentos ele pode fazer consultas nos seguintes formatos:

a)a partir de uma categoria semântica qual (quais) o(s) relacionamento(s) em que ela está envolvida;

b)a partir de duas categorias semânticas quais os relacionamentos em ambas estão envolvidas;

c)a partir de uma relação quais as categorias que se relacionam através dela;

d)a partir de uma categoria e uma relação qual (quais) a(s) outra(s) categoria(s) envolvida(s) neste relacionamento.

De acordo com a atividade a ser executada o usuário pode direcionar o resultado de uma consulta para algum vocabulário específico. Basicamente o usuário pode, especificando o vocabulário em que deseja as respostas:

a) consultar os sinônimos de um determinado conceito;

b) consultar os ancestrais ou os descendentes de um determinado conceito;

c) consultar relacionamentos em que um determinado conceito participe;

d) consultar sobre quais conceitos pertencem a uma dada categoria semântica.

4.9.2 Permissões exclusivas do Usuário Privilegiado

As atividades que devem ficar a cargo do usuário privilegiado são:

a) inclusão de novas categorias semânticas na Rede Semântica;

b) inclusão de novas relações na Rede Semântica;

c) inclusão de novos relacionamentos na Rede Semântica;

d) categorização de novos termos a serem incluídos no *Metathesaurus;*

e) estabelecimento de relações entre termos do *Metathesaurus.*

Cabe ressaltar que coube ao especialista a validação da rede semântica inicial do sistema.

4.10 Experimentação e Validação Preliminar do Protótipo

O protótipo desenvolvido ainda não está rodando de forma integrada a MEDLINE. Isso se deve ao fato de que não foi possível até o momento se obter a disposição a base de dados da MEDLINE para que fossem realizados testes de integração. Sem a base de dados da MEDLINE a validação fica comprometida, visto que ela depende da análise dos resultados obtidos através da consulta.

Pelo desempenho de sistemas que utilizaram as mesmas abordagens empregadas neste protótipo e pelos recursos que ele oferece acredita-se que ele será de grande auxílio na tarefa de definição e formulação da consulta, porém, não foi possível até o momento se obter valores numéricos que indiquem de quanto será a melhora.

Pelos estudos desenvolvidos e pela experimentação do protótipo pode-se concluir que a consulta resultante do sistema será melhor que a original pelos seguintes fatores:

a) a consulta resultante não conterá termos que a MEDLINE não reconheça como índices de suas referências;

b) com a inclusão de novos termos na consulta a necessidade de informação do usuário pode ser melhor definida, permitindo que ele especialize ou generalize os resultados esperados;

c) através do acesso ao *metathesaurus* o usuário pode livremente procurar os termos que definem a sua necessidade de informação. A utilização do metathesaurus auxiliará aquele usuário "não sei o que quero mas vou saber quando encontrar", ou seja, o usuário que encontra dificuldade em definir a informação requerida. Após algumas interações, mesmo que sem sucesso, eles passam a compreender como o domínio está organizado e aprendem sobre o assunto.

Conforme dito anteriormente, os usuários acabam se satisfazendo com a informação recuperada muito antes de atingirem os 100% de abrangência da recuperação desejados. Um dos fatores que contribuem para isso é a dificuldade na formulação da consulta, especificamente na escolha dos termos que a compõem.

Refinar uma consulta é uma atividade difícil de ser realizada se não houver nenhuma ferramenta que a apoie e a MEDLINE não oferece recursos de feedback de relevância. Este protótipo, pela sugestão de termos mais gerais, mais específicos e termos relacionados pretende oferecer subsídios ao usuário para que ele possa refinar a sua consulta. Cabe ao usuário a responsabilidade do refinamento.

4.11 Trabalhos Relacionados

4.11.1 Integração de Terminologias

O principal trabalho relacionado a área de integração de terminologias médicas é o UMLS, descrito na seção 3.3. Dele, esta aplicação herdou algumas qualidades, visto que ambos apresentam algumas das mesmas propostas e, por isso, nada mais justo do que considerar-se a experiência internacional no desenvolvimento de sistemas deste porte.

A partir dos estudos desenvolvidos achou-se conveniente seguir-se, conforme o UMLS, a aplicação das seguintes abordagens:

a) utilização de uma rede semântica como forma de representação do conhecimento médico;

b) utilização de um metathesaurus para armazenamento de cada termo e suas relações.

Além disso, considerou-se, para a construção da rede semântica, as seguintes definições do UMLS:

a) 115 categorias semânticas, as quais são utilizadas para categorização de conceitos na rede semântica;

b) 41 relações semânticas, as quais representam relações entre as categorias semânticas da rede.

4.11.2 Recuperação de Informações

Alguns sistemas especialistas para auxiliar os usuários na seleção dos termos corretos para realizar pesquisas em bancos de dados tem se destacado dentro da recuperação inteligente de informações ([PAR 89], [SHO 85]). Eles possuem uma base de conhecimento formada por palavras, conceitos, frases e relações semânticas representada na forma de uma rede semântica. Regras de decisão são utilizadas para localizar termos no vocabulário apropriado na rede semântica e eles são sugeridos ao usuário para possível expansão da consulta. Estas regras são baseadas em descrições e observações de especialistas na prática de pesquisa por informações. O termo inicial do usuário é localizado na rede semântica. Termos candidatos são identificados pela expansão do nodo através das ligações que partem dele para outros. Termos que estão ligados até a uma distância de dois nodos são recuperados e apresentados para o usuário. Ele então decide se estes termos candidatos são ou não relevantes e se devem ou não substituir os termos para os quais apontam. Este processo pode ser estendido até que nenhum novo termo possa ser sugerido.

5 Conclusões e trabalhos futuros

5.1 Conclusões

Um SRI tem como meta localizar documentos relevantes em resposta a uma consulta do usuário [KRO 92]. A recuperação de informações tem sido investigada por muitas décadas por pesquisadores das ciências da informação, biblioteca e computação. Este campo vem se tornando cada vez mais importante devido a quantidade crescente de informação eletrônica e aplicações baseadas em texto disponíveis e dos avanços nas tecnologias de redes de computadores, armazenamento de informações e multimídia. Métodos eficientes de recuperação de informações são extremamente importantes nos dias de hoje para garantir que a necessidade de informação do usuário será preenchida.

Os sistemas comerciais estão quase exclusivamente baseados no modelo de recuperação determinístico, as consultas são feitas através de expressões booleanas e o resultados são exibidos em uma lista ordenada de forma decrescente considerando-se a "similaridade" que o item recuperado possui com a consulta.

Alguns sistemas permitem que se atribua pesos aos termos que compõe a consulta, embora não seja muito comum, e o feedback de relevância, apesar de bastante comentado, não tem sido muito empregado na prática. Na verdade, os usuários encontram grande dificuldade em utilizar qualquer um destes dois recursos.

Com a popularização da Internet milhões de usuários do mundo todo passaram a ter acesso a uma quantidade extraordinária de informações. Esta situação fez com que ficasse mais do que comprovada a necessidade de SRI's capazes de desempenhar suas funções com o máximo de eficiência. Contrariamente a isso, o que se verifica nos dias de hoje é que os sistemas disponíveis, via Internet e para uso na Internet, têm muito ainda a evoluir a fim de preencher todos os requisitos ideais. Naturalmente, esse sistemas são bastante úteis e não há como dispensá-los. O problema é que muito ainda pode ser feito através da incorporação de funcionalidades que permitam a filtragem de informações, realizem a disseminação dessas informações, considerem o perfil do usuário na realização de qualquer operação, permitam a integração com outras aplicações, entre outras que irão surgir.

Este trabalho apresentou um estudo geral sobre os SRI, enfocando todos os processos envolvidos e relacionados ao armazenamento, organização e a própria recuperação. Posteriormente, foram destacados aspectos relacionados aos vocabulários e classificações médicas em uso, úteis para uma maior compreensão das dificuldades encontradas pelos usuários durante a interação com um sistema com esta finalidade. Por fim, foi apresentado o Protótipo do Sistema para Formulação de Consultas à MEDLINE, bem como seus componentes e funcionalidades.

Neste trabalho foi discutida a dificuldade que os usuários encontram em formular uma consulta à MEDLINE. Foi visto também que, em se tratando de um SRI

para a área médica, a dificuldade se expande em conseqüência das diferenças de terminologias utilizadas nesta área. Por este motivo, na maior parte das vezes, os usuários não alcançam resultados satisfatórios através das consultas formuladas tendo, então, que contar com o auxílio de um intermediário.

Neste momento cabe destacar que o sistema para Formulação de Consultas a MEDLINE não é destinado em hipótese alguma a intermediários, uma vez que, eles mesmos não necessitam das informações que o sistema presta. Nas diversas oportunidades em que se teve contato com pessoas que atuam como intermediários de pesquisa à MEDLINE, como bibliotecárias por exemplo, este fato ficou bastante claro pois eles conhecem profundamente o vocabulário DeCS em função do tempo que já vem utilizando essa ferramenta.

Através desse estudo realizado pode-se verificar que a linguagem de comunicação entre o usuário e o sistema é extremamente importante no sentido de permitir que o usuário tenha sucesso na realização da sua pesquisa. A sua escolha tem um impacto significante na eficiência do sistema.

O protótipo do Sistema para Formulação de Consultas à MEDLINE foi desenvolvido com o intuito de permitir que o usuário utilize qualquer termo na formulação de uma consulta destinada a MEDLINE. Ele possibilita a integração de diferentes terminologias médicas, originárias de vocabulários e classificações disponíveis em língua portuguesa e atualmente em uso.

Esta metodologia de unir terminologias vem sendo amplamente empregada e os resultados obtidos através de sua aplicação (habitualmente um *metathesaurus*), têm sido aproveitados por vários sistemas ou na integração de aplicações com diferentes propósitos. Muitos países têm investido em pesquisas neste campo, sobretudo os Estados Unidos. No Brasil, onde as mesmas necessidades podem ser verificadas, não foi encontrada nenhuma referência sobre pesquisas neste assunto.

Cabe ressaltar que os produtos desenvolvidos em outros países não se adequariam a nossa realidade, uma vez que nosso idioma é o português e, consequentemente nossos vocabulários também. Obviamente, a metodologia de desenvolvimento desses sistemas foi considerada.

Para o desenvolvimento deste protótipo buscou-se agregar abordagens de desenvolvimento empregadas em aplicações de integração de terminologias e aplicações de apoio a formulação de consultas que, pela validação dos respectivos sistemas, apontaram resultados positivos. O UMLS foi um desses sistemas. Ele foi projetado para ser uma fonte de conhecimento a ser usada por desenvolvedores de sistemas. Através de uma interface de consulta ele pode ser acessado, consultado e, dessa forma, utilizado um inúmeras aplicações.

Com este protótipo pretendeu-se definir uma aplicação que integrasse uma fonte de informação para a língua portuguesa e uma interface com funcionalidades especiais que auxiliassem o usuário na formulação de uma consulta à MEDLINE.

Dentre os componentes da aplicação desenvolvida a Rede Semântica de Conceitos Médicos e o Metathesaurus são especialmente responsáveis pela integração dos vocabulários. Essa integração não é, em momento algum, uma atividade simples. Muito pelo contrário, ela é uma atividade complexa e de muita responsabilidade, pois dela depende o correto funcionamento do sistema que posteriormente utilizará o vocabulário resultante.

Além disso, a integração de termos a partir de diferentes vocabulários consome muito tempo. A sua manutenção também deve ser levada em conta pois um vocabulário está sempre crescendo. Isto pode ser verificado pois, mesmo no vocabulário empregado pelas pessoas no dia-a-dia novas palavras estão sempre surgindo e, dentro de um domínio como a medicina, muito maior é a incidência de casos. Novas doenças vão surgindo e com elas novos diagnósticos, para combatê-las novas drogas vão sendo criadas, novos procedimentos vão sendo utilizados e tudo isto implica diretamente no aumento do vocabulário médico.

Segundo [SAL 75], um *thesaurus* nunca está completo. A mesma afirmação cabe a um metathesaurus, já que ele pode ser considerado uma agregação de vários *thesauri*. Ele deve ser atualizado continuamente baseando-se na experiência em sua aplicação prática a fim de refletir os mais recentes desenvolvimentos na área para a qual se destina.

Segundo [FOX 93], um terço do tempo do pesquisador é gasto em atividades de pesquisa, tais como: seleção da fonte a ser pesquisada, formulação da consulta, avaliação dos resultados obtidos e reformulação da consulta. Esta aplicação, destinada especificamente à MEDLINE, visa apoiar o usuário na etapa de formulação da consulta, pretendendo, dessa forma, contribuir para uma melhora global na performance da utilização desse sistema.

5.2 Trabalhos Futuros

Estudos sobre formas de integração da aplicação desenvolvida com a MEDLINE deve ser feitos, uma vez que não foi possível neste momento por questões de tempo hábil.

Na definição da Rede Semântica não foram esgotados todos os relacionamentos possíveis entre as categorias semânticas existentes, o que poderia ser feito para aumentar a semântica dos conceitos representados.

Novos vocabulários devem ser inseridos no Metathesaurus para que se tenha uma abrangência maior de terminologias.

Uma grande dificuldade encontrada no desenvolvimento de sistemas de informática para a área médica é a falta de padrões de linguagem, os quais permitiriam que desenvolvedores de tais sistemas compartilhassem dados sobre pacientes facilmente [AME 94]. A adoção de tais padrões possibilitaria a integração [FOW 92] e transferência automática de informações médicas, acelerando o atendimento e reduzindo a duplicação de exames e prescrições, permitiria a cobrança eletrônica, reduzindo custos administrativos, entre outras vantagens. Como eles atualmente não existem, uma alternativa seria a utilização de um vocabulário geral o bastante, como o contido no metathesaurus, o qual permitiria que as atividades acima mencionadas pudessem ser executadas.

Um trabalho futuro seria estudar a viabilidade do uso do Metathesaurus gerado na integração de sistemas dentro do domínio médico. Uma idéia inicial de como poderia ser esta integração pode ser vista na figura 5.1.

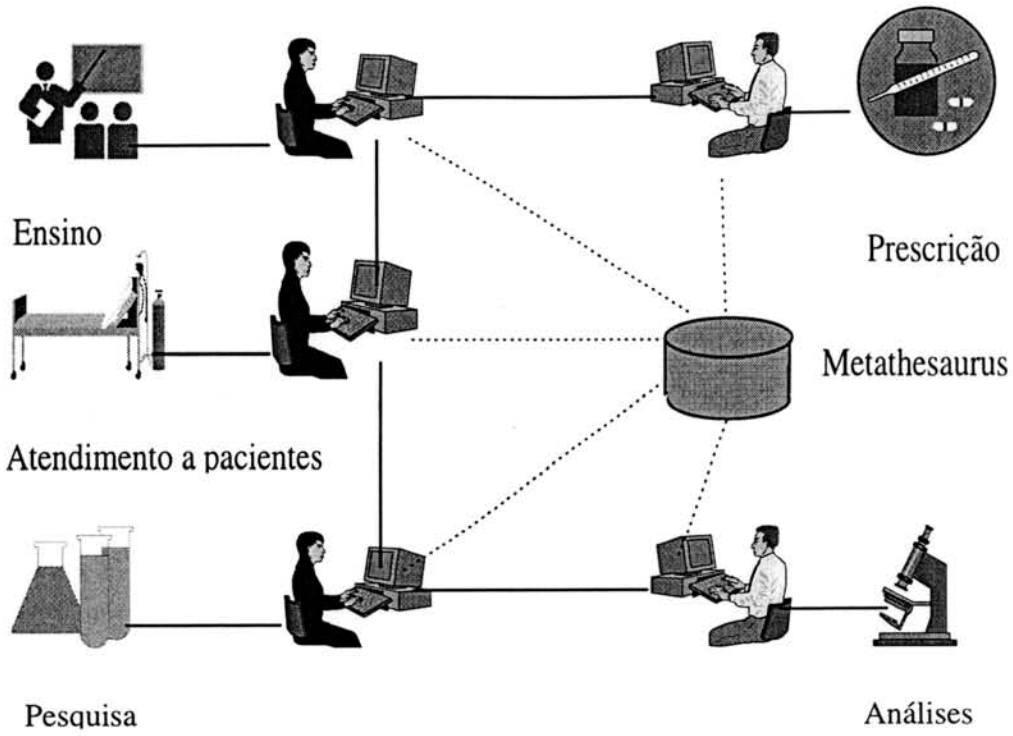

FIGURA 5.1 Integração de aplicações médicas através do Metathesaurus

Para finalizar, a validação desta aplicação seria uma tarefa bastante importante no sentido de comprovar-se a sua eficácia em um ambiente real, uma vez que ela está, até o momento, baseada na validação positiva de outras aplicações do gênero.

Anexo 1 Relações Básicas

A seguir são descritas as relações disponíveis para a construção da rede semântica seguidas pela sua definição e pela sua respectiva relação inversa. Estas relações foram extraídas da Rede Semântica do UMLS.

a) é_um: relação de generalização. Relaciona uma subclasse a sua superclasse.

   Relação Inversa: é_um (especialização)

b) equivalente_a : sinônimo de.

   Relação Inversa: equivalente_a

c) afeta: produz um efeito direto sobre. Implicado aqui é a alteração de uma condição, estado, situação ou entidade existente. Esta relação inclui as relações tem papel em, altera, influencia, predispõe, catalisa, simula, regula, diminui, impede, aumenta, contribui para, conduz a e modifica.

   Relação Inversa: é_afetado_por

d) avalia_efeito_de: analisa a influência ou conseqüência da função ou ação de.

   Relação Inversa: tem_seu_efeito_avaliado_por

e) associado_com: tem um relacionamento significativo com.

   Relação Inversa: associado_com

f) executa: executa um função ou desempenha um procedimento ou atividade. Ela inclui opera sobre, manipula e realiza.

   Relação Inversa: é_executado_por

g) causa: ocasiona uma condição ou efeito. Um agente, tal como uma substância farmacológica ou um organismo causa um efeito. Esta relação inclui induz, evoca e efetua.

   Relação Inversa: é_causado_por

h) coocorre_com: ocorre ao mesmo tempo de ou junto com. Inclui coincidente com, é concorrente com, é contemporâneo a, acompanha, coexiste com e é concomitante com.

   Relação Inversa: coocorre_com

i) complica: torna mais severo ou complexo ou resulta em efeito adverso.

   Relação Inversa: é_complicado_por

j) parte_conceitual_de: conceitualmente uma porção, divisão ou componente de alguma porção maior.

Relação Inversa: componente_conceitual_de

k) conceitualmente_relacionada_a: relacionado com algum conceito abstrato, pensamento ou idéia.

Relação Inversa: conceitualmente_relacionada_a

l) conectado_a: diretamente atado a outra unidade física como tendões estão conectados aos músculos.

Relação Inversa: conectado_a

m) consiste_de: estruturalmente composto de algum material ou substância no todo ou em parte. Esta relação inclui composto de, feito de e formado de.

Relação Inversa:compõe

n) contém: mantém ou é o receptáculo para fluidos ou outras substâncias. Esta relação inclui é preenchido com, mantém e é ocupado por.

Relação Inversa: está_contido_em

o) grau: a intensidade relativa de um processo ou a intensidade relativa ou quantidade de uma qualidade ou atributo.

Relação Inversa: grau

p) derivativo_de: na química, uma substância estruturalmente relacionada a uma outra ou que pode ser feita a partir de outra substância. Esta relação é utilizada apenas para relacionamentos estruturais e ela não inclui relacionamentos funcionais tais como: *meta*bólito de, pelo produto de, nem análogo de.

Relação Inversa: deriva

q) forma_desenvolvimental_de: estágio inicial na maturação individual de.

Relação Inversa: desenvolvido_por

r) diagnóstico: distingue ou identifica a natureza ou características de.

Relação Inversa: diagnóstico

s) rompe: altera ou influencia uma condição, estado ou situação já existente. Produz um efeito negativo sobre.

Relação Inversa: rompido_por

t) avaliação_de: julgamento do valor ou grau de algum atributo ou processo.

Relação Inversa: avaliação_de

u) exibe: mostra ou demonstra.

Relação Inversa: exibido_por

v) funcionalmente_relacionado_a: relacionado pela execução de alguma função ou atividade.

Relação Inversa: funcionalmente_relacionado_a

w) indica: dá evidência da presença em algum período de alguma entidade ou processo.

Relação Inversa: indicado_por

x) interage_com: age, funciona ou opera junto com.

Relação Inversa: interage_com

y) interconecta: serve para ligar ou juntar duas ou mais unidades físicas.

Relação Inversa: interconecta

z) enfoca: é um aspecto ou ponto de discussão, estudo, debate ou disputa.

Relação Inversa:

aa)localização_de: a posição, lugar ou região de uma entidade ou local de um processo.

Relação Inversa: se_localiza

bb)gerencia: administra ou contribui para o cuidado de um indivíduo ou grupo de indivíduos.

Relação Inversa: gerenciado_por

cc)manifestação_de: aquela parte de um fenômeno a qual é diretamente observável visivelmente expressada ou a qual dá evidência do processo fundamental, básico.

Relação Inversa: tem_como_manifestação

dd)medida_de: a dimensão, quantidade ou capacidade determinada pela medição.

Relação Inversa: tem_como_medida

ee)mede: descobrir ou marcar a dimensão, quantidade, grau ou capacidade de.

Relação Inversa:é_medido_por

ff) método_de: a maneira e seqüência de eventos em desempenhar uma ação ou procedimento.

Relação Inversa: tem_como_método

gg)ocorre_em: tem lugar em ou acontece sob dadas condições, circunstâncias ou período de tempo ou em um dado local ou população. Esta relação inclui aparece em, está presente em e existe em.

Relação Inversa:tem_ocorrência_de

hh)parte_de: compõe, com uma ou mais unidades físicas, alguma porção maior. Ela inclui componente de, porção de, fragmento de, seção de e camada de.

Relação Inversa: composto_de

ii) fisicamente_relacionado_a: relacionado em virtude de características ou atributos físicos.

Relação Inversa: fisicamente_relacionado_a

jj) pratica: desempenha habitualmente ou costumeiramente.

Relação Inversa: praticado_por

kk)precede: ocorre mais cedo, anteriormente, no tempo. Isto inclui antecede, vem antes, na frente de, é anterior a.

Relação Inversa: precedido_por

ll) previne: pára, esconde ou elimina uma ação ou condição.

Relação Inversa: previnido_por

mm)processo_de: ação, função ou estado de.

Relação Inversa: tem_como_processo_de

nn)produz: gera, cria, rende, dá.

Relação Inversa: produzido_por

oo)propriedade_de: características de ou qualidade de.

Relação Inversa: tem_como_propriedade

pp)resultado_de: a condição, produto ou estado ocorrendo como conseqüência, efeito ou conclusão de uma atividade ou processo. Isto inclui produto de, efeito de, seqüela de, conseqüência de, culminação de e conclusão de.

Relação Inversa: tem_como_resultado

qq)envolve: estabelece fronteiras para ou defini os limites de uma outra estrutura física. Ela inclui limita, confina, fecha, restringe e demarca.

Relação Inversa: está_envolvido_por

rr) temporalmente_relacionado_a: relacionado no tempo pela precedência, coocorrência ou seqüência.

Relação Inversa: temporalmente_relacionado_a

ss)atravessa: cruza ou se estende através de outra estrutura física ou área. Ela inclui cruzar por ou cruzar através.

Relação Inversa: é_atravessado_por

tt) trata: aplica um remédio com o objetivo de efetivar a cura ou controlar uma situação.

Relação Inversa: é_tratado_por

uu)usa: emprega na execução de alguma atividade. Esta relação inclui aplica, utiliza e emprega.

Relação Inversa: é_usado_por

Anexo 2 Categorias Semânticas Básicas

A seguir são descritas as categorias semânticas disponíveis para a construção da rede semântica seguidas pela sua definição. Estas categorias foram extraídas da Rede Semântica do UMLS.

a) **Achado**: algo que é descoberto pela observação direta ou medição de um atributo ou condição de um organismo, incluindo a história clínica do paciente.

b) **Ácido Nucleico, Nucleosídeo ou Nucleotídeo**: um composto complexo de alto peso molecular que ocorre em células vivas.

c) **Alga**: uma planta aquática que contém clorofila, não forma embriões durante o desenvolvimento e perde tecido vascular.

d) **Amino Ácido, Peptídeo ou Proteína**: amino ácidos e cadeias de amino ácidos conectadas por ligações peptídicas. Como exemplos cita-se: Glicoproteínas, Mioglobina, Alanina e Sulfatase.

e) **Anfíbios**: vertebrados de sangue frio, saem do ovo como uma larva aquática, respirando por guelras. Quando adultos, os anfíbios respiram pelos pulmões. Pode-se citar como exemplo: Salamandra, Urodela e Sapo.

f) **Animal**: um organismo com células eucariontes.

g) **Anormalidade Adquirida**: uma estrutura anormal ou uma estrutura anormal em tamanho ou localização, encontrada em ou derivada de uma estrutura previamente normal. Como exemplos pode-se citar: "Hérnia", "Fistula" e "Hemorróidas".

h) **Anormalidade Congênita**: uma estrutura anormal, ou uma que é anormal em tamanho ou localização, presente no nascimento ou desenvolvida com o tempo como resultado de um defeito na embriogênese.

i) **Área Geográfica**: uma localização geográfica, geralmente tendo fronteiras definidas. Como exemplo pode-se citar "Canadá", "Cidades" e "Regiões Árticas".

j) **Atividade**: uma operação ou uma série de operações que um organismo ou máquina executa ou participa.

k) **Atividade de Cuidado da Saúde**: uma atividade de ou relacionada com a prática da medicina ou envolvendo o cuidado de pacientes.

l) **Atividade de Máquina**: uma atividade executada primariamente ou exclusivamente por máquinas.

m) **Atividade de Pesquisa**: uma atividade executada como parte de uma pesquisa ou experimentação.

n) **Atividade Diária ou Recreacional**: uma atividade executada por recreação ou exercício.

o) **Atividade Educacional**: uma atividade relacionada a uma organização e provisão de educação.

p) **Atividade Governamental ou Regulatória**: uma atividade executada pelo governo ou uma atividade relacionada com a criação ou imposição de regras ou regulamentações do governo em algum campo. Pode-se citar como exemplo: "Assistência Pública", "Credenciamento" e "Regulamentação.

q) **Atividade Ocupacional**: uma atividade executada como parte de uma ocupação ou tarefa.

r) **Atributo de Grupo**: uma entidade conceitual a qual se refere a freqüência ou distribuição de certas características ou fenômenos em certos grupos. Como exemplos pode-se citar: "Mortalidade Neo-Natal", "Expectativa de Vida", "Tamanho Familiar", "Características Populacionais" e "Estrutura de Grupo".

s) **Atributo de Organismo**: uma propriedade de um organismo ou de sua parte maior.

t) **Bactérias**: um pequeno, normalmente unicelular, microorganismo procarionte.

u) **Carbohidratos**: um composto consistindo de carbono, hidrogênio e oxigênio no qual a proporção de hidrogênio/oxigênio é a mesma que da água. Carbohidratos são geralmente caracterizados como açúcares e incluem mono, di, oligo e polissacarídeos, glicosídeos, amido e glicanos. Incluem-se aqui fosfatos de açúcares e excluem-se os glicolipídios.

v) **Célula**: a unidade estrutural e funcional fundamental dos organismos vivos.

w) **Comida**: qualquer substância que contém nutrientes, tais como carbohidratos, proteínas e gorduras, que podem ser ingeridos por um organismo vivo e metabolizado em energia ou tecido corpóreo. Algumas comidas são encontradas naturalmente, outras são parcial ou inteiramente feitas por seres humanos.

x) **Componente Celular**: uma parte da célula ou a matriz intercelular, geralmente visível pelo microscópio óptico.

y) **Comportamento**: qualquer atividade de seres humanos ou animais que pode ser observada diretamente por outros ou que pode ser feita sistematicamente observável pelo uso de estratégias especiais.

z) **Comportamento Individual**: comportamento exibido por um humano ou animal que não é um resultado direto da interação com outros membros da espécie mas que pode ter um efeito sobre outros.

aa) **Comportamento Social**: comportamento que é o resultado ou função direta da interação de seres humanos ou animais com seus semelhantes.

bb)**Composto Inorgânico**: um composto simples, geralmente com uma ligação iônica, não contendo carbono como componente principal. A ligação entre elementos em compostos inorgânicos é geralmente iônica. Inclui-se aqui ácidos e sais inorgânicos, ligas e minerais e exclui-se os hidrocarbonos.

cc)**Composto Organofósforo**: um composto orgânico contendo fósforo como um constituinte. Estão incluídos aqui fosfínico orgânico, derivados do ácido fosfórico e fosfônico e seus tiofósforos correspondentes. Estão excluídos os fosfolipídeos e açúcares fosfatos.

dd)**Conceito Espacial**: uma localização, região ou espaço, normalmente tendo fronteiras definidas.

ee)**Conceito Funcional**: um conceito o qual é de interesse devido a sua participação na realização de um processo ou atividade.

ff) **Conceito Qualitativo**: um conceito que é uma avaliação de alguma qualidade, ao invés de uma medida direta.

gg)**Conceito Quantitativo**: um conceito que envolve as dimensões, quantidade ou capacidade de alguma coisa utilizando alguma unidade de medida, ou que envolve a comparação quantitativa de entidades.

hh)**Conceito Temporal**: um conceito que pertence ao tempo ou duração.

ii) **Disfunção Ambiental ou Comportamental**: uma disfunção clinicamente significante cuja maior manifestação é comportamental ou ambiental. Estas disfunções podem ter manifestações ou etimologias biológicas presumidas ou identificadas.

jj) **Disfunção Celular ou Molecular**: uma função patológica inerente as células, partes das células ou moléculas.

kk)**Dispositivo de Pesquisa**: um objeto manufaturado usado primariamente na realização de pesquisa ou experimentação científica.

ll) **Dispositivo Médico**: um objeto manufaturado usado no diagnóstico, tratamento ou prevenção de perturbações fisiológicas ou anatômicas.

mm)**Doença ou Síndrome**: uma condição que altera ou interfere com um processo, estado ou atividade normal de um organismo. Normalmente caracterizado pelo funcionamento anormal de um ou mais dos sistemas, partes ou órgãos hospedeiros. Incluído aqui é um complexo de sintomas descritivo de um doença.

nn)**Efeito Ambiental de Seres Humanos**: uma mudança no ambiente natural que é o resultado da atividades dos seres humanos.

oo)**Eicosanóide**: um composto estruturalmente relacionado ao ácido araquidônico. Incluído aqui estão ácido araquidônico, ácido eicosanóico e seus derivados saturados e insaturados.

pp)**Elemento ou Íon**: Um dos 109 tipos de substâncias conhecidas atualmente que constituim toda a matéria no nível atômico e acima dele. Ela inclui metais elementais, gases raros e elementos radioativos. Esta

categoria não inclui as formas isotópicas menos abundantes, para as quais a categoria "Isótopo" está atribuída.

qq)**Entidade**: uma entidade conceitual ou física.

rr) **Entidade Conceitual**: categoria ampla para agrupar entidades ou conceitos abstratos.

ss)**Enzima**: uma proteína complexa que é produzida por células vivas e que catalisa reações biomédicas específicas. Há seis tipos principais de enzimas: oxidoreductases, transferases, hydrolases, lyases, isomerases e ligases.

tt) **Espaço do Corpo ou Junção**: uma área cercada ou limitada pelas partes do corpo ou órgãos ou por um local onde duas estruturas anatômicas se encontram ou se conectam.

uu)**Esteróide**: um de um grupo de compostos policíclicos que ocorrem na forma natural ou sintética. Estão incluídos aqui os encontrados naturalmente ou sintéticos, bufanólides, cardanólides, homosteróides, norsteróides e secosteróides.

vv)**Estrutura Anatômica**: parte normal ou patológica da anatomia ou organização estrutural de um organismo. Como exemplo pode-se ter termos como: Penas, Guelras e Chifre.

ww)**Estrutura Anatômica Completamente Formada**: uma estrutura anatômica em um organismo completamente formado. Em mamíferos, por exemplo, pode ser uma estrutura presente no corpo após o nascimento do organismo.

xx)**Estrutura Embrionária**: uma estrutura anatômica que existe somente antes do organismo estar completamente formado, por exemplo, a estrutura que existe apenas anteriormente ao nascimento do organismo. Esta estrutura pode ser normal ou anormal.

yy)**Estrutura Macromolecular**: uma molécula muito grande cuja estrutura contribui para a fisiologia de uma célula.

zz)**Evento**: um tipo amplo para agrupar atividades, processos ou estados.

aaa)**Fator Imunológico**: um fator biológico cujas atividades afetam ou desempenham um papel no funcionamento do sistema imunológico.

bbb)**Fenômemo ou Processo**: um processo ou estado o qual ocorre naturalmente ou como um resultado de uma atividade.

ccc)**Função Biológica**: um estado, processo ou atividade do corpo ou de um de seus sistemas ou partes.

ddd)**Função Celular**: uma função fisiológica inerente a células ou componentes celulares.

eee)**Função de Organismo**: uma função fisiológica do organismo como um todo, de sistemas de múltiplos órgãos ou tecidos.

fff)**Função de Tecido ou Órgão**: uma função fisiológica de um órgão, sistema orgânico ou tecido específico.

ggg)**Função Fisiológica**: um processo, atividade ou estado normal do corpo.

hhh)**Função Genética**: Funções de ou relacionadas com a manutenção, tradução ou expressão do material genético.

iii)**Função Molecular**: uma função fisiológica que ocorre a nível molecular.

jjj)**Função Patológica**: um processo, atividade ou estado desordenado de um organismo como um todo de um sistema corporal, sistema, órgãos múltiplos ou tecidos. Estão incluídas aqui respostas normais a um estímulo negativo bem como estados ou condições patológicas que são menos específicas que um doença. Funções patológicas freqüentemente tem efeitos sistêmicos.

kkk)**Fungos**: organismo eucarionte caracterizado pela ausência de clorofila e a presença de uma parede celular rígida.

lll)**Gene ou Genoma**: uma seqüência específica, ou no caso no genoma a seqüência completa, de nucleotídeos ao longo da molécula de DNA ou RNA a qual representa as unidades funcionais da hereditariedade.

mmm)**Grupo**: uma entidade conceitual que se refere a classificação de indivíduos de acordo com certas características compartilhadas.

nnn)**Grupo de Paciente ou Desabilitado**: um ou mais indivíduos classificados de acordo com uma doença, condição ou tratamento.

ooo)**Grupo Etário**: um ou mais indivíduos classificados de acordo com a sua idade. Como exemplo cita-se: Adulto, Prematuro, Adolescente e Octogenário.

ppp)**Grupo Familiar**: um indivíduo ou indivíduos classificados de acordo com seu relacionamento familiar ou posição relativa na sua unidade familiar.

qqq)**Grupo Ocupacional ou Profissional**: um ou mais indivíduos classificados de acordo com sua vocação.

rrr)**Grupo Populacional**: um ou mais indivíduos classificados de acordo com seu sexo, origem racial, religião, local de habitação, posição social ou financeira ou algum outro atributo cultural ou comportamental.

sss)**Hormônio**: em animais trata-se de uma secreção química de uma glândula endócrina cujos produtos são liberados na corrente circulatória. Hormônios de plantas ou hormônios sintéticos os quais são utilizados apenas para alterar ou controlar vários processos fisiológicos, como os agentes de controle reprodutivo, são atribuídos ao tipo "Substância Farmacológica". Os hormônios agem como mensageiros químicos e regulam vários processos fisiológicos tais como o crescimento, reprodução e metabolismo. Eles, normalmente, caem em duas classes amplas: hormônios de esteróides e de peptídeos.

ttt)**Humano**: homem moderno, a única espécie remanescente do Homo genus. Se um termo descreve um ser humano do ponto de vista ocupacional, familiar, social, etc., então uma categoria da hierarquia "Grupo" é atribuído.

uuu)**Idéia ou Conceito**: um conceito abstrato, tal como um conceito social, religioso ou filosófico.

vvv)**Indicador ou Reagente**: uma substância usada em reações no laboratório ou testes e procedimentos de diagnóstico ou de laboratório para detectar, medir, examinar ou analisar outros processos ou condições químicas.

www)**Invertebrados**: animal sem coluna vertebral.

xxx)**Isótopos**: uma forma de elementos que tem o mesmo número atômico mas diferente massa atômica devido a presença de um ou mais nêutrons adicionais. Inclui-se os isótopos estáveis e radioativos.

yyy)**Lesão ou Envenenamento**: ferimento traumático, lesão ou envenenamento causado por agente ou força externa.

zzz)**Língua**: o sistema de comunicação utilizado por uma nação particular ou pessoas.

aaaa)**Lipídios**: uma gordura ou substância derivada da gordura. Incluem-se os glicolipídios e fosfolipídios.

bbbb)**Localização ou Região do Corpo**: uma área subdivisão ou região do corpo demarcada para uma proposta de descrição topológica.

cccc)**Mamífero**: um vertebrado que tem temperatura do corpo constante e é caracterizado pela presença de pelos, glândulas mamárias e glândulas sudoríparas.

dddd)**Material Dental ou Biomédico**: uma substância usada na biomedicina ou odontologia predominantemente por suas propriedades físicas, em oposição as químicas. Inclui-se aqui materiais biocompatíveis, adesivos de tecidos, cimentos de ossos, resinas, etc.

eeee)**Modelo Experimental de Doença**: Uma representação em um organismo não humano de uma doença humana com a proposta de pesquisa nos seus mecanismos ou tratamento.

ffff)**Normas ou Leis**: um produto intelectual resultante de uma atividade legislativa ou normativa.

gggg)**Objeto Físico**: um objeto perceptível ao sentido da visão ou tato.

hhhh)**Objeto Manufaturado**: um objeto físico feito por seres humanos.

iiii)**Ocupação Biomédica ou Disciplina**: uma vocação, disciplina acadêmica, ou campo de estudo relacionado a biomedicina.

jjjj)**Ocupação ou Disciplina**: uma vocação, disciplina acadêmica ou campo de estudo ou ainda uma subparte de uma ocupação ou campo de estudo. Se o termo se refere a indivíduos que possuem uma vocação, o tipo "Grupo Ocupacional ou Profissional" é atribuído.

kkkk)**Organismo**: geralmente um organismo vivo, incluindo todas as plantas e animais.

llll)**Organização**: o resultado da união para uma proposta ou função comum. A existência contínua de uma organização não é dependente de qualquer de seus membros, sua localização ou facilidade particular. Componentes ou subpartes da organização estão também incluídos aqui. Como exemplos pode-se citar "Universidades", "Nações Unidas" e "Instituto de Proteção ao Meio-Ambiente".

mmmm)**Organização de Socorro ou Auto-Ajuda**: uma organização cuja proposta e função é fornecer assistência a necessitados ou oferecer apoio para aqueles que compartilham problemas similares. Como exemplos pode-se citar "Alcoólicos Anônimos" e "Cruz Vermelha".

nnnn)**Organização Relacionada ao Cuidado da Saúde**: uma organização estabelecida a qual realiza funções específicas relacionadas ao cuidado da saúde ou pesquisa na ciência da vida. Termos para cuidado da saúde relacionados a sociedades profissionais são atribuídos ao tipo Sociedade Profissional.

oooo)**Parte do Corpo, Órgão ou Componente do Órgão**: uma coleção de células e tecidos os quais estão localizados em uma área específica ou combinam e executam uma ou mais funções especializadas de um organismo. Ele varia de estruturas grossas a pequenos componentes de órgãos complexos.

pppp)**Pássaro**: um vertebrado com temperatura do corpo constante e caracterizado pela presença de penas.

qqqq)**Peixe**: um vertebrado aquático de sangue frio caracterizado por ter escamas e respirar por guelras. Estão incluídos aqui peixes que tem esqueleto de ossos ou cartilagens.

rrrr)**Planta**: um organismo que possui a parede das células de celulose, cresce pela síntese de substâncias inorgânicas, geralmente distinguido pela presença de clorofila e perda do poder de locomoção. Parte das plantas estão incluídas também.

ssss)**Procedimento de Diagnóstico**: um procedimento, método ou técnica usada para determinar a natureza ou identidade de uma doença ou perturbação. Este tipo exclui procedimentos os quais são realizados em espécimes em um laboratório.

tttt)**Procedimento de Laboratório**: um procedimento, método ou técnica utilizada para determinar a composição, quantidade ou concentração de uma espécime e que é executada em um laboratório clínico. Incluem-se aqui os procedimentos que medem o tempo e as porções das reações.

uuuu)**Procedimento Preventivo ou Terapêutico**: um procedimento, método ou técnica projetada para prevenir um doença ou desordem, para melhorar uma função física ou utilizada no processo de tratamento de uma doença ou lesão.

vvvv)**Processo Mental**: uma função fisiológica envolvendo processamento mental ou cognitivo.

wwww)**Processo ou Fenômeno causado por Seres Humanos**: um processo ou fenômeno que é um resultado de atividades de seres humanos. Se o termo se refere a atividade, ao invés do resultado desta atividade, um tipo da hierarquia "Atividade" é atribuído.

xxxx)**Processo ou Fenômeno Natural**: um Processo ou Fenômeno que ocorre independente das atividades dos seres humanos.

yyyy)**Produto Intelectual**: uma entidade conceitual resultante do empenho humano. Termos atribuídos a este tipo geralmente se referem a informação criada por seres humanos para algum proposta. Como exemplo pode-se citar "Teorema de Bayes" e "Sistemas de Informação".

zzzz)**Prostaglandina**: um membro do grupo de compostos fisiologicamente ativos derivados a partir do ácido araquidônico. Membros do grupo desempenham um papel maior no processo reprodutivo, simulação de músculos, nível de pressão sangüínea, inflamação,etc. Estão incluídos aqui prostaciclinas, tromboxanes e leucotrienes.

aaaaa)**Química**: é vista a partir de duas perspectivas distintas na rede, funcionalmente e estruturalmente.

bbbbb)**Química Inorgânica**: a classe geral de substâncias incluindo os elementos, seus iônicos e isótopos correspondentes e qualquer componente químico cujas moléculas estão unidas ionicamente ao invés de covalentemente. Inclui todos os compostos que não contém carbono como um componente principal.

ccccc)**Química Orgânica**: a classe geral de composto que contém carbono, normalmente baseados em correntes ou anéis de carbono, também contendo hidrogênio com ou sem nitrogênio, oxigênio ou outros elementos. A ligação entre estes elementos é geralmente covalente.

ddddd)**Réptil**: um vertebrado de sangue frio que possui uma cobertura externa de escamas ou carapaça rígida. Os répteis respiram através de pulmões e são geralmente ovíparos.

eeeee)**Resultado de Teste ou Laboratório**: resultado de um teste específico para medir um atributo ou determinar a presença, ausência ou grau de uma condição. São considerados inerentemente quantitativos e, desta forma, não são atribuídos ao tipo adicional "Conceito Quantitativo".

fffff)**Rickettsia ou Chlamydia**: um organismo intermediário em tamanho e complexidade entre um vírus e uma bactéria e que é parasitário dentro das células de insetos e carrapatos.

ggggg)**Seqüência de Amino Ácidos**: a seqüência de amino ácidos, ordenados em cadeias, dentro da proteína molecular. É de fundamental importância na determinação da estrutura da proteína.

hhhhh)**Seqüência de Carbohidratos**: a seqüência de carbohidratos dentro de polissacarídeos, glicoproteínas e glicolipídios.

iiiii)**Seqüência de Nucleotídeo**: a seqüência de purinas e pirimidinas em ácidos nucleicos e polinucleicos. Estão incluídos aqui as regiões ricas em nucleotídeos, seqüências conservadas e regiões de transformação de DNA.

jjjjj)**Seqüência Molecular**: um tipo amplo para agrupar as seqüências coletadas de amino ácidos, carbohidratos e seqüências de nucleotídeos.

kkkkk)**Sinal ou Sintoma**: uma manifestação observável de uma doença ou condição baseada em julgamento clínico ou uma manifestação de uma doença ou condição que é experimentada pelo paciente e registrada como uma observação subjetiva.

lllll)**Sistema Corporal**: um complexo de estruturas anatômicas que desempenham uma função comum.

mmmmm)**Sociedade Profissional**: um organização unindo aqueles que possuem uma vocação em comum ou que estão envolvidos com um campo de estudo comum.

nnnnn)**Substância**: um material com composição química definida ou fracamente definida.

ooooo)**Substância Biologicamente Ativa**: uma substância produzida ou necessária por um organismo, de interesse primário devido a seu papel no funcionamento biológico do organismo que o produz.

ppppp)**Substância Corporal**: material extracelular ou misturas de células e material extracelular produzido, excretado, pelo corpo. Inclui-se aqui substâncias como saliva, ácido gástrico, suor e esmalte dental.

qqqqq)**Substância Farmacológica**: uma substância utilizada no tratamento, diagnóstico, prevenção ou análise de uma função corporal normal ou anormal. Estão incluídas substâncias que aparecem naturalmente no corpo e são administradas terapeuticamente.

rrrrr)**Substância Neuroativa ou Amina Biogênica**: um fator biológico cujas atividades afetam ou desempenham um papel no funcionamento do sistema nervoso. Incluem-se aqui neuroreguladores, neurofisinas, catecolaminas, etc.

sssss)**Substância Perigosa ou Venenosa**: uma substância de risco devido aos seus efeitos potencialmente perigosos ou tóxicos. Este tipo inclui a maioria das drogas abusivas, bem como agentes que requerem manipulação especial devido a sua toxicidade. A maioria dos agentes farmacêuticos, embora potencialmente prejudiciais, são excluídos daqui e são atribuídos ao tipo "Substância Farmacológica".

ttttt)**Tecido**: uma agregação de células similarmente especializadas e substância intercelular associada. Os tecidos não são relativamente localizados em comparação a partes do corpo, órgãos ou componentes orgânicos.

uuuuu)**Técnica de Pesquisa de Biologia Molecular**: qualquer das técnicas usadas no estudo ou modificação dirigida do gene de um organismo vivo.

vvvvv)**Vertebrado**: um animal que possui coluna vertebral.

wwwww)**Vírus**: um organismo que consiste de um núcleo de um único ácido nucleico cercado por uma camada protetora de proteína. Um vírus pode se

replicar somente dentro de uma célula hospedeira viva. Um vírus exibe algumas, mas não todas, características normais de seres vivos.

xxxxx)**Visão Química Funcional**: visão química de uma perspectiva de suas características funcionais ou atividades farmacológicas.

yyyyy)**Visão Química Estrutural**: visão química de uma perspectiva de suas características estruturais. Incluem-se aqui termos que podem significar um sal, um íon ou composto.

zzzzz)**Vitamina** : uma substância, normalmente um complexo químico orgânico, presente em produtos naturais ou feito sinteticamente, o qual é essencial na dieta do homem ou outros animais. Estão incluídas as vitaminas precursoras e pró-vitaminas.

Anexo 3 Relacionamentos Básicos

A seguir são apresentados alguns relacionamentos da Rede Semântica de Conceitor Médicos. Alguns deles foram extraídos da Rede Semântica do UMLS e outros foram definidos com o auxílio de um especialista.

Origem da Relacionamento	Destino da Relacionamento	Relação
Sinal ou Sintoma	Achado	é-um
Vertebrado	Animal	é-um
Invertebrado	Animal	é-um
Comportamento	Atividade	é-um
Atividade Recreacional ou Diária	Atividade	é-um
Atividade Ocupacional	Atividade	é-um
Atividade de Máquina	Atividade	é-um
Procedimento de Diagnóstico	Atividade de Cuidado a Saúde	é-um
Procedimento de Laboratório	Atividade de Cuidado a Saúde	é-um
Procedimento Terapêutico ou Preventivo	Atividade de Cuidado a Saúde	é-um
Técnica de Pesquisa de Biologia Molecular	Atividade de Pesquisa	é-um
Atividade de Cuidado à Saúde	Atividade Ocupacional	é-um
Atividade de Pesquisa	Atividade Ocupacional	é-um
Atividade Governamental ou Regulatória	Atividade Ocupacional	é-um
Atividade Educacional	Atividade Ocupacional	é-um
Comportamento Social	Comportamento	é-um
Comportamento Individual	Comportamento	é-um
Junção ou Espaço Corporal	Conceito Espacial	é-um
Região ou Localização Corporal	Conceito Espacial	é-um

Seqüência Molecular	Conceito Espacial	é-um
Área Geográfica	Conceito Espacial	é-um
Sistema Corporal	Conceito Funcional	é-um
Disfunção Comportamental ou Mental	Doença ou Síndrome	é-um
Objeto Físico	Entidade	é-um
Entidade Conceitual	Entidade	é-um
Idéia ou Conceito	Entidade Conceitual	é-um
Achado	Entidade Conceitual	é-um
Atributo de Organismo	Entidade Conceitual	é-um
Produto Intelectual	Entidade Conceitual	é-um
Língua	Entidade Conceitual	é-um
Ocupação	Entidade Conceitual	é-um
Atributo de Grupo	Entidade Conceitual	é-um
Grupo	Entidade Conceitual	é-um
Estrutura Embrionária	Estrutura Anatômica	é-um
Anormalidade Congênita	Estrutura Anatômica	é-um
Anormalidade Adquirida	Estrutura Anatômica	é-um
Estrutura Anatômica Completamente Formada	Estrutura Anatômica	é-um
Parte, Órgão ou Componente de Órgão do Corpo	Estrutura Anatômica Completamente Formada	é-um
Tecido	Estrutura Anatômica Completamente Formada	é-um
Célula	Estrutura Anatômica Completamente Formada	é-um
Componente Celular	Estrutura Anatômica Completamente Formada	é-um
Estrutura Macromolecular	Estrutura Anatômica Completamente Formada	é-um
Gene ou Genoma	Estrutura Macromolecular	é-um
Atividade	Evento	é-um
Processo ou Fenômeno	Evento	é-um
Processos ou Fenômenos Causados po Humanos	Processo ou Fenômeno	é-um
Processo ou Fenômeno	Processo ou Fenômeno	é-um

Natural		
Lesão ou Envenenamento	Processo ou Fenômeno	é-um
Função Fisiológica	Função Biológica	é-um
Função Patológica	Função Biológica	é-um
Função do Organismo	Função Fisiológica	é-um
Função de Órgão ou Tecido	Função Fisiológica	é-um
Função Celular	Função Fisiológica	é-um
Função Molecular	Função Fisiológica	é-um
Função Genética	Função Molecular	é-um
Doença ou Síndrome	Função Patológica	é-um
Disfunção Celular ou Molecular	Função Patológica	é-um
Modelo Experimental de Doença	Função Patológica	é-um
Grupo Ocupacional ou Profissional	Grupo	é-um
Grupo Populacional	Grupo	é-um
Grupo Familiar	Grupo	é-um
Grupo Etário	Grupo	é-um
Grupo de Paciente ou Desabilitado	Grupo	é-um
Conceito Temporal	Idéia ou Conceito	é-um
Conceito Qualitativo	Idéia ou Conceito	é-um
Conceito Quantitativo	Idéia ou Conceito	é-um
Conceito Funcional	Idéia ou Conceito	é-um
Conceito Espacial	Idéia ou Conceito	é-um
Humano	Mamífero	é-um
Organismo	Objeto Físico	é-um
Estrutura Anatômica	Objeto Físico	é-um
Objeto Manufaturado	Objeto Físico	é-um
Substância	Objeto Físico	é-um
Dispositivo Médico	Objeto Manufaturado	é-um
Dispositivo de Pesquisa	Objeto Manufaturado	é-um
Ocupação ou Disciplina Biomédica	Ocupação ou Disciplina	é-um

Planta	Organismo	é-um
Fungos	Organismo	é-um
Vírus	Organismo	é-um
Rickettsia ou Chlamydia	Organismo	é-um
Bactérias	Organismo	é-um
Animal	Organismo	é-um
Organização Relacionada ao Cuidado da Saúde	Organização	é-um
Sociedade Profissional	Organização	é-um
Organização de Socorro ou Auto-Ajuda	Organização	é-um
Alga	Planta	é-um
Efeitos Ambientais dos Seres Humanos	Processo ou Fenômeno Causado por Seres Humanos	é-um
Função Biológica	Processo ou Fenômeno Natural	é-um
Norma ou Lei	Produto Intelectual	é-um
Visão Química Estrutural	Química	é-um
Visão Química Funcional	Química	é-um
Química Orgânica	Visão Química Estrutural	é-um
Química Inorgânica	Visão Química Estrutural	é-um
Elemento ou Íon	Química Inorgânica	é-um
Isótopo	Química Inorgânica	é-um
Composto Inorgânico	Química Inorgânica	é-um
Esteróide	Química Orgânica	é-um
Eicosanóide	Química Orgânica	é-um
Lactam	Química Orgânica	é-um
Alcanóide	Química Orgânica	é-um
Ácido Nucleico, Nucleosídeo ou Nucleotídeo	Química Orgânica	é-um
Compostos Organofósforos	Química Orgânica	é-um
AminoÁcido, Peptídeo ou Proteína	Química Orgânica	é-um
Carbohidrato	Química Orgânica	é-um
Lipídios	Química Orgânica	é-um

Seqüência de Nucleotídeos	Seqüência Molecular	é-um
Seqüência de AminoÁcidos	Seqüência Molecular	é-um
Seqüência de Carbohidratos	Seqüência Molecular	é-um
Química	Substância	é-um
Substância Corporal	Substância	é-um
Comida	Substância	é-um
Substância Neuroativa ou Amina Biogênica	Substância Biologicamente Ativa	é-um
Hormônio	Substância Biologicamente Ativa	é-um
Enzima	Substância Biologicamente Ativa	é-um
Vitamina	Substância Biologicamente Ativa	é-um
Prostaglandina	Substância Biologicamente Ativa	é-um
Fator Imunológico	Substância Biologicamente Ativa	é-um
Substância Farmacológica	Visão Química Funcional	é-um
Material Dental ou Biomédico	Visão Química Funcional	é-um
Substância Biologicamente Ativa	Visão Química Funcional	é-um
Reagente ou Indicador	Visão Química Funcional	é-um
Substância Perigosa ou Venenosa	Visão Química Funcional	é-um
Mamífero	Vertebrado	é-um
Anfíbio	Vertebrado	é-um
Peixe	Vertebrado	é-um
Réptil	Vertebrado	é-um
Pássaro	Vertebrado	é-um
Achado	Achado	associado_com
Achado	Área Geográfica	associado_com
Achado	Parte, órgão ou componente de órgão	ocorre_em
Achado	Substância Farmacológica	resultado_de

Anormalidade Adquirida	Atividade	associada_com
Comportamento	Atividade Educacional	resultado_de
Objeto Manufaturado	Conceito Quantitativo	mede
Parte, órgão ou componente de órgão	Parte, órgão ou componente de órgão	conectado_a
Procedimento de Laboratório	Substância Farmacológica	avalia_efeito_de
Sinal ou Sintoma	Doença ou Síndrome	diagnóstico
Substância Perigosa ou Venenosa	Animal	afeta
Vírus	Função Patológica	causa
Vírus	Parte, órgão ou componente de órgão	afeta

Bibliografia

[ALL 95] ALLAN, James. **Relevance Feedback with too much data**. Disponível por WWW em http://ciir.cs.umass.edu/info/psfiles/irpubs/ (15/08/1995)

[AME 94] AMERICAN MEDICAL INFORMATION ASSOCIATION. Standards for Medical Identifiers, Codes and Messages Needed to Create an Efficient Computer-stored Medical Record. **Journal of the American Medical Informatics Association**, [S.l.], v.1, n.1, p.1-7, Jan/Feb.1994.

[BAS 95] BASE de Dados Medline. Escola Paulista de Medicina. Disponível por WWW em http://www.epm.br/bireme/BRM115.htm (30/04/96)

[BLE 93] BLEICH, Howard L. Critique of an Evaluation of Software for Searching MEDLINE. In: ANNUAL SYMPOSIUM ON COMPUTER APPLICATIONS IN MEDICAL CARE, 17., 1993, Washington. **Proceedings...** Washington: McGraw-Hill, 1993. p.591-595.

[BOR 95] BORLAND INTERNATIONAL INC. Delphi User's Guide. [S.l:s.n.], 1995. 452p.

[BOR 95a] BORLAND INTERNATIONAL INC. Delphi Component Writer's Guide. [S.l:s.n.], 1995. 156p.

[BOR 95b] BORLAND INTERNATIONAL INC. Delphi Database Application Developer's Guide. [S.l:s.n.], 1995. 189p.

[BRO 94] BROWN, Eric et al. **Supporting Full-Text Information Retrieval with a Persistent Object Store**. Disponível por WWW em http://ciir.cs.umass.edu/info/psfiles/irpubs/browncallanedbt94.ps (09/04/96)

[BUS 95] BUSSMANN, José Eduardo Carvalho. **BART - Uma Biblioteca Orientada a Objetos de Apoio à Recuperação Textual**. Campina Grande: Universidade Federal da Paraíba, 1995. Dissertação de Mestrado.

[CEU 95] CEUSTERS, W.; DEVLIES, J. The Anthem Representation Formalism for the Alphabetic Index of ICD. In: WORLD CONGRESS ON MEDICAL INFORMATICS, 8., 1995, Vancouver, Canada. **Proceedings...**Vancouver: Healthcare Computing & Communications Canada Inc, 1995. p. 113-116.

[CHE 94] CHEN, Hsinchun. The Vocabulary Problem in Collaboration. **IEEE Computer**, New York, v.27, n.5, p.2-10, 1994. Disponível por WWW em http://ai.bpa.arizona.edu/papers/cscw94.html (05/12/96)

[CHE 96] CHEN, Hsinchun et al. **A Concept Space Approach to Addressing the Vocabulary Problem in Scientific Information Retrieval: An Experiment on the Worm Community System.** Disponível por WWW em http://ai.bpa.arizona.edu/papers/ (05/12/96)

[CIM 94] CIMINO, James J. Integrating Clinical Systems by Integration Controlled Vocabularies. In: THE BRAZILIAN CONFERENCE ON MEDICAL INFORMATICS, 1994, Porto Alegre, Brasil. **Proceedings...**Porto Alegre: [S.l.:s.n.], 1994.

[CRO 93] CROFT, W.Bruce. Knowledge-Based and Statistical Approaches to Text Retrieval. **IEEE Expert**, Los Alamitos, CA, v.8, n.2,p.8-12, Apr.1993.

[DAN 95] DANIELS, J.J.; RISSLAND, E.L. A Case-Based Approach ot Intelligent Information Retrieval. In: INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 1995, Seattle. **Proceedings...**Seattle: ACM Press, 1995. p.238-245.

[DES 87] DESCRITORES em Ciências da Saúde: Índice Hierárquico. São Paulo: Bireme, 1987.

[DIC 96] DICIONÁRIO de Especialidades Farmacêuticas: edição 96/97. Rio de Janeiro: Ed. de Publicações Científicas, 1996. 964p.

[FLI 95] FLIER, F.J.; HIRS, W.M. The Challenge of an International Classification of Procedures in Medicine. In: WORLD CONGRESS ON MEDICAL INFORMATICS, 8., 1995, Vancouver, Canada. **Proceedings...**Vancouver: Healthcare Computing & Communications Canada Inc, 1995. p. 121-125.

[FLO 95] FLORANCE, Valerie; MARCHIONINI, Gary. Information Processing in the Context of Medical Care. In: INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 1995, Seattle. **Proceedings...** Seattle: ACM Press, 1995. p.158-163.

[FOW 92] FOWLER, Jerry et al. The Medline Retriever. In: ANNUAL SYMPOSIUM ON COMPUTER APPLICATIONS IN MEDICAL CARE, 16., 1992, New York. **Proceedings...**New York: McGraw-Hill, 1992. p.473-477.

[FOW 95] FOWLER, Jerry et al. The Architecture of a Distributed Medical Dictionary. In: WORLD CONGRESS ON MEDICAL INFORMATICS, 8., 1995, Vancouver, Canada. **Proceedings...** Vancouver: Healthcare Computing & Communications Canada Inc, 1995. p. 126-130.

[FOX 93] FOX, E.A. **Source Book on Digital Libraries**. Disponível por FTP anônimo em fox.cs.vt.edu no arquivo /pub/DigitalLibrary (10/08/95)

[FUR 87] FURNAS, G.W. et al. The Vocabulary Problem in Human-System Communication. **Communications of the ACM**, New York, v.30, n.11, p.964-971, Nov. 1987.

[GAU 91] GAUCH, Susan; SMITH, John B. Search Improvement via Automatic Query Reformulation. **ACM Transactions on Information Systems**, New York, v.9, n.3, p.249-280, July 1991.

[GIL 87] GILMAN, Alfred Goodman et al. **As Bases Farmacológicas da Terapêutica**. Rio de Janeiro: Ed. Guanabara, 1987. 1195p.

[GNA 93] GNASSI, John Angelo; BARNETT, G.Octo. A Survey of Eletronic Drug Information Resources and Identification of Problems Associated with the Differing Vocabularies Used to Key Them. In: ANNUAL SYMPOSIUM ON COMPUTER APPLICATIONS IN MEDICAL CARE, 17., 1993, Washington. **Proceedings...** Washington: McGraw-Hill, 1993. p.631-635.

[GUI 93] GUIDI, John N. Matching References with MEDLINE via TCP/IP. In: ANNUAL SYMPOSIUM ON COMPUTER APPLICATIONS IN MEDICAL CARE, 17., 1993, Washington. **Proceedings...** McGraw-Hill, 1993. p. 606-610 .

[GUI 96] GUIDI, John N.; FOX, E.A. Information Retrieval and Genomics-An Introduction. **Computers in Biology and Medicine**, Great Britain, v.26, n.3, p.179-182, May 1996.

[GUT 96] GUTHRIE, Louise et al. The Role of Lexicons in Natural Language Processing. **Communications of the ACM**, New York, v.39, n.1, p.63-79, Jan. 1996.

[HAI 95] HAINES, D; CROFT, W.D. **Relevance Feedback and Inference Networks**. Disponível por WWW em http://ciir.cs.umass.edu/info/ psfiles/irpubs (28/08/95).

[HER 95] HERSH, William R. et al. Towards New Measures of Information Retrieval Evaluation. In: ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 1995, Washington. **Proceedings...** Washington: ACM Press, 1995. p.164-170.

[HIR 95] HIRSCH, Morris; ARONOW, David. **Suggesting Terms for Query Expansion in a Medical Information Retrieval System**. Disponível por WWW em http://ciir.cs.umass.edu/info/psfiles/ irpubs (11/06/96).

[IIV 95] IIVONEN, Mirja. Searches and Searches: Differences Between the Most and Least Consistent Searches. In: ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 1995, Washington. **Proceedings...** Washington: ACM Press, 1995. p.149-156.

[ING 95] INGENERF, J. Taxonomic Vocabularies in Medicine: The Intention of Usage Determines Different Established Structures. In: WORLD CONGRESS ON MEDICAL INFORMATICS, 1995, Vancouver, Canada. **Proceedings...** Vancouver: Healthcare Computing & Communications Canada Inc, 1995. P.136-139.

[JIN 93] JING, Y; CROFT, B. **An Association Thesaurus for Information Retrieval**. Disponível por WWW em http:// ciir.cs.umass.edu/info/psfiles/irpubs/jingcrofassocthes.ps.

[JON 68] JONES, K.Spark; NEEDHAM, R.M. Automatic Term Classification and Retrieval. **IP&M**, [S.l.], v.4, p.91-100, 1968.

[JON 70] JONES, K.Spark; JACKSON, D.M. The Use of Automatically-Obtained Keyword Classifications for Information Retrieval. **IP&M**, [S.l.], v.5, p.175-201, 1970.

[KRO 92] KROVETZ, R.; CROFT, W.B. Lexical Ambiguity and Information Retrieval. **ACM Transactions in Information Systems**, New York, v.10, n.2, p.115-141, Apr. 1992.

[LEE 93] LEE, D.L.; CROFT, W.B. Artificial Intelligence in Text-Based Information Systems. **IEEE Expert**, Los Alamitos, CA, v.8, n.2, p.6-7, Apr. 1993.

[LEW 96] LEWIS, David D.; JONES, Karen Sparck. Natural Language Processing for Information Retrieval. **Communications of the ACM**, New York, v.39, n.1, p.92-101, Jan. 1996.

[LOP 96] LOPES, Gabriel Pereira. Combining Natural Language Understanding and Information Retrieval for Flexible Hypertext Navigation. In: SIMPÓSIO BRASILEIRO DE INTELIGÊNCIA ARTIFICIAL, 1996, Curitiba. **Proceedings**...Berlim: Springer Verlag, 1996. p.233.

[LOP 96a] LOPES, Gabriel Pereira. **Combining Natural Language Understanding and Information Retrieval for Flexible Hypertext Navigation**. [S.l.:s.n.], 1996. Trabalho apresentado no SIMPÓSIO BRASILEIRO DE INTELIGÊNCIA ARTIFICIAL, 13., 1996, Curitiba, Pr.

[MAT 96] MATCHO, Jonathan et al. **Usando Delphi 2 - O Guia de Referência Mais Completo**. Rio de Janeiro: Campus, 1996. 883p.

[MIL 95] MILLER, Elizabeth et al. The Development of a Controlled Medical Terminology: Identification, Collaboration, and Customization. In: WORLD CONGRESS ON MEDICAL INFORMATICS, 8., 1995, Vancouver, Canada. **Proceedings...** Vancouver: Healthcare Computing & Communications Canada Inc, 1995. p.148-152.

[MIL 92] MILLER, Randolph A. et al. CHARTLINE: Providing Bibliographic References Relevant to Patient Charts Using the UMLS Metathesaurus Knowledge Sources. In: ANNUAL SYMPOSIUM ON COMPUTER APPLICATIONS IN MEDICAL CARE, 16., 1992, New York. **Proceedings...**New York: McGraw-Hill, 1992, p.86-89.

[OGG 95] OGG, Nancy J. et al. The Missouri Informatics *Thesaurus*. In: WORLD CONGRESS ON MEDICAL INFORMATICS, 8., 1995, Vancouver, Canada. **Proceedings...**Vancouver: Healthcare Computing & Communications Canada Inc, 1995. p.153-156.

[OLI 94] OLIVEIRA, Flávio Moreira. **Critérios de Equilibração para Sistemas Tutores Inteligentes**. Porto Alegre: CPGCC da UFRGS, 1994. 119p.

[PAR 89] PARSAYE, K. et ali. **Intelligent Databases**. John Wiley & Sons, 1989. 479p.

[PAT 93] PATTISON-GORDON, Edward. **Thenetsys A Semantic Network System**. Technical Report, Julho, 1993, Boston, Massachusetts. Disponível por FTP anônimo em dsg.harvard.edu no arquivo /pub/vox/Thenetsys/ Thenetsys-TR.ps. (06/96)

[PAT 93a] PATTISON-GORDON, Edward**. Knowledge Representation for Medical Concepts**. Disponível por WWW em http://dsg.harvard.edu/public/general/DSGKR.html. (06/96)

[PEN 93] PENG, Ping et al. Generating MEDLINE Search Strategies Using a Librarian Knowledge-Based System. In: ANNUAL SYMPOSIUM ON COMPUTER APPLICATIONS IN MEDICAL CARE, 17., 1993, Washington. **Proceedings...** Washington: McGraw-Hill, 1993. p.596-600.

[RAJ 95] RAJASHEKAR, T.B.; CROFT, W.B. **Combining Automatic and Manual Index Representations in Probabilistic Retrieval**. Disponível por WWW em http://ciir.cs.umass.edu/info/psfiles/irpubs/raja.ps. 1995.

[RIC 93] RICH, Elaine. **Inteligência Artificial**. São Paulo: Makron Books, 1993. p.722.

[ROC 93] ROCHA, Roberto A. et al. Automated Translation Between Medical Vocabularies Using a Frame-Based Interlingua. In: ANNUAL SYMPOSIUM ON COMPUTER APPLICATIONS IN MEDICAL CARE, 17., 1993, Washington. **Proceedings...**Washington: McGraw-Hill, 1993. p.690-694.

[ROT 93] ROTHWELL, DJ. et al. Developing A Standard Data Structure For Medical Language-The SNOMED Proposal. In: ANNUAL SYMPOSIUM ON COMPUTER APPLICATIONS IN MEDICAL CARE, 17., 1993, Washington. **Proceedings...**Washington: McGraw-Hill, 1993. p.695-699.

[SAL 68] SALTON, Gerard. **Automatic Information Organization and Retrieval**. Pittsburg: McGraw-Hill, 1968.

[SAL 75] SALTON, Gerard. **Dynamic Information and Library Processing**. Englewood Cliffs: Prentice-Hall, 1975. 523p.

[SAL 83] SALTON, Gerard. **An Introduction to the Modern Information Retrieval**. New York: McGraw-Hill, 1983.

[SAL 87] SALTON, Gerard; BUCKLEY, C. **Term Weighting Approaches in Automatic Text Retrieval**. Ithaca: Department of Computer Science Cornell University, 1987.

[SAL 94] SALTON, Gerard et al. Automatic Structuring and Retrieval of Large Text Files. **Communications of the ACM**, New York, v.37, n.2, p.97-108, Feb. 1994.

[SHO 85] SHOVAL, P. Principles, Procedures and Rules in an Expert System for Information Retrieval. **Information Processing Management**, [S.l.], v.21, n.6, p.475-487, 1995.

[SOE 74] SOERGEL, Dagobert. **Indexing Languages and Thesauri: Construction and Maintenance**. Los Angeles: Melville Publishing Company, 1974. p. 632.

[SPE 96] SPECIALIST Lexicon. 1996. Disponível por WWW em http://www.nlm.nih.gov/publications/factsheets/umls_specialist_lexicon.html. (05/96)

[TER 96] TER HOFSTEDE, A.M.H. et al. Query Formulation as an Information Retrieval Problem. **The Computer Journal**, New York, v.39, n.4, 1996. p.255- 274.

[TUT 95] TUTTLE, M.S. et al. Merging Terminologies. In: WORLD CONGRESS ON MEDICAL INFORMATICS, 1995, Vancouver, Canada. **Proceedings...**Vancouver: Healthcare Computing & Communications Canada Inc, 1995. p.162-166.

[UML 93] UMLS(R) Knowledge Sources. 4th Experimental Edition-April 1993. Documentation. Disponível por FTP anônimo em nlmpubs.nlm.nih.gov no arquivos umls/metadoc.txt e umls/net.txt. (05/96)

[UML 96] Unified Medical Language System.1996. Disponível por WWW em http://www.nlm.nih.gov/publications/factsheets/umls.html. (05/96)

[UML 96a] UMLS Semantic Network. 1996. Disponível por WWW em http://www.nlm.nih.gov/publications/factsheets/umls_semantic_network.html. (05/96)

[UML 96b] UMLS Metathesaurus. 1996. Disponível por WWW em http://www.nlm.nih.gov/publications/factsheets/umls_metathesaurus.html. (05/96)

[UML 96c] UMLS Information Sources Map. 1996. Disponível por WWW em http://www.nlm.nih.gov/publications/factsheets/umls_info_sources_map.html. (05/96)

[UML 96d] UMLS Knowledge Source Server. 1996. Disponível por WWW em http://www.nlm.nih.gov/publications/factsheets/umls_know_source_server.html. (05/96)

[VAN 95] VAN DEN HEUVEL, Freek et al. Smart Classification System (SmaCS): Design, Implementation, and Application for Congenital Heart Disease Terminology. In: WORLD CONGRESS ON MEDICAL INFORMATICS, 8., 1995, Vancouver, Canada. **Proceedings...**Healthcare Computing & Communications Canada Inc, 1995. p.167-171.

[WEB 96] WEBBER, Carine G. **Estudo sobre Recuperação de Informações e Bibliotecas Digitais**. Porto Alegre: CPGCC da UFRGS, 1996. 91p.

[WIE 96] WIEDERHOLD, Gio. Glossary: Intelligent Integration of Information. **Journal of Intelligent Information Systems**, Dordrecht, v.6, n.2/3, p.193-203, June 1996.

[WIL 96] WILBUR, W.J.; YANG, Y. An Analysis of Statiscal Term Strength and Its Use in the Indexing and Retrieval of Molecular Biology Texts. **Computers in Biology and Medicine**, Great Britain, v.26, n.3, p.209-222, May, 1996.

[WOL 77] WOLFF-TERROINE, M. Terminology and Nomenclatures. In: IFIP WORKING CONFERENCE ON COMPUTACIONAL LINGUISTICS IN MEDICINE, 1977, Uppsala, Suécia. **Proceedings...**Uppsala: North-Holland Publishing Company, 1977. P.55-61.

[YAN 93] YANG, Yiming; CHUTE, C.G. Words or Concepts: the Features of Indexing Units and their Optimal Use in Information Retrieval. In: ANNUAL SYMPOSIUM ON COMPUTER APPLICATIONS IN MEDICAL CARE, 17., 1993, Washington. **Proceedings...**Washington: McGraw-Hill, 1993. p.685-689.

Informática UFRGS

CURSO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

O Estudo e Desenvolvimento do Protótipo de uma Ferramenta de Apoio a Formulação de Consultas a Bases de Dados na Área da Saúde

por

Carine Geltrudes Webber

Dissertação apresentada aos Senhores:

Profa. Dra. Beatriz de Faria Leão (UNIFESP)

Prof. Dr. José Palazzo Moreira de Oliveira

Dra. Zita Prates de Oliveira

Vista e permitida a impressão.
Porto Alegre, 14 / 05 / 97.

Prof. Dr. José Mauro Volkmer de Castilho,
Orientador.

Prof. Flávio Rech Wagner
Coordenador do Curso de Pós-Graduação
em Ciência da Computação - CPGCC
Instituto de Informática - UFRGS