

APLICAÇÃO DE MÉTODOS BIBLIOMÉTRICOS E DA
"CO-WORD ANALYSIS" NA AVALIAÇÃO DA LITERATURA
CIENTÍFICA BRASILEIRA EM CIÊNCIAS DA SAÚDE DE
1990 A 2002.

SUELI GONSALEZ SAES

Tese apresentada ao Programa de Pós
Graduação em Saúde Pública da
Faculdade de Saúde Pública da
Universidade de São Paulo para
obtenção do título de Doutor em
Saúde Pública.

Área de Concentração: Epidemiologia
Orientador: *Prof. AUGUSTO HASIAK
SANTO*

SÃO PAULO

2005

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Autorizo exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta tese, por processos fotocopiadores. Ao usá-lo, cite a fonte.

Assinatura:

Data:

*Aos meus queridos pais
Antonio e Luiza,*

*Pela dedicação e amor incondicional.
Tudo, sempre!*

"Existe somente uma idade para a gente ser feliz, somente uma época na vida de cada pessoa em que é possível sonhar e fazer planos e ter energia bastante para realizá-los a despeito de todas as dificuldades e obstáculos. Uma só idade para a gente se encantar com a vida e viver...Essa idade tão fugaz na vida da gente chama-se Presente e tem a duração do instante que passa".

Mario Quintana

*Às minhas filhas
Anna Carolina e Anna Luiza
Com todo meu amor.*

Ao meu orientador **Professor. Augusto Hasiak Santo**, agradeço profundamente pela oportunidade, orientação e apoio na realização deste trabalho.

**"El mejor homenaje que puede tributarse a las
personas buenas es imitarlas."**

Concepción Arenal

Ao **Prof. Pere Escorsa Castells**, da Universidade Politécnic da Catalunya, agradeço profundamente pela oportunidade, orientação e acolhida durante a realização do trabalho nessa Universidade.

Ao **Prof. Ramón Maspons Bosch**, da Universidade Politécnic da Catalunya, amigo de todos os momentos, agradeço pela dedicação e o apoio imprescindível, no acompanhamento das análises de "co-word", realizadas neste trabalho.

À **Dra. Elicet Cruz**, do IALE Tecnologia, agradeço pelo carinho e apoio profissional, durante minha permanência na Universidade Politécnic da Catalunya.

À **Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES**, agradeço pela bolsa de estudos concedida.

Ao **Dr. Júlio César Rodrigues Pereira**, do Departamento de Epidemiologia, da Faculdade de Saúde Pública, pela amizade e pelo auxílio na realização das análises bibliométricas deste trabalho.

Ao **Dr. Ricardo Oliva**, da Agência Nacional de Vigilância Sanitária, por possibilitar o acesso às bases de dados utilizadas para realização deste trabalho.

Ao **Marcello Sampaio Di Pietro**, da Coordenadoria de Avaliação e Integração de Dados Institucionais da Universidade Federal de São Paulo, pelo apoio profissional e sua generosa disponibilidade, que garantiram a utilização dos potenciais da informática na realização deste trabalho.

Ao **Dr. Luiz Jacintho da Silva**, meu chefe, que soube entender minhas ausências.

À **Maria Luisa dos Santos Marchi**, pela amizade e incentivo nos principais momentos.

À **Yeda Lopes Nogueira**, pela generosidade e amizade de sempre.

E todos aqueles que de alguma forma estiveram ao meu lado neste trabalho.

RESUMO

Saes, S.G. - Aplicação de Métodos Bibliométricos e da "Co-Word Analysis" na Avaliação da Literatura Científica Brasileira em Ciências da Saúde de 1990 a 2002. São Paulo; 2005. [Tese de Doutorado - Faculdade de Saúde Pública da Universidade de São Paulo]

Analisou-se neste trabalho a produção científica brasileira, no campo da Saúde, por meio da pesquisa de publicações, constantes da base de dados "web of Knowledge", no período de 1990 a 2002. A pesquisa incluiu 3.066 revistas desse campo do conhecimento. Nesse período foram contabilizadas 38.349 publicações. Os métodos bibliométricos tradicionais e a "co-word analysis" foram utilizados como instrumento de análise para medir a atividade científica no campo da saúde. O pacote estatístico, Statistical Package for Social Science - SPSS, foi utilizado para a obtenção dos indicadores bibliométricos tradicionais. Demonstraram-se as etapas de trabalho e a potencialidade da "co-word analysis". Foram estudadas as palavras que representam os descritores e identificadores na base de dados e estas foram relacionadas com outras variáveis como autores, ano de publicação e filiação da publicação. Para tal, utilizou-se o "Tétralogie", um software que aplica tratamentos estatísticos para conseguir que o grau de co-ocorrência entre as palavras seja representado como distâncias no espaço. Desta forma, as palavras podem ser visualizadas como pontos no espaço, por meio dos mapas tecnológicos e a distância entre elas será representativa de sua co-ocorrência ou afinidade. Os mapas tecnológicos permitiram a identificação de redes de cooperação entre pesquisadores e instituições, bem como identificação de "clusters" entre autores e as palavras descritoras do

campo de conhecimento estudado. Esta abordagem para a área de avaliação em Ciência e Tecnologia demonstrou ampliar as análises bibliométricas tradicionais, e pode incrementar as atividades de planejamento e definição de ações estratégicas indicando, para onde e como se poderiam dirigir esforços para desenvolvimento do campo estudado.

PALAVRAS CHAVE:

Análise Bibliométrica, Bibliometria, Co-ocorrência de palavras, "Tétralogie", Ciência e Tecnologia (C&T), Avaliação da Pesquisa Científica, Cienciometria, Ciências da Saúde, Brasil.

ABSTRACT

Saes, S.G.-Application of Bibliometric Methods and "Co-Word Analysis" of Brazilian Scientific Literature in Health Sciences, 1990-2002. São Paulo; 2005. [Tese de Doutorado - Faculdade de Saúde Pública da Universidade de São Paulo]

This work analyzed Brazilian scientific production in the health sciences field by analysis of publications present in the database "web of knowledge" in the period from 1990 to 2002. The study covered 3,066 journals in this field of knowledge. During this period a total of 38,349 Brazilian publications were found. Bibliometric analysis and "co-word analysis" were used to measure the scientific activity in health sciences. A statistical software package (Statistical Package for Social Science - SPSS) was used to determine traditional bibliometric analysis. The words representing descriptors and identifiers in the database had been studied and the relationship to other variables such as authors, publication year, and affiliation within the publication was determined. For this analysis, the "*Tétralogie*" software that applies statistical methods to measure the co-occurrence among words and represents this measure as distances in space was used. In this way, words can be visualized as points in space on a map. The maps and distances between words will represent the co-occurrence or affinity between keywords. The technological maps have shown networks of cooperation among researchers and institutions, as well as identification of "clusters" of authors and a hierarchy of the descriptive keywords. This approach for evaluation of science and technology demonstrated an extension of the traditional bibliometric analysis and showed that it could improve planning activities by identification of strategic actions

pointing to where and how efforts for improvement in the field of health sciences should be directed.

KEY WORDS:

Bibliometric Analysis, Bibliometrics, "Co-Word Analysis", Science and Technology (S&T), Scientific Research Evaluation, Scientometrics, Health Sciences, Brazil.

ÍNDICE DE ILUSTRAÇÕES

Lista de Quadros

Figura 1 - Número de publicações científicas brasileiras	49
Figura 2 - Crescimento da produção científica brasileira	50
Figura 3 - Número de citações em artigos científicos brasileiros ..	51
Figura 4 - Curva das publicações e citações	52
Figura 5 - Tendência de citações por ano, para publicações brasileiras	53
Figura 6 - Curva da média de autores, publicações brasileiras	54
Figura 7 - Curva do Percentual de publicações com colaboração	55
Figura 8 - Número de autores por publicação	56
Figura 9 - Distribuição anual das publicações brasileiras	56
Figura 10 - Distribuição do número de autores brasileiros, segundo número de publicações	57
Figura 11 - Distribuição das publicações no campo da saúde, segundo localidade onde são editadas	58
Figura 12 - Distribuição das publicações, segundo localidade	59
Figura 13 - Produção científica brasileira, de acordo com a "Lei de Bradford"	60
Figura 14 - Núcleo das revistas mais produtivas, de acordo com o Grupo 1 de Bradford	61
Figura 15 - Áreas do Conhecimento de maior produtividade de publicações brasileiras	62
Figura 16 - Produtividade por Áreas do Conhecimento	64
Figura 17 - Produção científica por Tipo de Documento	66
Figura 18 - Distribuição das publicações brasileiras, segundo tipo de documento	67
Figura 19 - Publicações científicas brasileiras, classificadas por Tipo de Documento	68

Figura 20 - Número de artigos segundo o idioma de publicação	68
Figura 21 - Percentual publicações brasileiras, segundo o idioma de publicação	69
Figura 22 - Composição dos "clusters" de acordo com os padrões da produção científica brasileira	71
Figura 23 - Relação do "cluster" 1 das publicações brasileiras, com as áreas do conhecimento	74
Figura 24 - Relação do "cluster" 2 das publicações brasileiras, com as áreas do conhecimento	75
Figura 25 - Relação dos "cluster" 3 e 4, das publicações brasileiras, com as áreas do conhecimento	76
Figura 26 - Percentual dos Descritores segundo a frequência de ocorrência	81
Figura 27 - Descritores e Identificadores, segundo as vinte primeiras maiores frequência de ocorrência	81
Figura 28 - Ocorrência da palavra "Dengue", nas publicações brasileiras	83
Figura 29 - Fragmento da matriz de co-ocorrência de descritores, segundo 1% das palavras de maior frequência	86
Figura 30 - Zoom da matriz de co-ocorrência de Descritores	87
Figura 31 - Fragmento da matriz de co-ocorrência de "DE", segundo 1% das palavras de maior frequência	87
Figura 32 - Descritores liderado pela palavra "epidemiology"	88
Figura 33 - Mapa segundo matriz de co-ocorrência de descritores ...	89
Figura 34 - Fragmento da matriz de descritores	90
Figura 35 - Descritores liderados pela palavra "anxiety"	90
Figura 36 - Artigos e citações segundo a palavra "anxiety"	91
Figura 37 - Artigos e citações que contém a palavra "anxiety"	92
Figura 38 - Fragmento da matriz de co-ocorrência de descritores ...	93
Figura 39 - Lista de descritores	93

Figura 40 - Mapa dos "clusters", segundo matriz de co-ocorrência de descritores	95
Figura 41- Mapa segundo matriz de co-ocorrência de descritores	95
Figura 42 - Identificadores, liderado pela palavra "expression" e "protein"	97
Figura 43 - Percentual dos Identificadores, segundo sua frequência de ocorrência	98
Figura 44 - Grupo 4 de descritores	99
Figura 45 - Cruzamento de descritores e identificadores	101
Figura 46 - Cruzamento de descritores e identificadores	102
Figura 47 - Cruzamento de descritores e identificadores	103
Figura 48 - Cruzamento de descritores e identificadores	104
Figura 49 - Cruzamento de descritores e identificadores	105
Figura 50 - Autores do "cluster ecology" segundo publicações brasileiras	107
Figura 51 - Mapa de autores do "cluster" ecology, segundo descritores	107
Figura 52 - Mapa de autores do "cluster" ecology, segundo descritores	108
Figura 53 - Mapa tecnológico de autores e descritores do "cluster ecology"	109
Figura 54 - Mapa tecnológico de autores e descritores do "cluster ecology", segundo publicações brasileiras	110
Figura 55 - Evolução do Grupo "Forattini" no Tempo	111
Figura 56 - Matriz de descritores e autores do "cluster ecology", segundo as publicações brasileiras	112
Figura 57 - Curva de carga para a contagem do campo de endereços, segundo as publicações brasileiras	113
Figura 58 - Rede de colaboradores do "cluster ecology", segundo as publicações brasileiras	114

SUMÁRIO

1 - INTRODUÇÃO	1
2 - SITUAÇÃO ATUAL DA CIÊNCIA E TECNOLOGIA (C&T) NO BRASIL	12
2.1 Histórico	12
2.2 Recursos Humanos em C&T no Brasil	15
2.3 Fomento	16
2.4 Setor produtivo	17
3 - AVALIAÇÃO EM CIÊNCIA E TECNOLOGIA	19
3.1 Bibliometria	19
3.2 Indicadores Bibliométricos	22
3.3 Co-ocorrência de Palavras - Co-word Analysis	25
4 - OBJETIVOS	31
4.1 Objetivo Geral	31
4.2 Objetivos Específicos	31
5 - MATERIAL E MÉTODOS	32
5.1 Universo do Estudo	32
5.2 Busca dos dados	34
5.3 Análises Bibliométricas	34
5.3.1 Tratamento do Dados	35
5.4 "Co-Word Analysis"	36
5.4.1 Tratamento dos Dados para a "Co-Word Analysis"	36
5.4.2 Características do Software "Tétralogie"	36
5.5 Processo de Tratamento da Informação	40
5.5.1 Pré-Tratamento dos Dados	40
5.5.2 Estrutura da "BASEDEDONNEES"	40

5.5.3 Tratamento do Dados.....	44
5.6 Matrizes e Mapas Tecnológicos.....	44
5.6.1 Métodos Estatísticos Integrados.....	46
5.7 Limitações do Estudo.....	47
6 - RESULTADOS E DISCUSSÃO.....	49
6.1 Análises Bibliométricas.....	49
6.2 "Co-Word Analysis".....	77
7 - CONSIDERAÇÕES FINAIS.....	116
8 - REFERÊNCIAS BIBLIOGRÁFICAS.....	127

1 - INTRODUÇÃO

"A ciência e a tecnologia constituem, hoje, fatores preponderantes no processo de desenvolvimento econômico e social de qualquer país". (Marcovitch-1986)

Esta frase, apesar de ter sido escrita há mais de uma década, guarda os mesmos fundamentos, é repetida em conferências e aberturas de fóruns de discussão de Ciência e Tecnologia (C&T). É consenso não só entre os atores envolvidos com as questões relacionadas com o desenvolvimento científico, tecnológico e de inovação. Para Arlington, 1998, em função das atuais características do processo econômico "... os investimentos em pesquisa básica, tecnologia de ponta e ensino da ciência e engenharia são fundamentais para o alcance de metas nacionais relativas ao desenvolvimento econômico e social de melhoria da saúde, do bem estar, da competitividade econômica e da segurança nacional".

Ciência e Tecnologia constituem a mola mestra que impulsiona o processo de desenvolvimento econômico e social das nações. Não há dúvida sobre a relevância da C&T para a sociedade contemporânea. O conhecimento é hoje o elemento fundamental para a geração de desenvolvimento. Neste contexto, a divulgação científica, além de contribuir para a democratização do conhecimento, aproxima o cidadão comum dos benefícios que ele tem direito de reivindicar para a melhoria do bem-estar social, dando-lhe uma visão mais clara sobre as verdadeiras causas e efeitos dos problemas que enfrenta no dia-a-dia. (Carneiro, 2002).

Nos países em desenvolvimento, como o Brasil, a necessidade de equacionar problemas sociais e impulsionar o crescimento econômico coloca o campo da Ciência e Tecnologia no centro de qualquer estratégia e política de desenvolvimento sustentado. Cresce, portanto, a necessidade de não somente fomentá-lo, nas esferas pública e privada, mas também de fazê-lo de forma ordenada e consistente, orientada por uma visão de futuro, perfeitamente articulada com as necessidades e demandas dos cidadãos.

Ao longo das últimas décadas, apesar de limitações e dificuldades enfrentadas pela comunidade científica brasileira, parece inegável que o país deu um salto considerável em matéria de ciência e tecnologia. O Brasil conta hoje com uma base significativamente sólida para empreender um novo salto e buscar objetivos mais ambiciosos, fazendo da C&T um poderoso instrumento de transformação social e desenvolvimento econômico.(UNESCO, 2003).

Algumas iniciativas, como a do próprio Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) quando desenvolveu em novembro de 1998 o seu Plano Institucional, evidenciam esta necessidade para o Brasil na introdução de seu planejamento estratégico quando afirma que *"... o conhecimento científico e tecnológico, neste final de século, constitui o cerne das próprias transformações que atingem todos os aspectos da vida humana. De um lado, a progressiva conversão da ciência em força produtiva e de competitividade internacional consolida suas vinculações com as formas de poder e torna inquestionável sua relevância econômica, social e política. Por outro lado, os custos crescentes da*

produção científica e tecnológica e a crise de financiamento público ampliam significativamente os desafios institucionais". (CNPq, 1998).

É ainda extremamente relevante o fato de que o desenvolvimento do conhecimento se distribui desigualmente. Leta, em seu trabalho de 1996, mostra como sendo indicativo desta distribuição o fato de que 7 países (USA, Grã-Bretanha, Alemanha, Japão, França, Canadá e Rússia) eram responsáveis por 75% dos artigos científicos publicados nos mais importantes periódicos científicos a cada ano, sendo que o restante do planeta, com 83% da população mundial, é composto de países "consumidores" desse conhecimento. (Leta e Meis-1998)

Em um artigo publicado em 15 de julho de 2004 no jornal "A Folha de São Paulo", Bonalume Neto, aponta para esta desigualdade "... *Que o mundo é desigual, muitos sabem. Noruegueses vivem bem mais e melhor... como costuma revelar o Índice de Desenvolvimento Humano, divulgado ontem pelo PNUD (Programa das Nações Unidas para o Desenvolvimento)*" e se reporta a um estudo, de autoria de David King (2004), sobre o estado da arte da ciência no planeta, onde evidencia que a situação tende a piorar cada vez mais, pois os países ricos, como a Noruega, investem muito mais e melhor em ciência e inovação tecnológica do que os mais pobres.

Dois estudos publicados recentemente apontam que o Brasil vem alcançando bom desempenho na divulgação de seus resultados de pesquisa e, conseqüentemente, aumentando a visibilidade entre os países empenhados em produzir ciência e tecnologia.

O estudo produzido pela *National Science Foundation*, agência do governo dos Estados Unidos,, em agosto de 2004, faz um balanço do número de artigos de ciência e engenharia publicados nas mais importantes revistas entre os anos de 1988 e 2001. O resultado é um notável crescimento da produção de países da Ásia e da América Latina. Nesta última região, o número de artigos publicados quase triplicou nesse período, com destaque para o Brasil que quadruplicou suas publicações e respondeu, em 2001, por 44% do desempenho regional. Em 2001, o Brasil produziu 7.205 artigos, contra 1.766 publicados em 1988. Uma análise per capita, no entanto, revela maior desempenho em C&T na Argentina e no Chile. Entre os anos de 1999 e 2001 esses países ultrapassaram 75 artigos publicados/milhão de habitantes, enquanto que no Brasil esse índice não alcançou 39 artigos/milhão de habitantes. (Rodrigues, 2004).

Por sua vez, o trabalho de autoria de King⁸ em julho de 2004, publicado na revista científica, "Nature", mostra que apenas 31 países, incluindo o Brasil, estão representados no 1% dos mais importantes artigos científicos, ou seja, aqueles que foram os mais citados por outros cientistas. O resultado foi um ranking dos 31 países que produzem as pesquisas mais citadas no mundo, no qual o Brasil ocupa a 23ª posição e a condição de único representante da América Latina. King⁸ fez comparações entre a produção de artigos desses 31 países mais significativos na ciência mundial, mostrando que esse grupo responde por 98% dos artigos mais citados dentre todos. Os outros 162 países do mundo contribuíram com os 2% restantes. No entanto, existem olhares críticos para esses resultados, como o de Rodrigues (2004) quando

diz "...A repercussão desses estudos, ao tempo que premia o esforço do pesquisador brasileiro, também reforça o papel dos índices bibliométricos como quase que exclusivos indicadores de produção de C&T. Assim, ao lado da comemoração pelas boas notícias, cabem algumas reflexões sobre o significado desses dados..." (Rodrigues, 2004)

Assim, numa sociedade competitiva em escala mundial, implementar o conhecimento técnico e científico é tarefa indispensável para o desenvolvimento econômico e social, principalmente para os países em desenvolvimento que são "consumidores" do conhecimento técnico e científico, de forma a identificar suas necessidades e peculiaridades.

Porém, o pano de fundo da implementação do conhecimento científico e tecnológico é a existência de recursos limitados e uma exigência cada vez maior de racionalidade e objetividade na aplicação dos poucos recursos disponíveis. (Black-1997)

Considerando que um fundamento básico para a implementação de qualquer sistema é "...que os resultados refletem os sistemas que o produzem" (Turnock-1997) e que, ainda nas palavras de Turnock, "...para melhorar algo devemos ser capazes de controlá-lo; para controlá-lo devemos ser capazes de entendê-lo e para entendê-lo devemos ser capazes de medi-lo", torna-se indispensável estudar, para conhecer e medir, a produção científica para a implementação desse conhecimento, superação das dificuldades e alcance dos pressupostos de racionalidade e objetividade.

A habilidade para avaliar o patamar científico das nações é vital para os governos, negócios e fundações que devem decidir as prioridades científicas e de investimento.

Nos Estados Unidos, a liderança do "*National Science Foundation*" uma agência independente do governo americano, em avaliar as fronteiras da ciência, pesquisa e educação é demonstrada, em parte, pelas importantes avaliações internas e externas de desempenho, produzidas pelo órgão, que são utilizadas de forma crescente pelo Congresso americano e pelos escritórios de gestão e orçamento de Ciência e Tecnologia para tomada de decisão.

Na Europa, o "*L'Observatoire des Sciences et des Techniques - L'OST*" tem o papel principal de produzir os indicadores quantitativos que se relacionam às atividades científicas, tecnológicas e de inovação. Seu relatório de 2002, "*Science & Indicateurs*" permitem interpretar e analisar a posição científica e tecnológica da França e de seus sócios europeus, em seus contextos, nacional e no mundo. O "OST" produz os indicadores que permitem os atores da ciência e da tecnologia discutir em sentido amplo a política da pesquisa e de elaborar sua estratégia. (www.obs.ost.fr).

O RICYT - Rede Ibero-americana de Indicadores de Ciência e Tecnologia, foi constituído pelo Programa Ibero-americano de Ciência y Tecnologia para o Desenvolvimento (CYTED) a partir de uma proposta surgida do Primeiro encontro Ibero-americano sobre Indicadores de Ciência e Tecnologia realizado na Argentina no final de 1994. Porém concretizou-se em abril de 1995. Atualmente executa o projeto de "Construção de um sistema

interamericano de ciência, tecnologia e inovação - Plataforma Básica", financiado pela Organização de Estados Americanos (OEA). Seu objetivo é construir um sistema de informação que reflita as necessidades dos países americanos, concebido como uma rede complexa que potencialize a escala regional em suas capacidades nacionais e sub-regionais, nesta matéria. (www.ricyt.edu.ar).

No Brasil, O Ministério da Ciência e Tecnologia - MCT, divulga os indicadores nacionais de ciência e tecnologia (C&T) por meio de sua página na Internet e em publicação impressa, cuja última versão foi editada em 2002. No segundo semestre de 2003 teve início o processo de atualização das informações com a reformulação de tabelas e gráficos, que se estendeu até 2004, com a inclusão de dados disponíveis até abril daquele ano. A discussão sobre os indicadores e a sua representatividade é uma preocupação do ministério que busca gerar informações mais completas e acuradas, para subsidiar os posicionamentos da sociedade e do governo.

Outras duas iniciativas que ocorreram em 2004, vêm ao encontro das necessidades na área de avaliação em C&T no país e devem ser creditadas à Fundação de Amparo à Pesquisa de São Paulo - FAPESP, que desenvolveu um serviço inédito, que facilita a busca pela internet de informações voltadas à produção e análise de indicadores de ciência, tecnologia e inovação. Trata-se do FAPESP.Indica. Este serviço tem como finalidade ajudar pesquisadores, gestores e outros interessados na consulta de informações para a produção e análise de indicadores de ciência, tecnologia e inovação CT&I. (Vogt e Gusmão, 2004) A segunda deveu-se ao lançamento da edição 2004 dos

"Indicadores FAPESP de CT&I", centrada na análise da evolução mais recente dos esforços de C&T paulistas, no contexto mais amplo da dinâmica nacional e internacional do setor, na entrada do novo século. (Gusmão, 2004).

Há várias maneiras de medir a produção científica dos países e seu impacto geral na ciência, nenhuma das quais é perfeita. Um artigo pode ser muito citado por outros cientistas de forma negativa, por exemplo. Mas, em geral, e em grandes conjuntos de dados, quanto mais citado, mais impacto tem uma pesquisa.

Os diálogos entre cientistas e pesquisadores vêm sendo estudados desde os anos 50 por meio das chamadas redes de comunicação científica. No entanto, a proliferação desses estudos deu-se a partir da década de 60 com Solla Price, pesquisando a mensuração da literatura científica por meio da análise de citações - bibliometria. Nessas análises percebe-se que a ciência é construída por meio de diálogos estabelecidos entre o autor citado e aquele que o cita; outro aspecto relevante abordado é o direcionamento evolutivo da ciência e tecnologia. (Pinto e Lima-2002)

A "Gestão de Ciência e Tecnologia" - enquanto um campo de conhecimento e de prática na busca permanente por maior racionalidade na aplicação de recursos e maior eficiência e eficácia nos resultados atingidos tanto na geração quanto na utilização do conhecimento científico, torna-se instrumento indispensável para os gestores públicos e privados, pois fornece as principais técnicas de avaliação do desenvolvimento do conhecimento, utilizando os recursos da **Bibliometria**. Estas são

questões fundamentais na gestão de C&T, sendo vasta a bibliografia relacionada. (Luukkonen-Gronow, 1987)

Como em qualquer outro campo social ou econômico, o diagnóstico de situação é instrumento fundamental para o estabelecimento, planejamento e implementação de políticas e atividades transformadoras. Para tanto é necessário avaliar e comparar todo momento, o desenvolvimento e a produção científica e tecnológica, estudando e identificando os determinantes e condicionantes desta produção, mapeando as instituições e os pesquisadores com relação a resultados alcançados, as potencialidades, as disponibilidades e as necessidades. (Kostoff, 1994).

Em documento elaborado para discussão de uma nova agenda em C&T para o Brasil, os analistas do Banco Mundial abordam a questão de o país estar preparado para as mudanças no setor "... *O ritmo acelerado da mudança tecnológica requer, rapidamente dos países, habilidade para produzir, avaliar, e selecionar o conhecimento. Os governos compreendem hoje as vantagens de uma estrutura e de políticas de ciência e tecnologia que, de um lado, permita a autonomia e a criatividade máximas para investigadores, e de outro, assegure a relevância da pesquisa e de seus resultados, em conformidade às necessidades do desenvolvimento econômico e social*". (Holm-Nielsen et al, 1996)

A moderna "inteligência" utiliza de forma crescente os métodos e os indicadores da Bibliometria e da Cienciometria, que permitem tratar, com ajuda da Informática, grandes quantidades de dados. Particularmente, estas ferramentas são imprescindíveis

para a elaboração de *mapas tecnológicos*. Estes mostram um panorama detalhado das linhas de investigação na área a ser estudada, por meio da análise do que se está publicando o patenteando. Tratados dinamicamente, ou seja, comparando mapas de diferentes períodos, podemos seguir a evolução no tempo das tecnologias e das linhas de investigação, "... *conhecer as relações e interações entre os diferentes elementos bibliográficos: investigadores, campos do conhecimento, setores..., tentando descrever o conteúdo das atividades e sua evolução, e seguindo fronteiras movediças*". (Callon et al, 1993)

Uma das formas de análise ainda não muito difundida no nosso meio, pelo menos no que diz respeito à avaliação da produção científica, tecnológica, é a co-ocorrência de palavras ou "*co-word analysis*". No entanto, está mais comumente associada ao tema indexação de trabalhos científicos, biblioteconomia e trabalhos relacionados com o termo "*webmetria*". Este termo foi proposto por Almind e Ingwersen em 1997, como sendo um ramo da informetria voltado à análise quantitativa do conteúdo e da estrutura das "home-pages" da "Word Wide Web". A co-ocorrência de palavras estuda o aparecimento conjunto de duas ou mais palavras representativas em campos tais como: títulos de artigos científicos ou patentes, resumos ou "*abstracts*", palavras-chave ou "*key-words*" em descritores e identificadores ou mesmo, no texto. As análises bibliométricas em geral, nos permitem detectar as áreas, "*clusters*" ou programas de investigação em que se subdivide um campo determinado, mas não nos permitem entrar no conteúdo dos documentos. Para consegui-lo é indispensável abordar a análise de

palavras nos títulos, *abstracts*, descritores, identificadores e texto, já que estas palavras explicam o conteúdo dos próprios documentos. Com a análise de "co-word", mapas tecnológicos podem ser obtidos e assim visualizarmos a estrutura do conhecimento que existe por trás do documento, bem como o que o tornou possível.

2-SITUAÇÃO ATUAL DA CIÊNCIA E TECNOLOGIA NO BRASIL

2.1. HISTÓRICO

O desenvolvimento formal da ciência brasileira tem suas raízes no Instituto Oswaldo Cruz, no Rio de Janeiro. A história da ciência e da tecnologia no Brasil remonta ao início do século XX, com as pesquisas de dois grandes cientistas: o médico sanitarista Oswaldo Cruz, pioneiro na medicina experimental no País e responsável pelas primeiras iniciativas na área de saúde e saneamento, e Carlos Chagas, que atribuiu à pesquisa básica a grande descoberta da doença que leva o seu nome. Assim, a ciência brasileira tem fortes raízes na área médica aplicada. (Tundisi-2003) Na área agrícola, o Instituto Agrônomo de Campinas, IAC, fundado por D. Pedro II em 1887, voltou às suas atividades, no início do século passado, para a resolução de problemas do setor, e, na década de 20, a pesquisa básica realizada pelo IAC já contribuía para o melhoramento genético das culturas de café e de algodão. Hoje, o Brasil tem mais de 60 mil cientistas e tecnólogos em atividade e está entre os 20 países que mais publicam artigos científicos em revistas internacionais. (Almeida et al, 1999)

Este universo de C&T somente começa a ganhar estrutura quando é criado o Conselho Nacional de Pesquisas, hoje Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e a Coordenação de Aperfeiçoamento do Ensino Superior (Capes), no início de

1950. Nos três decênios seguintes, o Brasil, em comparação com outros países de industrialização recente, construiu um expressivo parque de pesquisa. O modo como ele se construiu acompanhou, em vários aspectos, o modelo de industrialização em sua etapa de substituições de importações. Algumas das características básicas da pesquisa e desenvolvimento (P&D) naquele momento eram a horizontalidade e pouca seletividade e estavam vinculadas ao modelo então predominante na produção científica, que buscava, prioritariamente, criar uma massa crítica de recursos humanos qualificados.

Entretanto, o predomínio do modelo nacional-desenvolvimentista gerou a necessidade de buscar alguma articulação entre a produção técnico-científica e a produção agrícola e industrial. São exemplos dessa articulação para o desenvolvimento tecnológico, a criação da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e dos departamentos de P&D de empresas estatais, como o CENPES da Petrobrás, bem como a articulação entre o Instituto Tecnológico da Aeronáutica (ITA) e a Empresa Brasileira de Aeronáutica (Embraer).

No campo dos mecanismos de fomento, devem ser lembrados, além da Finep e CNPq, o Fundo de Tecnologia do Banco Nacional de Desenvolvimento Econômico (Funtec/BNDE). Porém, praticamente não se verificou extensão deste modelo para o campo das políticas sociais, a não ser em raras exceções como o Programa de Auto-Suficiência Nacional em Imunobiológicos (PASNI).

Desde 1980, vem se fortalecendo a articulação entre países em torno da idéia de que a pesquisa em saúde é uma ferramenta importante para a melhoria da situação

de saúde das populações, bem como, para a tomada de decisões na definição de políticas e no planejamento em saúde. Isso tem contribuído para a melhoria das ações de promoção, proteção, recuperação e reabilitação da saúde e a diminuição das desigualdades sociais. Organizações internacionais na área de saúde, com destaque para a Organização Mundial da Saúde (OMS), vêm desempenhando papel importante nesse movimento, no qual o Brasil deve buscar maior participação.

Investindo cerca de 0,7% de seu Produto Interno Bruto (PIB) em Ciência e Tecnologia, o Governo Federal consolidou, em meados dos anos 80 do século passado, a pesquisa acadêmica no País. Foram cerca de 60 mil os cientistas e tecnólogos em atuação no Brasil. Embora ainda seja um número pequeno em relação à população brasileira, representa um significativo avanço quando comparados há décadas anteriores e às nações do terceiro mundo. Toda esta organização institucional resultou no aumento da produtividade científica e em uma continuidade na formação de recursos humanos. Nos anos 90, o Brasil passou a formar 2.500 doutores por ano, em todas as áreas de ciência. Há um amplo suporte do governo federal aos programas de formação de mestres e doutores, no Brasil e no exterior. Este esforço resultou na diversificação das áreas de ciência básica, no aumento de programas de pós-graduação e no aumento significativo da produção científica publicada no País e no exterior em revistas indexadas.

O recente estudo realizado por King⁸ em 2004, mostrou que vários países melhoraram em muito sua classificação. Por exemplo, entre 1993 e 1997, o Brasil produziu 0,84% dos artigos científicos indexados no

mundo, isto é, registrados em banco de dados. De 1997 a 2001 esta produção elevou-se ao percentual de 1,21%.

Apesar dos avanços ocorridos, o país ocupa ainda uma modesta posição no panorama internacional da produção científica. Ainda assim, conseguiu construir uma tradição que se caracteriza pela capacidade de gerar internamente a imensa maioria dos recursos financeiros utilizados para o funcionamento da capacidade instalada de pesquisa e formar a quase totalidade dos recursos humanos para a pesquisa, de técnicos a doutores, dentro de suas fronteiras.

Esses dois fatos distanciam claramente o Brasil do panorama de pesquisa em saúde existente na maioria dos países em desenvolvimento. Como ocorre em vários países, a área da saúde também representa o maior componente de toda a produção científica e tecnológica. (2ª. Conferência Nacional de C&T/I, 2004).

2.2. RECURSOS HUMANOS EM C&T NO BRASIL

Entre os fatos mais promissores ocorridos no panorama da formação de recursos humanos para a pesquisa no país, na última década, destacam-se a criação do Programa Institucional de Bolsas de Iniciação Científica (PIBIC) e a descentralização geográfica dos programas de doutorado. Esta descentralização se for acompanhada de fluxo sustentado de recursos para as regiões Nordeste, Norte e Centro-Oeste, como está previsto na operação dos fundos setoriais do Ministério da Ciência e Tecnologia (MCT), poderá contribuir para a correção de uma das

sérias distorções na distribuição de recursos humanos em pesquisa, que é sua intensa concentração geográfica.

Ainda há carências importantes no que se refere ao desenvolvimento tecnológico no Brasil, sobretudo as relacionadas com a escassez de centros de excelência, profissionais e instituições capacitados para a gestão de processos de inovação, que se ajustem às exigências de qualidade e segurança dos órgãos reguladores.

A área de pesquisa em saúde em geral não difere das outras áreas quanto à distribuição de recursos humanos, porém apresenta alguns componentes mais concentrados que a média, como a pesquisa médica e odontológica em São Paulo, e outros menos concentrados, como a saúde coletiva, em que a presença da região Nordeste situa-se acima da média de participação¹⁹. (2ª. Conferência Nacional de C&T/I, 2004).

2.3. FOMENTO

O esforço governamental para fomentar a pesquisa em saúde é bastante significativo, mas insuficiente. No plano federal destacam-se as atuações do Ministério de Ciência e Tecnologia, por meio das suas agências de fomento (CNPq e Finep), do Ministério da Saúde por meio de suas instituições (Fundação Oswaldo Cruz, Instituto Nacional de Câncer, Instituto Evandro Chagas) e da contratação de projetos com grupos de pesquisa em diversos centros do país. Cabe ainda mencionar, a atuação do Ministério da Educação, especialmente na formação de recursos humanos e na disseminação de informações científicas por meio da Capes. No âmbito estadual

destacam-se o papel dos institutos de pesquisa vinculados às secretarias de saúde e algumas agências de fomento (FAPs), em particular a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) que, a partir da última década, vem desenvolvendo programas de apoio à pesquisa estratégica, de alto impacto nacional e internacional, em saúde. O Ministério da Saúde participa com cerca de 20% do total de desembolso público na pesquisa em saúde, enquanto o Ministério da Agricultura por meio da Embrapa comparece com quase o dobro (39%). Esse quadro mostra a necessidade de um deslocamento do papel do Ministério da Saúde para uma posição central na estruturação do fomento a pesquisa em saúde. Isso significa aumentar a capacidade indutora em P&D em Saúde aproximando-a das necessidades da política de saúde.

2.4. SETOR PRODUTIVO

Não é fácil quantificar os esforços de Ciência, Tecnologia e Inovação em Saúde no país. Para as atividades de Pesquisa e Desenvolvimento (P&D) em empresas, os dados são bastante precários, havendo pouca informação sobre o campo da saúde. Calcula-se que no Brasil, em 2000, foram investidos cerca de US\$ 13 bilhões em P&D, recursos estes majoritariamente do governo, aplicados em atividades desenvolvidas por instituições de ensino superior. O baixo investimento por parte do setor privado é atribuído ao caráter fortemente internacionalizado do complexo produtivo da saúde. Esta característica levou as empresas que vieram se instalar no país a optar pela realização de atividades de P&D em

suas matrizes no exterior. (2^a. Conferência de C&T e Inovação-2004).

Segundo Guimarães (2002) "*... O empresariado brasileiro nunca teve uma participação consistente no esforço nacional de C&T. Estima-se que, hoje, essa participação situe-se entre 10% e 20% do orçamento total de C&T no país, o que é bem pouco diante do tamanho do parque industrial*".

O complexo produtivo da saúde é formado por três grandes componentes, as indústrias químicas e de biotecnologia (fármacos, testes diagnósticos, vacinas e hemoderivados), as indústrias mecânicas, eletrônicas e de materiais (equipamentos, órtese e prótese e materiais de consumo) e as organizações de prestação de serviços. Nos últimos anos, os segmentos dos dois primeiros componentes apresentaram déficit comercial significativo, atingindo cerca de US\$3,5 bilhões em 2001. Deste déficit na balança comercial, 70% decorreu de relações com países desenvolvidos e 30% de relações com países que apresentam nível de desenvolvimento compatível com o brasileiro.

3 - AVALIAÇÃO EM CIÊNCIA E TECNOLOGIA

"A palavra avaliar vem do latim "valere". Esta apresenta, entre outras acepções, a de ser merecedor ou digno de alguma coisa. A avaliação, dentro de um determinado ramo do conhecimento, permite dignificar o saber, quando métodos confiáveis e sistemáticos são utilizados para mostrar à sociedade como tal saber vem-se desenvolvendo e de que forma tem contribuído para resolver os problemas que se apresentam dentro de sua área de abrangência". (Mugnani, Jannuzi e Quonian, 2004)

3.1. BIBLIOMETRIA

As técnicas quantitativas de avaliação podem ser subdivididas em bibliometria, *cienciometria*, *informetria* e, mais recentemente, *webmetria*. Todas têm funções semelhantes, mas, ao mesmo tempo, cada uma delas propõe medir a difusão do conhecimento científico e o fluxo da informação sob enfoques diversos. Existe, ainda, muita dificuldade em estabelecer onde termina uma e começa a outra.

Freqüentemente os termos *bibliometria* e *cienciometria* são utilizados indistintamente. Muitos são os autores que definem este conjunto de ferramentas para o estudo e avaliação da produção de publicações.

As primeiras noções do conceito de bibliometria remontam o início do século XIX. O primeiro estudo sobre os problemas desta disciplina se atribui a Cole e Eales em 1917. Gross e Gross em 1927 foram os primeiros a contabilizar, além dos documentos científicos, as

citações que os investigadores faziam em seus próprios documentos de trabalhos publicados anteriormente. Anos mais tarde, surgiu a palavra "bibliometria", termo atribuído a Pritchard para referir-se ao método de análise da informação escrita baseado na estatística de indicadores bibliográficos, ainda que alguns autores franceses atribuam a Paul Otlet várias décadas antes, em sua obra intitulada *Traité de documentati6n* em 1934. (Escorsa e Maspons-2001)

A Bibliometria se interessa principalmente pelos problemas de gestão das bibliotecas e dos centros de documentação, o que inclui a contagem de artigos e publicações assim como estudos mais detalhados sobre uma disciplina determinada, segundo Callon em 1993.

Segundo Pritchard em 1993 a Bibliometria é a ciência que estuda a natureza e o curso de uma disciplina, mediante a estatística e a análise das diversas facetas da comunicação escrita.

A Bibliometria é a exploração estatística das publicações. (Quoniam, 2001)

Considera-se Bibliometria o conjunto de estudos que tratam de quantificar o processo da comunicação escrita e a natureza e evolução das disciplinas científicas, tal e como se refletem na literatura, mediante a contagem e análise de diversas características da referida comunicação.

A Bibliometria é uma ferramenta de medida baseada na aplicação de métodos estatísticos e matemáticos que têm por objeto facilitar a comparação e a compreensão de conjuntos de referências bibliográficas. (Rostaing, 1995)

A Cienciometria, no entanto, designa aqueles trabalhos dedicados a análise quantitativa da atividade científica e técnica. (Callon¹⁵ et al, 1993)

Essa disciplina é dedicada exclusivamente à análise dos documentos relatados por investigadores e técnicos. De alguma forma, tenta identificar as leis que regem a atividade científica. A Cienciometria tem por objetivo o estudo dos aspectos quantitativos da criação, difusão e utilização da informação científica e técnica e a compreensão dos mecanismos da investigação como atividade social. A cienciometria busca as leis que regem a ciência, daí sua denominação de "*ciência das ciências*" por De Solla Price em 1963.

A Cienciometria é o resultado da convergência de dois movimentos, o americano, centrado no estudo da *ciência das ciências*; e o dos países do Leste Europeu, especialmente a União Soviética, com uma corrente denominada *Naukovodemia*. No primeiro caso, os estudos iniciais desenvolvidos estão associados ao nome de Derek de Solla Price, graças a seus livros publicados nos anos sessenta, em especial "*Little Science, Big Science*". O movimento da *Naukovodemia* tem um objetivo similar ao dos americanos, ou seja, estudar cientificamente a atividade de investigação para favorecer seu desenvolvimento. A primeira publicação desse grupo aconteceu em 1926 com um artigo de Borichevski, onde ele anuncia a constituição de um novo campo de investigação focado para o estudo da natureza *intrínseca da ciência*. (Escorsa e Maspons, 2001)

As ferramentas cienciométricas ajudam emergir, com maior ou menor êxito e precisão, as fronteiras de um campo de investigação, os limites de uma disciplina, as

variações que afetam essas disciplinas, as relações de influência que tem lugar no curso da elaboração de uma informação. Ajudam emergir a partir de alguns indícios deixados por seus atores no curso de suas práticas, as redes que constituem os campos da ciência.

A terminologia, no entanto, continua sendo confusa, sem que os especialistas se coloquem totalmente de acordo. Não obstante, nos últimos anos estas técnicas são objeto de um interesse crescente entre a comunidade científica assim como no mundo industrial.

A Cienciometria se baseia nas análises e computo de determinados indicadores bibliométricos, como autores de artigos, citações que aparecem na bibliografia de cada artigo, palavras contidas nos títulos, resumos, textos, descritores e identificadores das publicações. O termo bibliometria é amplamente utilizado para quantificar os processos de comunicação escrita.

A bibliometria estaria então inserida na cienciometria, *"... que nasce na confluência da documentação científica, a sociologia da ciência e a história da ciência, com objetivo de estudar a atividade científica como fenômeno social e mediante indicadores e modelos matemáticos"* (Bordons; Zulueta, 1999, p. 791)

3.2. INDICADORES BIBLIOMÉTRICOS

Para Escorsa e Maspons, 2001, existem três grupos principais de indicadores, primeiro, os que mostram o tamanho e as características da produção científica ou tecnológica; segundo os que estudam o impacto das publicações, medido através das citações que recebem; e

terceiro os que evidenciam os aspectos estruturais da ciência, que são os indicadores relacionais de primeira e segunda geração. Enquanto este último grupo serve para a elaboração dos *mapas tecnológicos*, os dois primeiros, denominados indicadores de atividade, constituem o núcleo em torno do qual se avalia a investigação.

O indicador bibliométrico básico mais simples consiste na contagem do número de publicações de determinados grupos, instituições ou países e sua distribuição. A simples contagem dos elementos bibliográficos como, autores, organismos, revistas, artigos, empresas, patentes, temas, datas de publicação, é geralmente considerado como uma medida da produtividade deste elemento. Em cifras absolutas estes indicadores podem ser interessantes, pelas evoluções temporais destas medidas são sempre muito mais significativas. Conhecer a velocidade ou a aceleração que experimentam certos indicadores pode ser de capital importância na tomada de decisões.

O primeiro grupo de indicadores pode também se denominar indicadores de publicação e medem a qualidade e o impacto das publicações. (Spinak, 1998) No âmbito dos indicadores de publicações existem três leis básicas em bibliometria: (Dahal 1998, Sancho 1990)

- Lei de Zipf que trata da freqüência da ocorrência de palavras no texto.

Esta lei, formulada claramente por George Kingsley Zipf, professor de lingüística da Universidade de Harvard, teve seus precursores nos trabalhos de Pareto e Estoup, respectivamente em 1896 e 1916, (Rousseau, 2002) afirma que "se as palavras que ocorrem, num texto de

tamanho considerável, forem listadas em ordem decrescente de frequência, então a graduação de uma palavra na lista será inversamente proporcional à frequência da ocorrência da palavra".

- Lei de Lotka que trata da produtividade dos autores em termos de publicações científicas.

Esta lei afirma que a proporção de autores que contribuem com um único trabalho deve ser 60% do total de autores.

- Lei de Bradford que trata da distribuição dos artigos pelas diferentes revistas.

Em 1934 Bradford formulou uma lei empírica para literatura científica em periódicos nos seguintes termos: "Se revistas científicas são organizadas em ordem decrescente de produtividade de artigos em determinado assunto, eles podem ser divididos dentro de núcleos de periódicos mais dedicados ao assunto e vários grupos de áreas que contem o mesmo número de artigos como o núcleo". (Dahal, 1998)

O segundo grupo, os indicadores de citação, mede o impacto das relações entre as publicações científicas por meio das medidas de fator de impacto, análise de citações, análises de co-citações, índice de impacto, índice de auto citação entre outros.

Do terceiro grupo, os indicadores relacionais mais importantes são as *citações comuns* ou *co-citações*, as *análises de colaboração científica* e a *co-ocorrência de palavras*, "*co-word analysis*". Em atenção à data em que se começou a utilizar estes indicadores, os primeiros são

denominados indicadores relacionais de primeira geração enquanto que os segundos se chamam de segunda geração.

3.3. "CO-WORD ANALISYS"

A "co-word analysis" estuda a co-ocorrência de palavras, ou seja, a aparição conjunta de duas ou mais palavras representativas em campos como títulos de artigos ou de patentes, resumos de artigos, descritores e identificadores - palavras-chave ou "key words" - códigos de classificação, reivindicações de patentes ou diretamente no texto livre. Os Descritores, objeto deste estudo, são palavras ou expressões utilizadas em indexação e tesouro para representar, sem ambigüidade, um determinado conceito. Por sua vez "tesouro", vem do latim "*thesaurus*", é uma palavra latina e foi empregada, a partir de 1500, para indicar um acervo ordenado de informações e conhecimentos.

Uma das primeiras obras a incluir esta expressão no seu título foi o "*Thesaurus linguae romanae et britannicae*", publicada em 1565, de autoria de Cooper. Em 1852 foi publicado por Peter Mark Roget, um estudioso do léxico, o "*English Thesaurus of Words and Phrases*", o qual reunia palavras pela ordem alfabética, de acordo com as "idéias que exprimiam, pelo seu significado". (Cavalcanti, 1978)

Mais especificamente, um tesouro atualmente reúne palavras escolhidas, destinadas à indexação e recuperação de documentos e dados num determinado campo de saber. Não é dicionário, nem vocabulário controlado. Trata-se de um instrumento que garante aos documentalistas e

pesquisadores a padronização da linguagem documentária. (Bruschini et al, 2001)

Um tesouro distingue-se de um simples vocabulário controlado por duas características. A primeira é que as palavras nele listadas não descrevem, mas antes significam: **cada palavra é um conceito**. Sendo assim, essas palavras que designam conceitos não são mais simples palavras, tornam-se "termos", ou ainda "descritores". A segunda característica é que todos os termos estão relacionados entre si; nenhum termo pode figurar no tesouro sem estar relacionado a algum outro, sendo essa relação determinada pelo seu significado. (Fundação Carlos Chagas)

Em co-ocorrência de palavras a repetição de duas palavras juntas, por exemplo, indica também uma relação de proximidade entre elas. Isto é, os conceitos que representam estão estreitamente associados, enquanto que se nunca aparecem juntos, significa que estas palavras não têm pontos de contato, ou seja, estão distantes uma das outras. A repetição de duas palavras juntas com fraca intensidade, pode indicar a aparição de uma nova tecnologia ou de um novo agrupamento ou "*cluster*". (Escorsa e Maspons, 2001)

Segundo Robredo e Cunha em 1998, em um artigo que pode ser considerado como um clássico na matéria, Whitaker definia, em 1989, a análise da co-ocorrência de palavras, a "co-word analysis", "*...como a utilização do 'comportamento' das palavras como um meio para elucidar as estruturas das idéias e outros problemas representados em conjuntos adequados de documentos*". Além de Whitaker, pode-se encontrar na obra de Callon, Law e Rip em 1986,

uma descrição bastante completa do desenvolvimento dos métodos de análise das co-ocorrências entre pares de palavras. Dentre os numerosos autores que nos últimos anos aplicaram a análise de co-ocorrência de palavras-chave ao estudo da situação ou da evolução de diversas áreas da ciência, merecem destaque, além dos já citados, Jean King em 1987, Leydesdorff em 1991, Callon, Courtial e Laville em 1991.

A co-ocorrência de palavras consiste, primeiramente, na detecção das palavras que caracterizam o conteúdo de determinados trabalhos sobre um tema e a contagem das aparições das palavras. Posteriormente verifica-se a co-ocorrência dessas palavras, ou seja, que palavras aparecem juntas e com que frequência. Os conceitos de proximidade ou distância entre essas palavras podem ser representados graficamente e constituem a base para a laboração dos "mapas tecnológicos".

Os mapas tecnológicos representam graficamente estas *proximidades* ou *distâncias*, dando uma visão da estrutura de relações que existe em um âmbito dado. A distância no mapa entre duas palavras, referentes a tecnologias, produtos, autores, empresas, assinala a maior ou menor relação entre elas. Esta proximidade pode ser quantificada mediante diversos índices e medidas e, graças a isto, visualizá-las graficamente por meio dos mapas tecnológicos.

Estes mapas mostram um panorama detalhado das linhas de investigação em uma área considerada, através da análise do que se está publicando ou patenteando. Tratados dinamicamente, ou seja, comparando mapas de

diferentes períodos, permitem seguir a evolução, no tempo, das tecnologias e das linhas de investigação.

Como afirma Van Raan em 1993, a "...representação visual de grandes volumes de dados proporciona em pouco tempo um panorama completo que se recorda facilmente. Além disso, há uma redução da informação, o que é muito importante. Atualmente encontramos uma enorme quantidade de "ruído" em uma enorme quantidade de dados disponíveis. É crucial filtrar os dados significativos." Os mapas agregam os dados de uma forma que nenhum perito é capaz de fazer, dada a impossibilidade de assimilar a quantidade de informação disponível.

Mapas baseados na "co-word analysis" ou co-ocorrência de palavras

Os mapas baseados nas citações permitem detectar as áreas, "clusters" ou programas de investigação em que se subdivide um campo do conhecimento, mas não nos permitem entrar no conteúdo dos documentos. Para consegui-lo é indispensável abordar a análise de palavras, já que estas explicam o conteúdo dos próprios documentos. Com a "co-word analysis" e os mapas podemos visualizar a estrutura do conhecimento que há por trás do documento e o que o tornou possível.

O método "co-word analysis" começa surgir em meados dos anos oitenta, a partir de um projeto francês desenvolvido pelo "Centre de Sociologie de l'Innovation (CSI) de l'École des Mines de Paris" e l'Institut National d'information Scientifique et Technique "(INIST) com a finalidade de processar os termos de indexação de Pascal, uma base de dados científica e multidisciplinar que agrupa os trabalhos de autores europeus,

principalmente como Callon, Courtial e Laville em 1991. Como parte deste projeto também se desenvolveu uma ferramenta informática, o *Leximappe* que permitia desenvolver este tipo de análise de uma forma automatizada e confiável.

A técnica "co-word" não se limita estudar a co-ocorrência entre palavras, presta-se também ao estudo da co-ocorrência entre palavras e empresas, palavras e datas, palavras e códigos de classificação, isto é, entre os diversos campos de informação de uma base de dados.

Os constantes avanços da informática têm favorecido o desenvolvimento de ferramentas estatísticas úteis para a cienciometria. Estão disponíveis no mercado, programas que agilizam as tarefas de processamento de documentos e multiplicam as possibilidades de análises, combinando múltiplos indicadores. (Dousset, 1999) Entretanto, especificamente para análise cienciométrica, existem até o momento, escassos programas informáticos.

Dentre os programas mais conhecidos temos, o *Leximappe*, *Dataview* e o *Tétralogie*, do *Institut de Recherche Informatique* (IRIT) de Toulouse. Estes três programas têm facilitado o processamento da informação e apresentação dos resultados, tanto para a geração de listas e matrizes, contagem simples, como para a análise da co-ocorrência das palavras, chegando inclusive à elaboração de mapas tecnológicos. (Maspons, 1996)

Assim, o principal objetivo deste trabalho é estudar o campo da ciência da saúde, no Brasil por meio de métodos de análise, devidamente validados, em avaliação de C&T.

Demonstrar as etapas de trabalho, a potencialidade da "*co-word analysis*" e verificar se esta técnica de análise incrementa os resultados observados, com as técnicas bibliométricas, já tradicionalmente utilizadas, de maneira a poder qualificar as instituições, as disciplinas e os profissionais produtores do conhecimento técnico e científico, nas diferentes áreas que compõe este campo.

Discutir uma nova abordagem para a área, ampliando as análises bibliométricas tradicionais e consolidar no nosso meio científico a utilização dessa área de estudos quantitativos. Explorar as bases de dados também como um instrumento para desenvolver atividades de análise, enfatizando as possibilidades que este tipo de avaliação tem, para elevar o nível das informações, assim como para explorar as técnicas bibliométricas no gerenciamento de políticas e de tomadas de decisão.

4 - OBJETIVOS

4.1. OBJETIVO GERAL

O objetivo deste trabalho é avaliar a produção científica brasileira no campo da saúde e sua repercussão internacional no período de 1990 a 2002.

4.2. OBJETIVOS ESPECÍFICOS

Analisar a produção científica no campo da saúde brasileira utilizando-se de indicadores bibliométricos, bem como estudá-la através da co-ocorrência de palavras-chave, descritoras do campo, com a finalidade de:

1- Responder o que é o campo do conhecimento no campo da saúde atualmente no país: estudar os indicadores de produção e impacto no tempo; qual a complexidade da produção científica no campo da saúde brasileira; quais os processos de geração e divulgação dessa produção; mecanismos da investigação científica enquanto atividade social. Os autores serão estudados como padrões de comportamento.

2- Demonstrar a potencialidade da técnica de "*co-word analysis*" e verificar de que forma ela complementa os resultados observados, com as técnicas bibliométricas tradicionalmente, utilizadas.

5 - MATERIAL E MÉTODOS

5.1. UNIVERSO DO ESTUDO:

A pesquisa das publicações, indexadas, de autores brasileiros, no período de 1990 a 2002, foi realizada utilizando-se duas bases de dados: **ISI-Science Citation Index Expanded (SCI-EXPANDED)** e **ISI-Social Sciences Citation Index (SSCI)** ambas da **Web of Knowledge, do Institute for Scientific Information (USA)**. Essas bases contem publicações indexadas a partir de 1945 e de 1956 respectivamente.

O Institute for Scientific Information, (ISI) mantêm, hoje, a maior base de dados bibliográfica, multidisciplinar de informação em pesquisa do mundo. Esta base inclui mais de 8 mil jornais e revistas em 164 áreas do conhecimento, indexando dados bibliográficos completos, referências citadas pelo autor e resumo. Dentre esses estão 15 revistas brasileiras. O *ISI* foi criado em princípios da década de sessenta por Eugene Garfield, discípulo de De Solla Price, na Filadélfia.

Os produtos e serviços do ISI incluem além de dados atualizados constantemente, citações retrospectivas indexadas, serviços ao usuário, informações em produtos químicos.

Estas bases de dados foram escolhidas para garantir uma padronização das informações visando à construção do banco de dados compatível com os pacotes estatísticos que foram utilizados na análise, ou seja, a base garante a obtenção de resultados em estruturas

padrões, especialmente projetados para dados bibliométricos.

Foi utilizada a totalidade das informações disponíveis no campo da saúde, no período de 1990 a 2002 e foram definidas, para busca do tema a ser estudado, todas as áreas relacionadas que contem as publicações do campo. Para tal foram selecionadas as revistas indexadas que se distribuem entre as 50 áreas relacionadas ao campo da saúde, de acordo com a classificação ISI- De Luxe.

As revistas selecionadas para busca de dados, pertencem ao campo saúde, cujas áreas estão relacionadas no Quadro 1.

Quadro 1 - Áreas do Conhecimento - ISI

1- Anaesthesia e Intensive Care	26- Medical Research, Diagnosis and Treatment
2- Biochemistry e Biophysics	27- Medical Research, Organs and Systems
3- Biology	28- Medical Research General Topics
4- Biotechnology e Applied Microbiology	29- Microbiology
5- Cardiovascular e Hematology research	30- Molecular Biology and Genetics
6- Cardiovascular e Respiratory System	31- Neurology
7- Cell e Developmental Biology	32- Neuroscience and Behavior
8- Chemistry	33- Oncogenesis and Cancer Research
9- Chemistry e Analysis	34- Oncology
10- Clinical Immunology e Infectious Disease	35- Ophthalmology
11- Clinical Psychology e Psychiatry	36- Orthopedics, Rehabilitation and Sport Medicine
12- Dentistry Oral surgery medicine	37- Otolaryngology
13- Dermatology	38- Pediatrics
14- Endocrinology, Metabolism e Nutrition	39- Pharmacology and Toxicology
15- Endocrinology, Nutrition e Metabolism	40- Physiology
16- Entomology Pest Control	41- Psychiatry
17- Environment Ecology	42- Psychology
18- Environmental Medicine e Public Health	43- Public Health and Health Care Science
19- Experimental Biology	44- Radiology, Nuclear Medicine and Imaging
20- Food Science Nutrition	45- Rehabilitation
21- Gastroenterology and Hepatology	46- Reproductive medicine
22- General and Internal Medicine	47-Research Lab.Medicine and Medical Technology
23- Health Care Science and services	48- Rheumatology
24- Hematology	49- Surgery
25- Immunology	50- Urology and Nephrology

Algumas revistas estão indexadas em mais de uma área relacionada ao campo da saúde. Por este motivo, o total de revistas selecionadas somaram 3.543. Após terem sido identificadas duplicações, 477 revistas foram subtraídas e o arquivo final para busca contou com 3.066 revistas.

5.2. BUSCA DE DADOS:

Para padronizar a pesquisa na base de dados ISI, utilizou-se o mecanismo de busca "GENERAL SEARCH" tendo sido pesquisadas as 3.066 revistas nos campos: SOURCE TITLE/Fonte. A seleção de publicação para o país foi feita utilizando-se o campo "ADDRESS"/Endereço, colocando-se o nome do país, "BRAZIL or BRASIL" e o período dessa busca foi fixado de 1990 a 2002, no campo "TIME"/Tempo. Com este critério definido, foi realizada a pesquisa das publicações brasileiras, no campo da saúde, no período fixado.

Resumo de busca: Source Title = total das 3.066 revistas do campo da saúde e áreas relacionadas.
Address=Brazil or Brasil; **DocType**=All document types;
Language=All languages; **Databases**=SCI-EXPANDED, SSCI;
Publication data= 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998,1999,2000,2001 e 2002. (sorted by latest date)

5.3. ANÁLISES BIBLIOMÉTRICAS

Os pacotes estatísticos utilizados para este tipo de análise exigiram padrões rígidos na construção dos

bancos de dados, sem os quais tornar-se-ia inviável o processamento e análise de cada variável. Os registros oriundos das duas bases pesquisadas passaram por um processo de correção e recodificação para obtenção de uniformidade na sua apresentação.

5.3.1. TRATAMENTO DOS DADOS:

Os dados obtidos foram submetidos a um tratamento onde foi fixado em 6 o número máximo de autores por artigo, conforme recomendam os requisitos de Vancouver (1997), para a inclusão de autores nas referências bibliográficas, bem como os respectivos endereços para a identificação de parcerias. Foi acrescentado um campo para registro do total de autores, campo (TA), de forma a garantir essa informação para a análise de frequência de autores por artigo. Todos os dados migrados em plataforma Access, uma vez corrigidos e codificados, foram exportados para planilhas de Microsoft Excel de forma padronizada, dando origem às variáveis necessárias para a análise.

O total de 38.349 registros desse banco de dados foi analisado por meio do Statistical Package for Social Science - SPSS. Para estudar os indicadores de publicação e de citação, ou seja, o número médio de publicações por autor, por cidade, por ano, por área, número médio de referências, taxas de crescimento para as publicações entre outros, foram utilizadas técnicas de extensão bibliométrica e técnicas de análise de citações. Aplicou-se a lei do quadrado inverso, Lei de Lotka, para verificar a contribuição dos autores. Para estudar a distribuição dos artigos pelas diferentes revistas,

aplicou-se a lei de dispersão, de Bradford. Técnica de análise de "cluster" foram utilizadas para estudar os padrões da produção científica brasileira.

5.4. "CO-WORD ANALYSIS"

O software utilizado para as análises de "co-word" também exigiu padrões rígidos na construção de um arquivo a partir das informações obtidas. Para este grupo de análises foi utilizado o software "**Tétralogie**". Este software foi desenvolvido pelo Institut de Recherche Informatique (IRIT) de Toulouse, França. Seu idealizador foi o pesquisador Bernard Dousset.

5.4.1. TRATAMENTO DOS DADOS PARA A "CO-WORD ANALYSIS"

As informações obtidas das bases de dados, em formato html (Hyper Text Markup Language), foram migradas para um arquivo em formato texto (ASCII). Foram selecionados somente dados de interesse para o trabalho e como exemplo, um registro do arquivo em formato de texto (txt), é demonstrado na sua íntegra no quadro 2 abaixo, onde:

AU1...AU6- nome de autor da publicação; **TA**- total de autores por publicação; **TI**- título; **SO**- fonte, revista; **LA**- idioma; **DT**- tipo de arquivo; **NR**- número de referências; **ISSN**- número de identificação internacional da revista; **PU**- editor; **CO1...C06**- endereço de autores; **DE**- descritores; **ID**- identificadores; **TC**- número de citações recebidas; **BP**- número da primeira página; **EP**- número da última página; **P**- páginas; **PY**- ano de

publicação; **PD-** mês de publicação; **VL-** Volume; **RV-** dados da localização da publicação na revista; **PI-** informações da editora.

Quadro 2 - Exemplo de registro do arquivo em formato texto.

```

AU1 ROCHA MD -
AU2 COSTA NSS
AU3 COSTA JCG
AU4 ANGELO MTDL
AU5 ANGELO JRL
AU6 SONODA L
TA 8
TI IDENTIFICATION OF ASCARIS IN THE BILIARY-TRACT
SO ABDOMINAL IMAGING
LA English
DT Article
NR 13
ISSN 0942-8925
PU SPRINGER VERLAG
C01 ROCHA MD, UNIV SAO PAULO, FAC MED, BR-05508 SAO PAULO, BRAZIL.
DE ASCARIASIS, BILIARY TRACT, CT
ID PANCREATIC ASCARIASIS
TC 10
BP 317
EP 319
PG 3
PY 1995
PD AUG
VL 20
RV 20: (4) 317-319 JUL-AUG 1995
PI NEW YORK

```

5.4.2. CARACTERÍSTICAS DO SOFTWARE "TÉTRALOGIE"

Para a contagem, identificação e análise de palavras-chave - que se refere ao campo conceitual no qual a publicação se enquadra - foi utilizado o software "**Tétralogie**". Utilizou-se a metodologia de "Co-word Analysis" que permite observar a freqüência da ocorrência de palavras isoladamente e detectar palavras que aparecem juntas freqüentemente nas publicações, evidenciando uma proximidade.

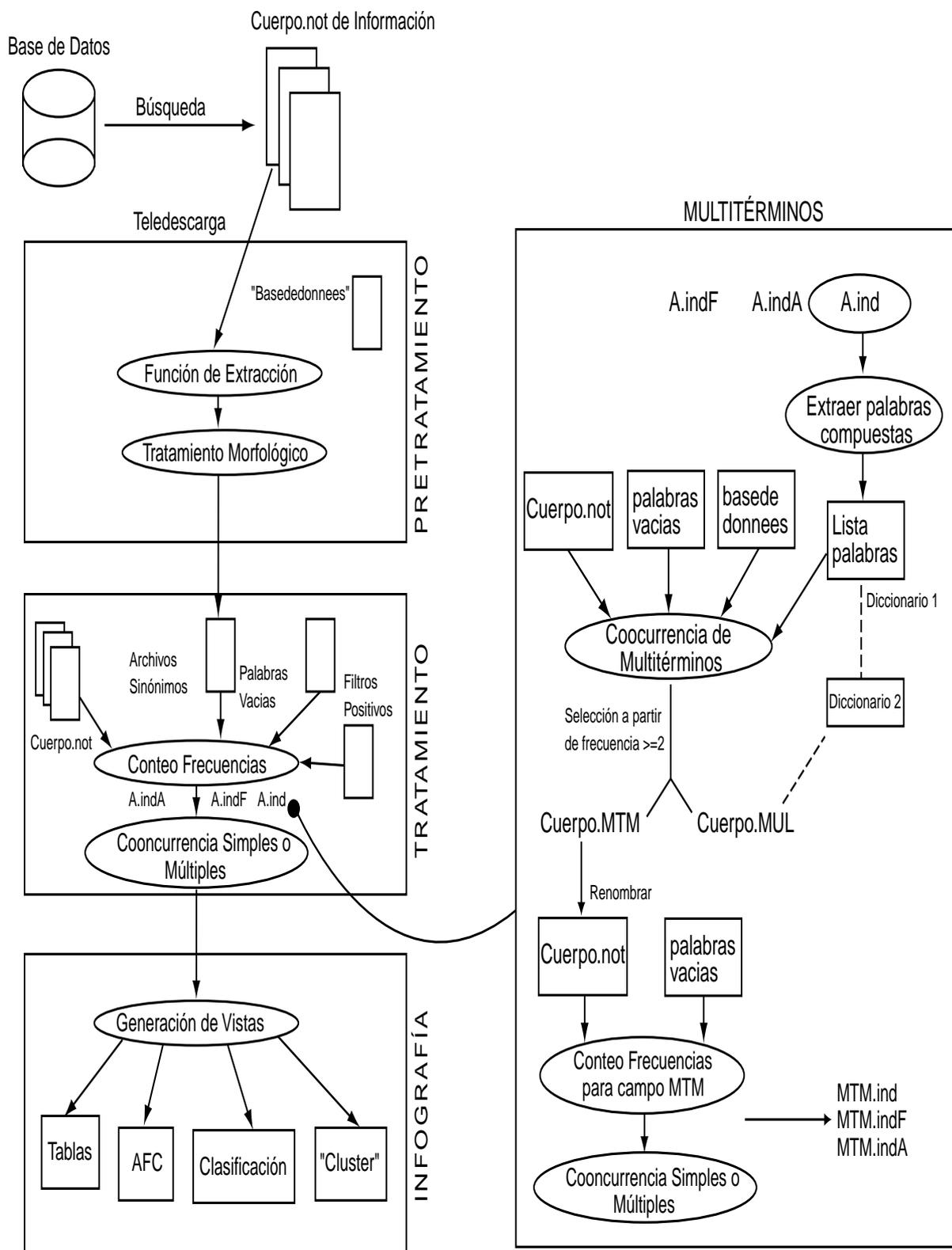
A partir dessas análises é possível visualizar a informação em forma de mapas e agregá-la, utilizando

técnicas de análise hierárquica. Com este software é possível, além de contar autores, instituições e publicações, analisar colaborações entre autores, entre autores e instituições, instituições e palavras-chave, autores e palavras-chave, entre outras. Isto nos possibilita compreender com detalhe a estrutura da atividade científica.

O *Tétralogie* (mimeo 1997) supõe que a fase de obtenção de dados foi realizada, que os dados estejam disponíveis e que se investigará um tema. Supõe também que a fonte de informação é conhecida e que sua descrição está disponível. Quando os dados provêm de uma base de dados não estruturada e existe o interesse de estudar um campo texto completo, dentro de uma base de dados estruturada (Título, Abstract ou Descritores ou Identificadores) o pacote *Tétralogie* dispõe de uma fase de Pré-tratamento, com modelos Estatísticos e modelos Morfológicos. A versão utilizada neste trabalho ainda não incorpora modelos lingüísticos, semânticos (análises dos diferentes sentidos de uma palavra) e sintáticos (relações entre palavras por seu significado), os quais permitiriam um tratamento mais completo de documentos em formato texto, mas não foram objeto do presente estudo. No *Tétralogie* o processo de tratamento da informação se dá como no esquema 1.

Esquema 1 - Metodología de trabajo del Software TÉTRALOGIE

Fonte: (Dousset e Dkakis, 1997)



5.5. PROCESSO DE TRATAMENTO DA INFORMAÇÃO

5.5.1. PRÉ-TRATAMENTO DOS DADOS

Esta etapa compreende analisar dados a partir do corpo de informação oriunda do arquivo em formato texto, contando 38.349 registros, base do estudo cienciométrico em questão, elaborado a partir dos registros obtidos de uma ou de várias bases de dados.

Esses registros foram unificados, as duplicidades foram eliminadas e os campos foram renomeados entre outras operações. Este arquivo junto com uma nova estrutura da base de dados, a "*basededonnées*", permitirá iniciar o processo de tratamento da informação.

5.5.2. ESTRUTURA DA "BASEDEDONNEES"

Para que o programa lesse o arquivo texto, foi necessário alimentá-lo com uma espécie de roteiro das informações, denominado "*Basededonnées*". Trata-se de arquivo criado a partir das informações contidas no arquivo em formato texto, para descrever a estrutura dos campos da base de dados e isolar a informação pertinente aos campos de interesse para o estudo, mediante a identificação e uso dos separadores. A seguir observa-se um exemplo de alguns campos na "*Basededonnées*", utilizada para o processamento do banco de dados.

Basededonnées para o processamento do banco de dados.

Exemplo:

descripteurs des champ #					
#	nom	abrev	champ	visible	Separateurs #
AU1	AU1	AU1	True	/n	"AU2"AU3"AU4"AU5"AU6"
TA		TA	TA	False	/n"
TI		TI	TI	False	/n"
SO		SO	SO	True	/n"

O arquivo da "*basededonnées*" pode sofrer modificações sucessivas de acordo com as necessidades que surjam durante a etapa de tratamento e os objetivos da análise. Após a definição do corpo de informações a serem tratadas, a primeira etapa de análise ou de pré-tratamento foi a extração dos dados e a primeira contagem simples de frequências de palavras, a partir da qual criaram-se os dicionários de sinônimos. Foram realizadas contagens simples para alguns campos como os que continham os descritores, identificadores, autores, instituições, fonte e ano de publicação. Os resultados foram expressos em arquivos que podem ser visualizados com um software editor de texto chamado UEDIT32 - UltraEdit-32 - Professional Text/Hex Editor (IDM Computer Solutions, Inc.)

Por meio das análises realizadas obtivemos, inicialmente, uma contagem de 42.233 diferentes descritores e 41.281 diferentes identificadores, distribuídos entre os 38.349 registros do banco de dados. Os arquivos gerados nessa contagem contem as palavras encontradas nos artigos e pode apresentar-se em ordem

alfabética ou de frequência de aparições, como mostra o exemplo a seguir, referente à frequência de descritores onde, "BRAZIL" é a primeira palavra mais encontrada e "DAY-CARE SURGERY" finaliza a lista de 42.233 descritores, com uma ocorrência.

Exemplo de arquivo de descritores e sua respectiva frequência

FREQ.	DESCRITORES
721	BRAZIL
360	EPIDEMIOLOGY
202	NITRIC OXIDE
201	RAT
187	TRYPANOSOMA CRUZI
121	HYPERTENSION
114	RATS
112	RISK FACTORS
111	ULTRASTRUCTURE
109	CYTOKINES
108	INFLAMMATION
103	FISH
101	DIAGNOSIS
101	ANXIETY .. até..
1	"DAY-CARE" SURGERY

Uma das fases, extremamente importante, desta etapa de pré-tratamento é a de **Tratamento Morfológico**. Esta é uma função que permite a equiparação de palavras ou a criação de sinônimos para as palavras. Depois de criado o dicionário de sinônimos, cria-se os demais arquivos que são os filtros. O software prevê uma forma automática de dicionários, porém insuficiente para os problemas encontrados. Desta forma, estes sempre devem ser revisados e ajustados, como no exemplo a seguir.

Tratamento Morfológico e Sinônimos para os Descritores

DESCRITORES	SINÔNIMOS
IMMUNODEFICIENCY VIRUS	HIV/AIDS
BRAZIL	BRAZIL
BR-	BRAZIL
APOPTOSIS	APOPTOSIS
T CELL APOPTOSIS	APOPTOSIS
MACROPHAGE APOPTOSIS	APOPTOSIS
IMMUNOMODULATION AND APOPTOSIS	APOPTOSIS
INDUCED APOPTOSIS	APOPTOSIS
AIDS	HIV/AIDS
HIV/AIDS	HIV/AIDS
IMPACT OF AIDS	HIV/AIDS
RESPONSES TO AIDS	HIV/AIDS

Após tratamento com filtros de sinônimos, os novos arquivos apresentaram uma listagem por ordem de frequência de todas as palavras presentes nos campos, a serem estudados, sem duplicações. O arquivo de descritores passou a contar com 39.455 palavras enquanto o de identificadores foi de 40.160. O produto inicial da contagem realizada para o campo de autores continha 148.971 registros, com os autores das publicações registrados no banco de dados, até o sexto autor. Neste caso, além das duplicações, em alguns registros as informações presentes no campo endereço, devido sua formatação original na base de dados, "invadiram" o campo de autores. Após a retirada de dados não pertinentes e tratamento morfológico este campo livre de duplicações e interferências somou 73.644 autores.

Desta forma foram realizadas as análises e os cruzamentos necessários para o estudo dos indicadores bibliométricos desejados. O mesmo procedimento foi

realizado para todas as informações contidas nos diferentes campos de interesse para o estudo.

5.5.3. TRATAMENTO DOS DADOS

A simples contagem permitiu análises estatísticas dos diferentes campos objeto do estudo. Uma segunda contagem dos arquivos listados por frequência, desta vez com os arquivos de filtros e sinônimos, geraram novos arquivos que foram utilizados na etapa das análises de "co-word".

Algumas das análises realizadas foram feitas por amostragem, a saber: dos descritores e dos identificadores foram selecionadas 1% das palavras de maior ocorrência respectivamente, 39.455 e 40.160 palavras. Para autores foi fixado um mínimo de 3 publicações no período estudado; e as análises de "clusters", um mínimo de 10 publicações no período.

Para as etapas de Pré-tratamento e Tratamento é muito importante o critério dos peritos nas áreas do conhecimento, objeto do estudo. Da qualidade dos dicionários, filtros, sinônimos, palavras vazias, multitermos, dependerá a qualidade das análises e interpretações posteriores.

5.6. MATRIZES E MAPAS TECNOLÓGICOS

Nesta fase de elaboração dos mapas tecnológicos, foram utilizadas matrizes com os resultados da co-

ocorrência de palavras, geradas do cruzamento das variáveis estudadas e com as devidas aplicações de dicionários, filtros e sinônimos.

- **Matrizes**

As matrizes oriundas das análises de cruzamentos das informações desejadas, permitiram organizar, transformar e refinar os dados, por meio das eliminações de duplicidades ou aplicando-se operações matemáticas nas linhas ou colunas quando a tabela original se apresentava muito dispersa. Os dados das tabelas são representados nas linhas, pelos indivíduos e nas colunas pelas variáveis relativas a cada indivíduo. As variáveis podem ser quantitativas (valores absolutos) ou qualitativas (valores relativos).

As tabelas podem receber diferentes formas de ordenação da informação de modo a conseguir melhores resultados das análises de cruzamento dos dados. As formas de ordenação utilizadas foram por afinidade bem como, por consistência de colunas e linhas. Outro recurso disponível no software, é a ordenação por ordem alfabética, dos indivíduos ou variáveis. A ferramenta de "Zoom" permite visualizar a qualidade da estrutura das ordenações obtidas e os "Histogramas" explicam a intensidade das relações entre linhas e colunas.

- **Mapas tecnológicos**

Os mapas tecnológicos são mapas fatoriais que resultaram das transformações matemáticas das tabelas ou da aplicação de métodos estatísticos integrados ao software conforme explicamos a seguir.

5.6.1 MÉTODOS ESTATÍSTICOS INTEGRADOS AO SOFTWARE "TÉTRALOGIE" UTILIZADOS NAS ANÁLISES DE "CO-WORD"

- **Análise dos Componentes Principais (ACP)**

É um método de análise de dados que permite identificar os eixos principais de uma nuvem de pontos. Quando os dados são da mesma natureza, cruzamento de autor-autor, por exemplo, foi utilizada a métrica euclidiana. Se dois indivíduos aparecem próximos no mapa, sua proximidade encontrará uma explicação no círculo de correlação, com a posição das variáveis no mesmo setor, correlação positiva ou no setor oposto, correlação negativa.

- **Análise dos Componentes Principais Reduzida (ACPR)**

Este método foi utilizado para normalizar os dados e equilibrar as variáveis analisadas. A redução dos dados permitiu considerar as variáveis de menor valor absoluto.

- **Análise Fatorial de Correspondência (AFC)**

Este método também permite identificar os eixos principais de uma nuvem de pontos, é realizado sobre os valores relativos dos dados, é recomendado para parâmetros qualitativos. No mesmo mapa podem ser visualizados tanto os indivíduos como as variáveis. A distância entre um indivíduo e uma variável é interpretada como uma relação de proximidade e de dependência. Aplica-se como método de análise tanto em tabelas de contingência como a matrizes de cruzamento.

- **Classificação Hierárquica Ascendente (CHA)**

Esta classificação gera uma árvore plana que se pode cortar a diferentes níveis para determinar as classes mais ou menos extensas e numerosas. Estas classes são obtidas mediante a agregação dos elementos mais próximos, distância mínima, considerados como uma mesma classe. A métrica utilizada também é a euclidiana.

5.7. LIMITAÇÕES DO ESTUDO

Não obstante a qualidade das informações pesquisadas nas bases do ISI e desta ser um dos mais importantes sistema de informações referente à produção bibliográfica em âmbito internacional, é necessário reconhecer que é limitada em sua abrangência geográfica, além de ter rígidos critérios de inclusão de revistas.

Para Garfield, criador da base de dados ISI, a seletiva cobertura oferecida pela base, não é devido tão somente a motivos econômicos, mas, ao fato de que, em algumas áreas do conhecimento, as revistas realmente importantes que são lidas e citadas, são relativamente pouco numerosas, refletindo os preceitos da Lei de Bradford. (Pellegrini Filho e Goldbaum, 1997) Grande parte das publicações brasileiras está registrada em periódicos não indexados na base ISI, especialmente em determinadas áreas do conhecimento mais direcionadas a temas do interesse nacional.

Assim, a produtividade científica brasileira é, em parte, prejudicada, mas esta base foi escolhida porque apesar de representar um corte dessa produção, apresenta as publicações de grande impacto e maior repercussão.

Também colabora para as limitações o fato de haver outras formas de produção científica e bibliográfica além da publicação de artigos em periódicos. As formas de divulgação dos resultados da produção científica variam segundo áreas do conhecimento, de modo que a contabilização destes resultados, limitada ao número de artigos publicados, introduz um viés em seu cálculo, privilegiando algumas destas áreas em detrimento de outras.

O motivo de não termos utilizado os dados de fontes específicas para publicações da América Latina e Caribe, como a base LILACS em Ciências da Saúde bem como, bases de reconhecimento mundial neste campo como a Medline, deve-se ao fato destas não oferecerem as informações importantes para este estudo, como os indicadores de impacto das publicações, por exemplo o número de citações por artigo. Modelo semelhante ao sistema de informação da "web of Knowledge" é o SciELO que, no entanto, teve seu início em 1997.

Desta forma, fica registrado que não se estudou o universo da produção científica brasileira na área da saúde e que se admite algumas limitações no que tange à adequada mensuração dessa produção.

Em função dos objetivos deste trabalho, apresentaremos os resultados obtidos por meio das análises bibliométricas tradicionais, com o auxílio do software Statistical Package for Social Science - SPSS.

A "co-word analysis" será apresentada de forma a evidenciar o potencial de análises e resultados que este tipo de ferramenta pode oferecer, mostrando exemplos com as diferentes variáveis estudadas.

6 - RESULTADOS E DISCUSSÃO

6.1. ANÁLISES BIBLIOMÉTRICAS

6.1.1 INDICADORES DE PRODUÇÃO E IMPACTO NO TEMPO

Na tabela 1 e gráfico 1, a produção intelectual na área da Saúde no Brasil, reflete o número de publicações distribuídas entre 3.066 revistas pesquisadas, de diversas áreas do conhecimento relacionados com o campo da saúde e de acordo com os registros da base de dados da ISI, "web of Knowledge", num período estudado de 13 anos, de 1990 a 2002.

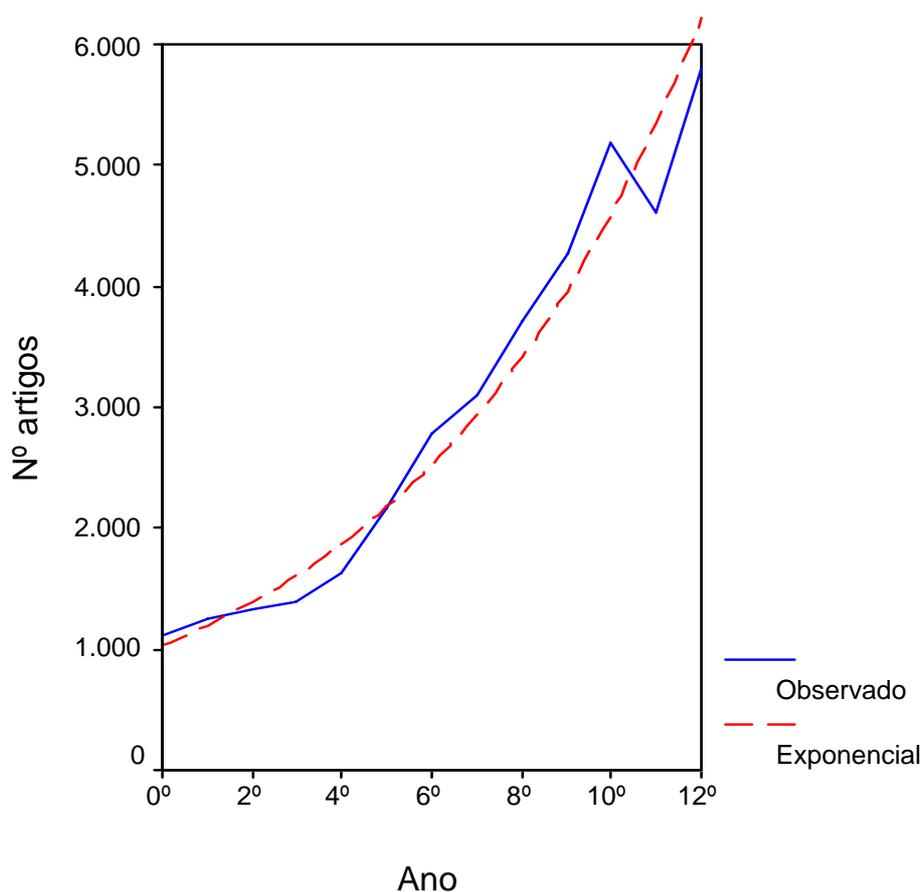
Tabela 1 - Número de publicações brasileiras, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002, no campo da saúde.

Ano	Frequência	Percentual	Percentual Cumulativo
1990	1111	2,9	2,9
1991	1242	3,2	6,1
1992	1334	3,5	9,6
1993	1396	3,6	13,3
1994	1638	4,3	17,5
1995	2162	5,6	23,2
1996	2772	7,2	30,4
1997	3092	8,1	38,5
1998	3718	9,7	48,1
1999	4278	11,2	59,3
2000	5190	13,5	72,8
2001	4607	12,0	84,9
2002	5809	15,1	100
Total	38349	100	100

Os resultados encontrados evidenciam que a produção científica em saúde no Brasil cresceu a uma taxa

anual média de 14,98% nesse período. Em termos comparativos, no curto intervalo de duas décadas, entre 1981 e 2000, o Brasil passou da 28ª para 17ª posição no ranking mundial de produção de ciência. Nesta posição, o Brasil está à frente da Bélgica, Escócia e Israel, entre outros, e bem próximo da Coreia do Sul, Suíça, Suécia, Índia e Holanda. (IBICT 2004)

Gráfico 1 - Crescimento da produção científica brasileira, registrado na base de dados da "web of Knowledge", no período de 1990 a 2002, no campo da saúde.



6.1.2. Impacto das publicações no tempo - Citações por ano

A busca do reconhecimento entre os pares na carreira de pesquisadores e cientistas é percebida com

seriedade. A citação na ciência é um veículo de comunicação, onde os diálogos se dão de forma não presencial, podem ser considerados "a teia invisível ou virtual da ciência" a qual é responsável pela disseminação da ciência.

Os diálogos entre cientistas e pesquisadores vêm sendo estudados desde os anos 50 através das chamadas redes de comunicação científica. No entanto, a proliferação destes estudos deu-se a partir da década de 60 com Derek de Solla Price, que pesquisou a mensuração da literatura científica através das análises de citações. Nessas análises percebe-se que a ciência é construída por meio de diálogos estabelecidos entre o autor citado e aquele que o cita; outro aspecto relevante abordado é o direcionamento evolutivo da ciência e tecnologia. (Pinto e Lima, 2002)

Tabela 2 - Número de citações em artigos brasileiros, no campo da saúde, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.

Ano	No. de Citações	No. de Artigos	Média de citação por artigo
1990	8524	1111	7,67
1991	10759	1242	8,66
1992	10395	1334	7,79
1993	11813	1396	8,46
1994	12302	1638	7,51
1995	13177	2162	6,09
1996	13541	2772	4,88
1997	11376	3092	3,68
1998	10788	3718	2,9
1999	7445	4278	1,74
2000	11135	5190	2,15
2001	5848	4607	1,27
2002	1560	5809	0,27

De acordo com a tabela 2 e gráfico 2, a análise das citações para os artigos brasileiros, publicados no período, revela que enquanto o número de artigos aumenta, o número de citações é mais ou menos estável com valores que variam entre 8 e 13 mil citações por ano. Em consequência a taxa média de citação por artigo diminui.

Os valores encontrados para os anos de 2001 e 2002 necessitam um período maior de tempo para inferirmos algum julgamento, porque o número de citações das publicações mais recentes é menor, pois estas necessitam de tempo para serem citadas.

Gráfico 2 - Curva das publicações e citações no campo da saúde, registrados na base de dados da "web of Knowledge", no Brasil, no período de 1990 a 2002.

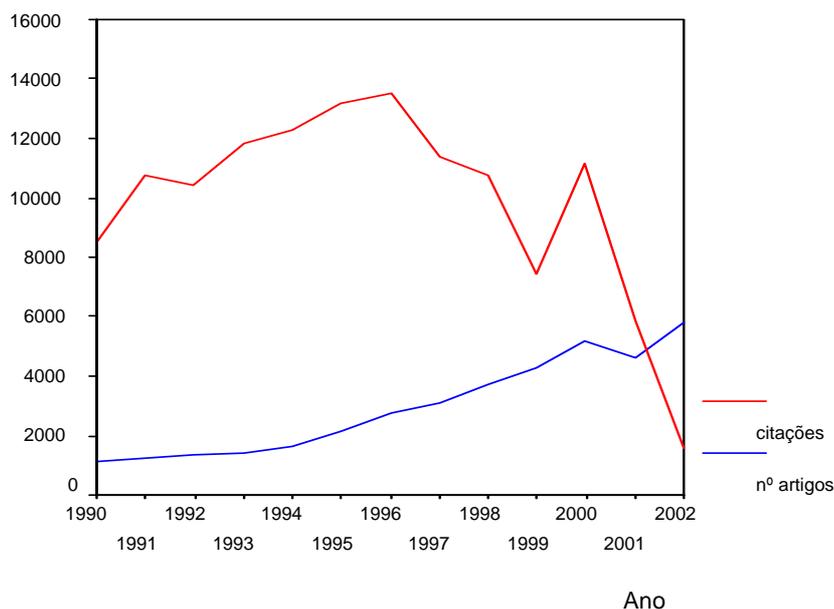
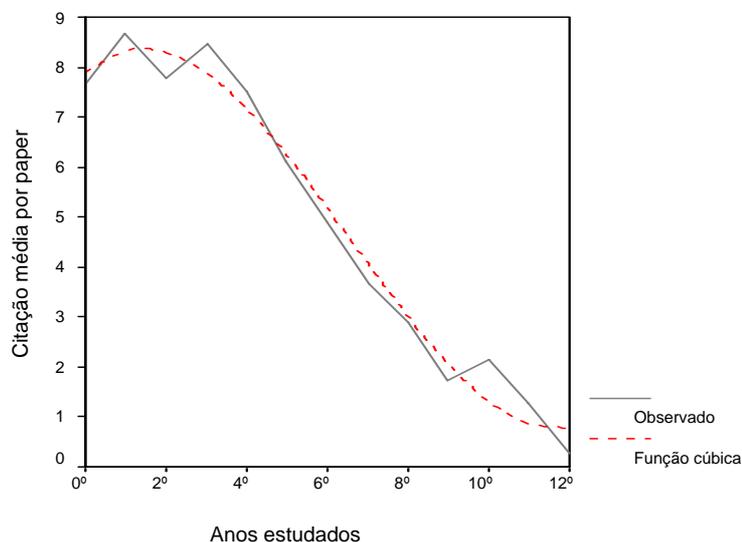


Gráfico 3 - Tendência de citações por ano, para publicações brasileiras, no campo da saúde, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.



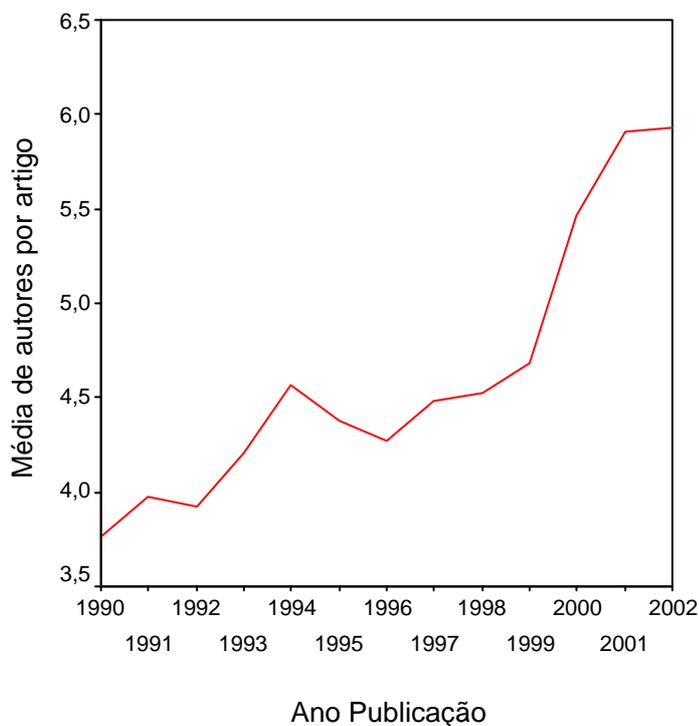
A taxa de citação, gráfico 3, para os artigos brasileiros publicados na área da saúde, mostra tendência de diminuição de 0,56 citações por artigo a cada ano, para o período estudado. Ainda assim, embora a publicação de King em 2004, na revista "Nature", refira-se a todas as áreas do conhecimento, seu artigo mostra que apenas 31 países, incluindo o Brasil, estão representados no 1% dos mais importantes artigos científicos, ou seja, aqueles que foram os mais citados por outros cientistas. Isto significa o impacto ou a visibilidade das publicações dentro de uma comunidade científica.

Enquanto o aumento do número de artigos publicados num determinado período, significa o crescimento do campo da ciência, o envelhecimento dos campos científicos, podem ser observados segundo a vida média das referências de suas publicações. (Sancho, 1990)

Podemos inferir que a queda no número de citações pode significar o esgotamento de determinada área e ainda assim, é verificado um aumento de produção ou se a qualidade da produção científica em algumas áreas não corresponde ao desejado.

6.1.3. COMPLEXIDADE: N° MÉDIO DE AUTORES POR ARTIGO

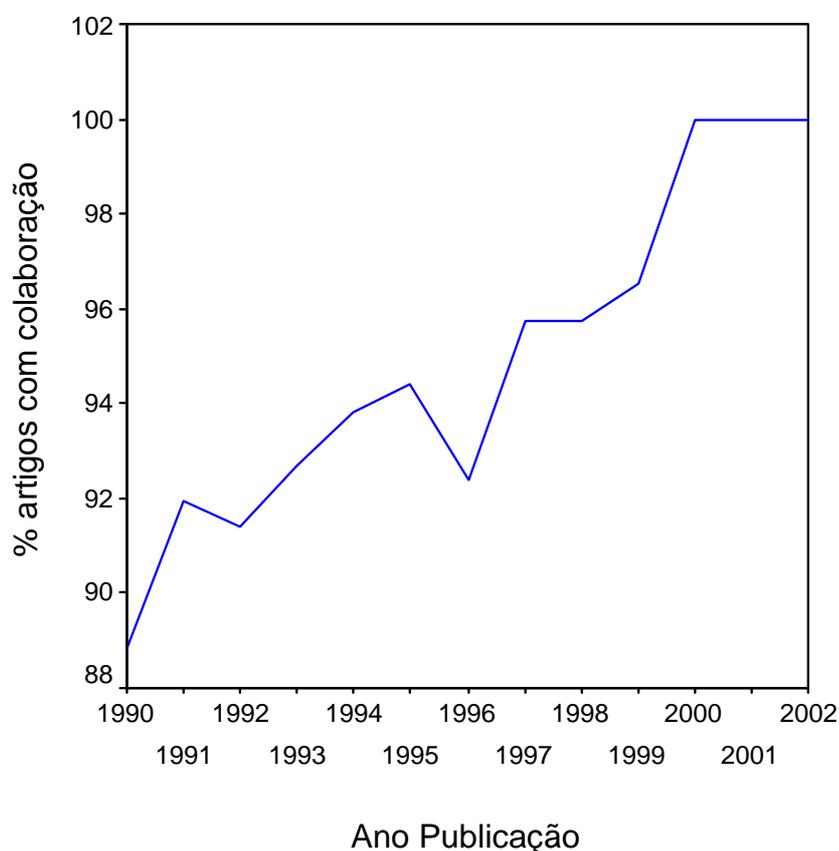
Gráfico 4 - Curva da média de autores, para publicações brasileiras, no campo da saúde, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.



Segundo resultados apontados nos gráficos 4 e 5 e nas tabelas 3 e 4, se o número de autores necessários para um artigo pode indicar a complexidade de um estudo, podemos dizer, de acordo com os dados recuperados da base de dados da ISI, que a complexidade dos trabalhos científicos tem aumentado. O número médio de autores por artigo cresceu de 3,7 autores para 5,9 aproximadamente,

no período estudado. Este número também pode indicar a busca de melhor desempenho, ou seja, a necessidade de conjugar esforços financeiros e competência técnica e científica, conseqüentemente aumentando o grau de parcerias, rumo a maior produtividade e competitividade, que por sua vez facilitam a obtenção de recursos para o financiamento de pesquisas.

Gráfico 5 - Curva do Percentual de publicações com colaboração, no campo da saúde, no Brasil, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.



A tabela 3 mostra que a colaboração entre autores concentra, até 10 autores, um percentual de 96,75% dos artigos publicados no período, sendo maior o número de artigos com 2 a 6 autores. Publicações com mais de 10 autores significam apenas 3,15% dos artigos, com 1 autor,

somente 3,4% dos artigos foram realizados sem colaboração.

Tabela 3 - Número de autores por publicação, no Brasil, no campo da saúde, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.

<i>No. de Autores</i>	<i>No. de Publicações</i>	<i>% Publicações</i>
1	1310	3,4%
2	4230	11,0%
3	6538	17,0%
4	7170	18,7%
5	6103	15,9%
6	4694	12,2%
7	3264	8,5%
8	1885	4,9%
9	1164	3,2%
10	741	1,9%
Sub total	37.099	96,75%
>10	1208	3,15%
0	42	0,10%
Total	38.349	100%

Tabela 4 - Distribuição anual das publicações brasileiras, com um autor e com colaboração, no campo da saúde, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.

Ano	Um único autor		Com colaboração		Total Publicações
	Publicações	%	Publicações	%	
1990	124	11,20%	986	88,80%	1110
1991	100	8,10%	1140	91,90%	1240
1992	115	8,60%	1219	91,40%	1334
1993	102	7,30%	1292	92,70%	1394
1994	101	6,20%	1536	93,80%	1637
1995	121	5,60%	2037	94,40%	2158
1996	210	7,60%	2557	92,40%	2767
1997	132	4,30%	2953	95,70%	3085
1998	157	4,20%	3546	95,80%	3703
1999	148	3,50%	4125	96,50%	4273
2000	0	0,00%	5190	100,00%	5190
2001	0	0,00%	4607	100,00%	4607
2002	0	0,00%	5809	100,00%	5809

A tabela 4 evidencia a distribuição das publicações com e sem colaboração, por ano. Os artigos com colaboração são responsáveis por 88,80% das publicações no ano de 1990 e num crescimento constante chega a 100% de 2000 a 2002. Vale lembrar que em 0,1% do total das publicações no período não constam autores, o que pode ocorrer quando as publicações são institucionais.

6.1.4. LEI DE LOTKA

Tabela 5 - Distribuição do número de autores brasileiros, segundo número de publicações, no campo da saúde, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.

No. de Publicações	Autores	%
1	45788	62,17
≥ 2	27856	38,83
	73644	100

Segundo Urbizagástegui (2002) para estabelecer a parte com que homens de diferentes qualidades contribuem ao progresso da ciência, Lotka em 1926 estabeleceu os fundamentos da lei do quadrado inverso, afirmando que o número de autores que fazem " n " contribuições em um determinado campo científico é proporcional a $1/n^2$ daqueles que fazem uma só contribuição e que a proporção daqueles que fazem uma única contribuição é de 60%.

Desde 1926, muitos estudos têm sido conduzidos para investigar a produtividade dos autores em distintas disciplinas. Até dezembro de 2000, mais de 200 trabalhos,

entre artigos, monografias, capítulos de livro, comunicações em congressos e literatura gris, foram produzidas tentando criticar, replicar e/ou reformular esta Lei bibliométrica, ainda segundo Urbizagástegui (2002). Apesar das numerosas pesquisas realizadas sobre este assunto, os resultados parecem contraditórios e inconclusivos. No entanto no seu trabalho para avaliar o ajuste da distribuição teórica à distribuição observada, o autor conclui que, das dez literaturas revisadas, somente quatro se ajustaram ao modelo da Lei de Lotka, entre elas a da medicina.

Na tabela 5, onde 62,17% dos 73.644 autores, presentes na base de dados, publicaram um único artigo no campo estudado, fica comprovado o princípio da Lei de Lotka.

Gráfico 6 - Distribuição das publicações no campo da saúde, segundo localidade onde são editadas, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.

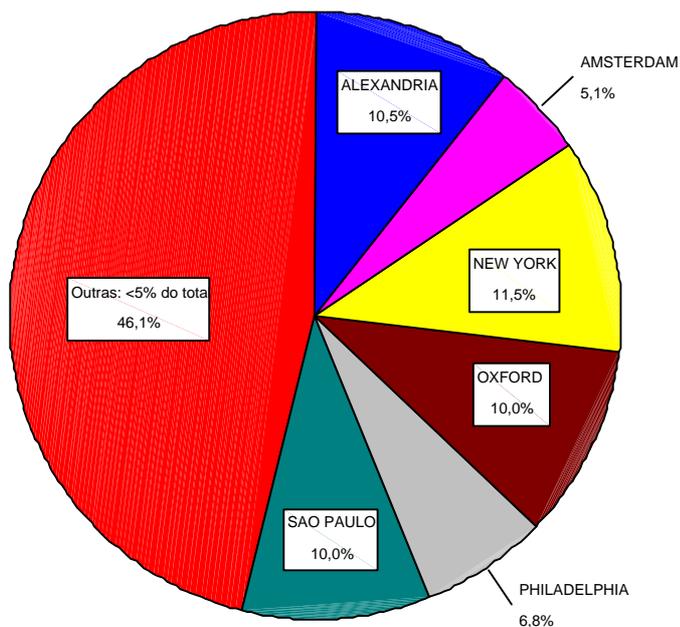


Tabela 6 - Distribuição das publicações no campo da saúde, segundo localidade onde são editadas, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.

<i>Local de Publicação</i>	<i>No. de Publicações</i>	<i>% de Publicações</i>
NEW YORK	4410	11,5
ALEXANDRIA	4027	10,5
SAO PAULO	3834	10,0
OXFORD	3835	10,0
PHILADELPHIA	2608	6,8
AMSTERDAM	1956	5,1
TOTAL	20670	53,9

As informações sobre as cidades onde se editam as revistas são oriundas do campo de informações dos publicadores. Mais de 300 cidades de vários países do mundo constam dessa listagem onde estão destacados no Brasil, além de São Paulo com 10% das publicações, gráfico 6 e tabela 6, as cidades do Rio de Janeiro, Curitiba e Ribeirão Preto que elevam para quatro o número de cidades brasileiras e para 13,41% o percentual de publicações no país. (anexo 1)

6.1.5. REVISTAS SEGUNDO A "LEI DE BRADFORD"

A "Lei de Bradford", ou Lei de Dispersão, permite, mediante a medição da produtividade das revistas, estabelecer o núcleo e as áreas de dispersão sobre um determinado assunto em um mesmo conjunto de revistas, portanto 1/3 das publicações concentram-se em 1 grupo de revistas, outro 1/3 em outro grupo com ' n ' revistas, e o último terço em um grupo com n^2 revistas,

ou seja, Bradford expressou sua lei com a seguinte fórmula: $1:n:n^2$.

Podemos observar estes resultados na tabela 7, onde no primeiro grupo, 24 revistas são responsáveis por 33,7% das publicações. Com 12.946 artigos em 209 revistas, ou seja, 33,8% das publicações temos o segundo grupo e as restantes, 1813 revistas, com 32,5% das publicações formam o terceiro grupo, que representam 12.482 artigos.

Tabela 7 - Produção científica brasileira, de acordo com a "Lei de Bradford", registrada na base de dados da "web of Knowledge" no período de 1990 a 2002.

<i>Grupos de Bradford</i>	<i>Publicações</i>	<i>% de Publicações</i>	<i>No.de Revistas</i>	<i>Média de Publicações por Revista</i>
Grupo 1	12921	33,7	24	538,4
Grupo 2	12946	33,8	209	61,94
Grupo 3	12482	32,5	1813	6,8
Total	38.349	100	2.046	

Tabela 8 - Núcleo das revistas mais produtivas, de acordo com o Grupo 1 de Bradford, relacionadas ao campo da saúde, pesquisadas na base de dados da "web of Knowledge", no período de 1990 a 2002.

<i>REVISTA</i>	<i>Artigos</i>	<i>%</i>
JOURNAL OF DENTAL RESEARCH	3953	0,1031
BRAZILIAN JOURNAL OF MEDICAL AND BIOLOGICAL RESEARCH	1721	0,0449
REVISTA DE SAUDE PUBLICA	995	0,0259
JOURNAL OF THE BRAZILIAN CHEMICAL SOCIETY	611	0,0159
REVISTA DE MICROBIOLOGIA	513	0,0134
BRAZILIAN ARCHIVES OF BIOLOGY AND TECHNOLOGY	472	0,0123
MEMORIAS DO INSTITUTO OSWALDO CRUZ	440	0,0115
BLOOD	435	0,0113
FASEB JOURNAL	426	0,0111
GENETICS AND MOLECULAR BIOLOGY	397	0,0104
INTERNATIONAL JOURNAL OF PSYCHOLOGY	301	0,0078
HEPATOLOGY	275	0,0072
JOURNAL OF BIOLOGICAL CHEMISTRY	274	0,0071
TRANSPLANTATION PROCEEDINGS	236	0,0062
AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE	218	0,0057
BIOPHYSICAL JOURNAL	215	0,0056
GASTROENTEROLOGY	210	0,0055
BRAIN RESEARCH	188	0,0049
TETRAHEDRON LETTERS	188	0,0049
AMERICAN JOURNAL OF MEDICAL GENETICS	186	0,0049
PEDIATRIC RESEARCH	180	0,0047
EUROPEAN JOURNAL OF PHARMACOLOGY	167	0,0044
TRANSFUSION	160	0,0042
WATER SCIENCE AND TECHNOLOGY	160	0,0042
TOTAL	12921	0,3371

As 24 revistas do Grupo 1 de Bradford ou "hard core", da tabela 8, representam as que contêm o maior número de publicações no campo da saúde, (12.921) ou seja, 33,7% do total das publicações analisadas e uma média de 538,7 publicações por revista. No anexo 2, encontram-se listadas as revistas do Grupo 2, "soft core" com 209 revistas e 12.946 publicações, que significam uma média de 61,94 publicações por revista e o Grupo 3, "periférico" com 12.482 publicações em 1.813 revistas com

a média de 6,8 publicações por revista. As publicações brasileiras pertencem a 2.046 títulos de periódicos, embora tenham sido pesquisados 3.066 títulos o que significa que destas, 1.020 não contém nenhuma publicação brasileira, no período estudado.

Tabela 9 - Áreas do Conhecimento de maior produtividade, segundo revistas classificadas no Grupo 1 de Bradford, para publicações brasileiras, relacionadas ao campo da saúde, de acordo com as bases de dados da "web of Knowledge", no período de 1990 a 2002.

<i>Áreas do Conhecimento</i>	<i>Publicações</i>	<i>%</i>
1. Medical Research, Organs and Systems	4438	34,30%
2. Dentistry & Oral Surgery Medicine	3953	30,60%
3. Medical Research General Topics	2559	19,80%
4. Public Health and Health Care Science	995	7,70%
5. Biology	985	7,60%
6. Chemistry	611	4,70%
7. Cardiovascular e Hematology research	595	4,60%
8. Hematology	595	4,60%
9. Molecular Biology and Genetics	583	4,50%
10. Biochemistry e Biophysics	489	3,80%
11. Gastroenterology and Hepatology	485	3,80%
12. Experimental Biology	426	3,30%
13. Physiology	301	2,30%
14. Medical Research, Diagnosis and Treatment	236	1,80%
15. Surgery	236	1,80%
16. Environmental Medicine e Public Health	218	1,70%
17. Chemistry e Analysis	188	1,50%
18. Neuroscience and Behavior	188	1,50%
19. Otolaryngology	180	1,40%
20. Pharmacology and Toxicology	167	1,30%
21. Environment Ecology	160	1,20%

O prestígio das revistas onde se publicam os resultados das investigações científicas representa uma

medida da influência que podem exercer os artigos nelas publicados.

Como cada revista pode corresponder de 1 a 3 áreas de conhecimento, pelos grupos de Bradford podemos conhecer as áreas de maior produtividade em ciências da saúde.

De acordo com as tabelas 9 e 10, observa-se a relação entre diferentes áreas de especialização. Temos que as áreas de Pesquisa Médica em Órgãos e Sistemas, Odontologia e Cirurgia Oral e Pesquisa Médica Geral somam o maior número de publicações da área. Juntos são responsáveis por 10.950 artigos, isto é, 28,55% de toda a produção do campo, no período estudado.

Na tabela 9 estão relacionados os grupos de áreas com maior produtividade, ou seja, grupo "hard core". Na tabela 10 temos a distribuição da produção científica total, no período estudado, segundo a área do conhecimento ao qual está relacionada. No anexo 5, estão relacionados os grupos de áreas do conhecimento, que de acordo com a produtividade das revistas do campo da saúde, formam o que denominamos "soft core" e "periféricos".

Tabela 10 - Produtividade por Áreas do Conhecimento, relacionados ao campo da saúde, publicadas em revistas classificadas nos Grupos 1, 2 e 3 de Bradford, pesquisadas na base de dados da "web of Knowledge", no Brasil, no período de 1990 a 2002.

Área do Conhecimento	Artigos	%
1. Medical Research, Organs and Systems	6117	16,00%
2. Dentistry Oral surgery Medicine	4674	12,20%
3. Medical Research General Topics	3508	9,10%
4. Biochemistry and Biophysics	2203	5,70%
5. Neuroscience and Behavior	2203	5,70%
6. Cardiovascular and Hematology research	1901	5,00%
7. Biology	1862	4,90%
8. Environment Ecology	1707	4,50%
9. Pharmacology and Toxicology	1655	4,30%
10. Microbiology	1600	4,20%
11. Molecular Biology and Genetics	1518	4,00%
12. Cardiovascular and Respiratory System	1517	4,00%
13. Immunology	1266	3,30%
14. Medical Research, Diagnosis and Treatment	1255	3,30%
15. Public Health and Health Care Science	1180	3,10%
16. Chemistry	1090	2,80%
17. Chemistry and Analysis	1020	2,70%
18. Gastroenterology and Hepatology	861	2,20%
19. Entomology Pest Control	803	2,10%
20. Hematology	762	2,00%
21. Endocrinology, Nutrition and Metabolism	690	1,80%
22. Food Science Nutrition	686	1,80%
23. Biotechnology and Applied Microbiology	682	1,80%
24. Surgery	654	1,70%
25. Experimental Biology	647	1,70%
26. Physiology	592	1,50%
27. Cell and Developmental Biology	581	1,50%
28. Environmental Medicine and Public Health	579	1,50%
29. Research Laboratory Medicine and Medical Technology	505	1,30%
30. Oncogenesis and Cancer Research	467	1,20%
31. Endocrinology, Metabolism and Nutrition	423	1,10%
32. Urology and Nephrology	421	1,10%
33. Psychology	412	1,10%
34. Clinical Immunology and Infectious Disease	404	1,10%
35. Otolaryngology	360	0,90%
36. Neurology	343	0,90%
37. General and Internal Medicine	342	0,90%
38. Pediatrics	280	0,70%
39. Radiology, Nuclear Medicine and Imaging	264	0,70%
40. Anaesthesia and Intensive Care	254	0,70%
41. Oncology	246	0,60%
42. Psychiatry	223	0,60%
43. Reproductive Medicine	220	0,60%

(continua...)

Tabela 10 - Produtividade por Áreas do Conhecimento, relacionados ao campo da saúde, publicadas em revistas classificadas nos Grupos 1, 2 e 3 de Bradford, pesquisadas na base de dados da "web of Knowledge", no Brasil, no período de 1990 a 2002. (continuação)

Áreas do Conhecimento	Artigos	%
44.Ophthalmology	195	0,50%
45.Rheumatology	173	0,50%
46.Dermatology	152	0,40%
47.Orthopedics, Rehabilitation and Sport Medicine	82	0,20%
48.Clinical Psychology and Psychiatry	62	0,20%
49.Rehabilitation	27	0,10%
50.Health Care Science and Services	10	0,00%
Total de Publicações	38349	100,00%

6.1.6. TIPO DE DOCUMENTOS

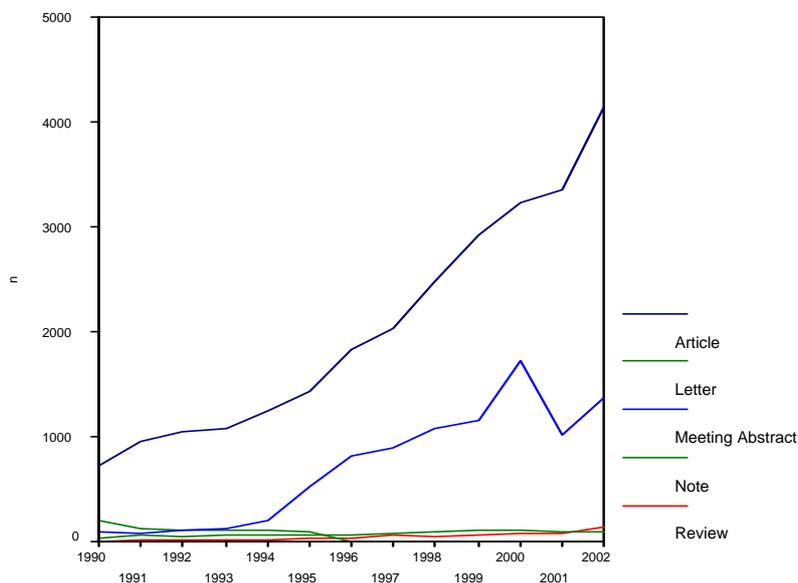
Os resultados da maioria das investigações em ciência e tecnologia são transmitidos por meio de um processo de comunicação escrita, em artigos de revista, livros, revisões, anais de congressos, teses, patentes e cartas entre outros. Portanto, os trabalhos publicados compõem um dos produtos finais de toda a atividade científica e tecnológica e representam um indicador do volume da pesquisa produzida.

Tabela 11 - Produção científica por Tipo de Documento, no campo da saúde, no Brasil, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.

<i>Tipo de Documento</i>	<i>Número de Publicações</i>	<i>%</i>
Article	26460	69,0
Meeting Abstract	9168	23,9
Letter	930	2,4
Note	725	1,9
Review	581	1,5
Editorial Material	327	0,9
Book Review	57	0,1
Correction	39	0,1
Correction, Addition	13	0
Discussion	13	0
Biographical-Item	12	0
Item About an Individual	8	0
Reprint	3	0
News Item	2	0
Software Review	1	0
Sub-Total	38339	100
**MISSING	10	
TOTAL GERAL	38349	

O crescimento observado para as publicações, de acordo com o tipo de documento, na tabela 11 e gráfico 7, é tanto para artigos quanto para participações em congressos.

Gráfico 7 - Distribuição das publicações brasileiras, segundo tipo de documento, no campo da saúde, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.



Tradicionalmente os artigos são o principal veículo de difusão da ciência e tem maior expressão numérica. Representaram 69% de toda a produção estudada, seguidos de participações em congresso com 23,9% no período, conforme mostram a tabela 11 e o gráfico 7.

A tabela 12 apresenta o número médio de autores, de acordo com os tipos documentos científicos publicados. Os artigos e as participações em congresso apresentam uma média de 5 autores por publicação.

Tabela 12 - Publicações científicas brasileiras, classificadas por Tipo de Documento, segundo número médio de autores, no campo da saúde, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.

<i>Tipo de Documento</i>	<i>Número de Autores Média</i>
Article	5
Biographical-Item	3
Book Review	1
Correction	6
Correction, Addition	4
Discussion	9
Editorial Material	3
Item About an Individual	2
Letter	4
Meeting Abstract	5
News Item	11
Note	4
Reprint	12
Review	4
Software Review	1

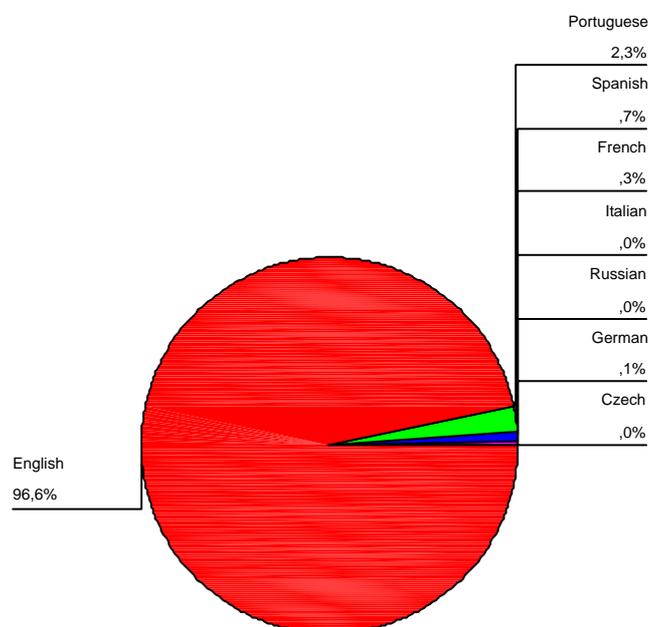
6.1.7. IDIOMA E LOCAL DE PUBLICAÇÃO

A tabela 13 e o gráfico 8 indicam o idioma que as publicações foram divulgadas e sua freqüência.

Tabela 13 - Número de artigos segundo o idioma de publicação, na área da saúde, no Brasil, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.

<i>Idioma</i>	<i>Frequência</i>	<i>%</i>
Inglês	37037	96,6
Português	900	2,3
Espanhol	252	0,7
Francesa	117	0,3
Alemã	28	0,1
Russo	3	0
Italiano	2	0
Checo	1	0
Sub Total	38340	100
Sem Informação	9	
Total	38349	100

Gráfico 8 - Percentual de publicações brasileiras, segundo o idioma de publicação, no campo da saúde, registrados na base de dados da "web of Knowledge", no período de 1990 a 2002.



O perfil das publicações brasileiras quanto ao idioma preferido para publicação evidencia a língua inglesa, que fica a frente, com a grande maioria das publicações, 96,6%. Já as publicações em português aparecem com apenas 2,3%. Publicações em língua espanhola representam 0,7% do total de 38.340 publicações. Este é o resultado esperado, pois os autores têm preferência pelo idioma inglês para divulgar seus trabalhos em revistas estrangeiras.

A produção científica brasileira, no campo da saúde, segundo a filiação da editora, soma 33.191, ou seja, 87% do total, no período, para as editoras estrangeiras, uma vez que as publicações em revistas internacionais são a maioria.

6.1.8. PADRÕES DA PRODUÇÃO CIENTÍFICA BRASILEIRA.

Para conhecer os padrões da produção científica brasileira, foram realizadas análises de "cluster" das áreas do conhecimento, usando os como indicadores, o número total das publicações, a tendência de crescimento da produção científica, o número médio de autores por artigo, a proporção de artigos publicados em revistas no exterior, o número médio de referências de cada publicação e o número médio de citações. A análise de citações que se fazem a artigos e revistas, proporciona uma visão fundamental, quantitativa e objetiva do desenvolvimento da investigação científica e sua tendência, em um determinado campo do conhecimento.

O objetivo primário da análise de "cluster" é buscar entender a estrutura dos dados, aglomerando observações similares em grupos. A sua aplicação é, portanto, bastante imediata e pode conferir um bom poder explanatório à pesquisa, supondo que decisões ponderadas sejam tomadas sobre formas alternativas para mensuração de similaridades, critérios para formação de grupos e número ideal de grupos a serem formados.

Estudos similares realizados por Pereira (1996 e 1999), demonstram a utilização de métodos de análise multivariada, como a análise de agrupamento, "*cluster analysis*", em avaliação de Ciência e Tecnologia, pois esta é uma área que lida com evento multidimensional e exige uma abordagem multivariada.

O fator de impacto dos artigos e revistas, idealizado por Garfield, baseia-se na frequência média de citações num determinado período.

É evidente que o fator de impacto indica a categoria de um artigo ou de uma revista. Daí a importância para os autores de publicarem em revistas com um elevado fator de impacto, já que conseguiram obter grande visibilidade.

Na atualidade, discute-se o valor das citações e existe um consenso sobre as citações nem sempre medirem a qualidade da pesquisa científica, mas o impacto das publicações.

Para visualizar os resultados, as tabelas 14, 15, 16 e 17, evidenciam a composição dos "clusters" em conjunto e separadamente. A tabela 14, mostra o resultado global das análises, e apresenta como forma de comparação os 4 "clusters".

Tabela 14 - Composição dos "clusters" de acordo com os padrões da produção científica brasileira, no campo da saúde, segundo registros contidos na base de dados da "web of Knowledge", no período de 1990 a 2002.

<i>Cluster</i>	<i>Média de publicações</i>	<i>Tendência de crescimento</i>	<i>Nº médio de autores</i>	<i>Nº médio de citações</i>	<i>Proporção artigos publicados no exterior</i>	<i>Nº médio de referências</i>
1	406,12	0,24	5,23	3,56	1	10,86
2	1569,79	0,05	5,1	4,06	0,84	12,41
3	4091	0,01	4,55	2,22	0,69	9,26
4	6117	0,07	4,88	0,93	1	2,62
Média	993,56	0,17	5,16	3,6	0,94	11,07

Da análise dos indicadores bibliométricos estudados para a definição do comportamento da produção

científica, em "clusters" ou agrupamentos, como o número total das publicações, a tendência de crescimento da produção científica, o número médio de autores, o número médio de citações, a proporção de artigos publicados no exterior e o número médio de referências de cada publicação, resultou a caracterização de 4 "clusters", distinguidos por diferenças significantes em número de publicações, taxa de crescimento da produção científica, citações e proporção de publicações no exterior.

Desta forma, podemos caracterizar o campo da ciência da saúde em dois padrões distintos definidos pelos "clusters" 1 e 2, tabelas 15 e 16, que contemplam 33 e 14 áreas do conhecimento respectivamente. E, em 3 áreas de comportamentos característicos, de acordo com os "clusters" 3 e 4, tabela 17. O "cluster" 3, com apenas 2 áreas do conhecimento, fica melhor analisado como sendo áreas afastadas, com alguma relação entre si que permitiu seu agrupamento e o "cluster" 4 é representado por uma única área do conhecimento. Assim entendemos os "clusters" 3 e 4 como três áreas de comportamentos característicos.

O **"cluster" 1**, representado na tabela 15, apresenta as áreas do conhecimento menos produtivos, com um total de 13.402 artigos publicados durante os treze anos estudados, distribuídos em 33 áreas, o que significa uma média de 406,12 publicações por área do conhecimento. No entanto, possui a maior tendência de crescimento, 0,24 por ano, de publicações integralmente no exterior. Como característica adicional, tende a ter maior complexidade,

se tomarmos em conta o alto número de referências, uma média de 10,86 referências por artigo e maior número médio de autores por publicação. A alta taxa de citação desse "cluster", a média de 3,56 citações por artigo, pode significar a existência de áreas emergentes, bem como um impacto ou visibilidade de suas publicações para a comunidade científica internacional.

Tabela 15 - Relação do "cluster" 1 das publicações brasileiras, com as áreas do conhecimento, no campo da saúde, registradas na base de dados da "web of Knowledge", no período de 1990 a 2002.

"cluster" 1					Proporção	
Área do Conhecimento	Nº Médio	Tendência de crescimento	Nº médio de autores	Nº médio de citações	artigos publicados no exterior	Nº médio de referências
Anaesthesia and Intensive Care	254	0,16	5,02	3,09	1	6
Biotechnology and Applied Microbiology	682	0,13	4,3	3,52	1	15,5
Cell e Developmental Biology	581	0,11	4,51	3,92	1	15,9
Clinical Immunology and Infectious Disease	404	0,1	7,18	3,99	1	7,47
Clinical Psychology and Psychiatry	62	0,78	4,68	2,27	1	13,84
Dermatology	152	0,32	5,22	1,64	1	9,11
Endocrinology, Metabolism and Nutrition	423	0,16	6,11	6,53	1	11,71
Endocrinology, Nutrition and Metabolism	690	0,11	5,47	3,81	1	13,34
Entomology Pest Control	803	0,08	3,76	2,3	1	12,22
Environmental Medicine and Public Health	579	0,14	5,55	6,39	1	17,79
Experimental Biology	647	0,09	4,41	2,22	1	8,84
Food Science Nutrition	686	0,11	4,26	3,06	1	12,91
Gastroenterology and Hepatology	861	0,08	6,51	1,3	1	3,77
General and Internal Medicine	342	0,2	5,27	5,9	1	9,14
Health Care Science and Services	10	1,2	4,5	1,9	1	10,4
Hematology	762	0,07	6,75	1,48	1	3,93
Neurology	343	0,2	5,41	5,36	1	13,39
Oncogenesis and Cancer Research	467	0,17	6,85	7,26	1	15,29
Oncology	246	0,25	6,89	9,93	1	13,2
Ophthalmology	195	0,33	5,42	4,07	1	9
Orthopedics, Rehabilitation and Sport Medicine	82	0,89	3,96	2,67	1	9,3
Otolaryngology	360	0,11	5,41	1,17	0,99	5,84
Pediatrics	280	0,18	5,2	2,99	0,99	14,24
Psychology	412	0,23	4,62	3,85	1	18,62
Psychiatry	223	0,03	4,81	3,47	1	14,62
Physiology	592	0,23	3,1	1,63	1	6,31
Radiology, Nuclear Medicine and Imaging	264	0,22	5,55	3,14	1	7,19
Rehabilitation	27	0,39	3,78	1,07	1	10,07
Reproductive medicine	220	0,22	6,44	1,86	1	5,85
Research Laboratory Medicine and Medical Technology	505	0,09	5,48	4,99	1	12,36
Rheumatology	173	0,32	5,11	4,13	1	12,9
Surgery	654	0,08	5,57	2,56	1	8,29
Urology and Nephrology	421	0,1	5,39	4,1	1	10,12
No. Médio	406,120,24		5,23	3,56	1	10,86

Tabela 16 - Relação do "cluster" 2 das publicações brasileiras, com as áreas do conhecimento, no campo da saúde, registradas na base de dados da "web of Knowledge", no período de 1990 a 2002.

"Cluster" 2						
Área do Conhecimento	Nº	Tendência	Nº	Nº	Proporção	Nº
	Médio	de	médio	médio	artigos	médio
	de	de	de	publicados	no exterior	de
	referências	referências	referências	referências	referências	referências
Biology	1862	0,04	3,94	2,02	0,47	8,25
Cardiovascular and Hematology research	1901	0,03	6,61	2,82	1	6,43
Cardiovascular and Respiratory System	1517	0,04	6,42	3,5	1	7,92
Chemistry	1090	0,06	4,16	2,5	0,44	12,49
Chemistry and Analysis	1020	0,08	4,44	4,96	1	12,65
Environment Ecology	1707	0,05	3,92	3,69	1	13,64
Immunology	1266	0,05	6,44	6,4	1	14,33
Medical Research, Diagnosis and Treatment	1255	0,03	5,41	3,32	1	10,88
Microbiology	1600	0,05	5,46	4,55	1	14,55
Molecular Biology and Genetics	1518	0,05	5,21	5,57	0,74	12,39
Neuroscience and Behavior	2203	0,03	5,34	5,79	1	16,15
Pharmacology and Toxicology	1655	0,05	5,12	4,56	1	14,66
Public Health and Health Care Science	1180	0,12	3,69	1,06	0,16	14,06
Número Médio	1569,8	0,05	5,1	4,06	0,84	12,41

O "cluster" 2, representado na tabela 16, possui uma alta produção, com uma média de 1596,76 publicações distribuídas pelas 14 áreas do conhecimento, o que representa um total de 21.977 artigos publicados no período. Têm menor tendência de crescimento, 0,05, e menor expressão, em relação ao "cluster" 1, no que diz respeito às publicações no exterior, uma proporção de 0,84. Tende a uma maior complexidade com o maior número de referências por artigo e um alto número médio de 5,10 autores, por trabalho publicado.

Este é o "cluster" mais citado, tem o maior número médio de 4,06 citações por artigo publicado e,

portanto pode sugerir áreas emergentes, maior impacto e visibilidade das suas publicações, uma vez que atualmente é consenso que número de citações não se traduz, obrigatoriamente, em qualidade da investigação científica. Os "clusters" 1 e 2 são áreas estáveis que talvez melhor representem as ciências da saúde.

Tabela 17 - Relação dos "cluster" 3 e 4, das publicações brasileiras, com as áreas do conhecimento, no campo da saúde, registradas na base de dados da "web of Knowledge", no período de 1990 a 2002.

Áreas conhecimento	Nº Médio	Tendência de crescimento	Nº médio de autores	Nº médio de citações	Proporção artigos publicados no exterior	Nº médio de referências
"cluster" 3 Dentistry Oral surgery medicine	4674	0,01	4,3	0,3	1	1,52
Medical Research General Topics	3508	0,01	4,8	4,15	0,38	17,01
TOTAL	4091	0,01	4,55	2,22	0,69	9,26
"cluster" 4 Medical Research, Organs and Systems	6117	0,07	4,88	0,93	1	2,62
Número Médio	6117	0,07	4,88	0,93	1	2,62

Os "clusters" 3 e 4 estão representados na tabela 17. As áreas que compõem os "clusters" 3 e 4 estão muito afastadas das demais devido sua alta produtividade. Apresentam uma produção média por área de 4091 e 6117 artigos, respectivamente, ao longo do período estudado. Diferenciam-se um do outro porque o "cluster" 3 recebe maior número de citações, uma média de 2,22 citações por publicação. No entanto, se isolarmos a área "Dentistry Oral Surgery Medicine", chama a atenção no "cluster" 3 pois é a área que menos citações recebe, uma média de 0,3 por artigo, dentre as 50 áreas estudadas no campo do conhecimento da saúde.

Enquanto o "cluster" 4 recebe o menor número de citações de todos os grupos, com uma média de apenas 0,93 citações por artigo, apresenta maior taxa de crescimento, ainda que modesta, de 0,07 ao ano, em relação ao "cluster" 3. Como são áreas de grande produção não se espera que tenham elevada tendência de crescimento. Estas, parecem ser áreas, do campo das ciências da saúde, já consolidadas.

6.2. "CO-WORD ANALYSIS"

O grupo de análises que serão apresentadas nesta segunda parte trata de estudar a co-ocorrência de palavras. Esta consiste na detecção das palavras que caracterizam o conteúdo dos trabalhos sobre um determinado tema, na contagem da aparição destas e posterior estudo de sua proximidade e significância. Os conceitos de proximidade ou distância podem ser representados graficamente e constituem a base para a elaboração dos "mapas tecnológicos".

Os mapas tecnológicos representam graficamente estas *proximidades ou distâncias*, dando uma visão da estrutura de relações que existe em um âmbito dado. A distância no mapa entre duas palavras, referentes a tecnologias, produtos, autores, empresas, assinala a maior ou menor relação entre elas. Esta proximidade pode ser quantificada mediante diversos índices e métricas e, graças a isto, visualizá-las graficamente por meio dos mapas. "O fundamento metodológico da "co-word analysis" é a idéia que a co-ocorrência de palavras-chave descreve o conteúdo de um documento em um arquivo. Do ponto de vista

metodológico, é portanto uma questão de usar um ou mais índices para medir a intensidade relativa dessas ocorrências e efetuar uma representação simplificada das redes as quais elas podem evidenciar". (Callon, 1991)

As palavras que foram estudadas estão contidas no campo descritores da base de dados e foram relacionadas com outras variáveis, da mesma base, como autores, identificadores, ano de publicação do artigo ou filiação da publicação.

Os descritores foram escolhidos porque representam o assunto do artigo, são chamados de descritores de assunto, mas estes termos também são conhecidos como palavras-chave. Com base no artigo ou documento propriamente dito, os profissionais que realizam indexações fazem uma análise temática e identificam os descritores que representam os assuntos enfocados no artigo.

Já os identificadores, que também representam os assuntos enfocados no artigo, são designados pelo próprio autor, de acordo com sua interpretação do que melhor identificará seu trabalho. Um identificador pode, ou não, pertencer ao dicionário ou ao vocabulário controlado de descritores de um determinado campo do conhecimento, pois, muitas vezes, são palavras novas ou que melhor representam novas descobertas.

O fato de escolher trabalhar com descritores e identificadores está fundamentado no princípio, segundo Whitaker, de que *"os autores dos artigos científicos escolhem com cuidado os termos técnicos que utilizam;*

quando diversos termos são utilizados no mesmo artigo, isso acontece, de fato, porque o autor reconhece ou supõe que existe algum tipo de relação não trivial entre seus referentes; se um número significativo de autores reconhece o mesmo tipo de relacionamento entre os termos, pode-se admitir que esse relacionamento possui algum significado dentro da área da ciência considerada; que as palavras-chave escolhidas, por indexadores competentes, como descritores do conteúdo dos artigos são de fato uma indicação confiável dos conceitos científicos a que se referem, o que torna possível o uso dessas palavras como o elemento base para a análise das co-ocorrências das palavras". (Robredo e Cunha, 1998)

Se os pressupostos acima fazem sentido, nada impede utilizar as frequências com que ocorrem os possíveis pares de palavras relacionadas, em cada artigo integrante de um conjunto de artigos, como um meio para descrever a estrutura dos conceitos contidos nos artigos.

Na impossibilidade de apresentar a totalidade dos agrupamentos de palavras e todas suas inter-relações, as análises de co-ocorrências de palavras foram centradas em alguns exemplos significativos, de forma a evidenciar as possibilidades e potencialidades desse método para "mapear" diferentes áreas do conhecimento, sem, necessariamente, caracterizar o estado da arte da produção científica brasileira, na área da saúde, no período estudado.

6.2.1 ANÁLISE DOS DESCRITORES

Os descritores encontrados somaram 39.455 palavras diferentes, no total de 38.349 registros em todo o banco de dados. Para algumas análises dos descritores, como a realização de matrizes de co-ocorrência simples, foram selecionadas 395 palavras, que significam 1% do total de descritores, em ordem decrescente de frequência de ocorrência. Um total de descritores e seu respectivo número de aparições, até 15 ocorrências, consta de uma lista e está apresentado na seção de anexos. (Anexo 3)

Polanco (1993) e (Basevi 1993), em suas análises de co-ocorrência de palavras, aplicaram critérios que consiste em eliminar termos e expressões com frequência (f) muito baixa, ($f < 2$), e que são muito numerosos, e os de frequência muito elevada quando são muito genéricos ou obviamente redundantes com o tema escolhido. Para o presente estudo foram eliminados alguns termos do grupo de palavras escolhido para análise, como "infection", "formation", "transmission", "brazil", pois estes são demasiado genéricos e, na grande maioria das vezes, são comuns a diferentes agrupamentos de palavras, sem nenhuma relação entre si.

A tabela 18 mostra o percentual dos descritores segundo sua frequência, onde observamos que 72,03% deles aparecem uma única vez como descritor de assunto de um artigo. Apenas 27,97% dos 39.455 descritores, têm frequência maior que 1. No entanto, se tomarmos os descritores que se repetem até no mínimo 10 vezes, no

período estudado, somente 1142 (2,89%) ocorrem numa frequência igual ou maior que 10.

Tabela 18 - Percentual dos Descritores segundo sua frequência de ocorrência, nas publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

<i>Frequência</i>	<i>Descritores</i>	<i>%</i>
1	28.418	72,03
>1	11.037	27,97
Total	39.455	100

Tabela 19 - Descritores e Identificadores, segundo as vinte primeiras maiores frequências de ocorrência, nas publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

<i>Descritores</i>	<i>Frequência</i>	<i>Identificadores</i>	<i>Frequência</i>
BRAZIL	743	PROTEIN	835
EPIDEMIOLOGY	407	RAT	765
RAT	401	EXPRESSION	741
INSULIN/DIABETES	359	INFECTION	631
TRYPANOSOMA	327	CELLS	582
HIV/AIDS	294	MICE	570
NITRIC OXIDE	275	DISEASE	562
AMAZONIA	249	IDENTIFICATION	547
SCHISTOSOMA	217	RECEPTOR	468
LEISHMANIA	192	MUTATIONS	438
CALCIUM	189	MECHANISM	436
ANTIBODY	188	CHILDREN	383
INFLAMMATION/INFLAMM	178	BINDING	378
GLUTAMINE/GLUTAMATE	167	DNA	374
HYPERTENSION	160	INHIBITION	361
MORTALITY	156	BRAZIL	361
DNA	155	BRAIN	357
CHAGAS	155	GROWTH	342
PCR	146	POPULATION	336
STRESS	134	CANCER	323

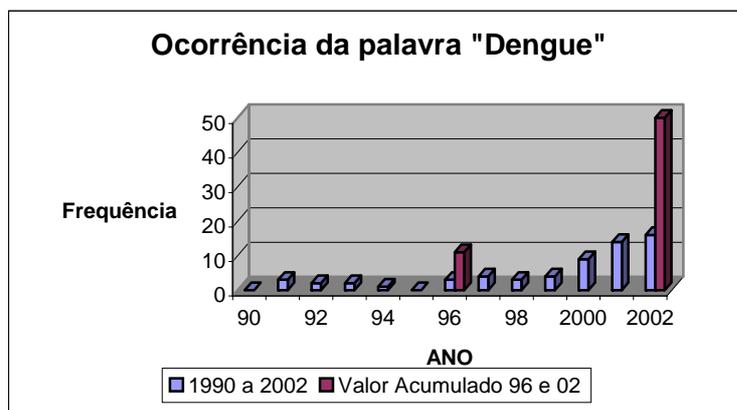
A tabela 19 é um exemplo do resultado de uma contagem simples de palavras. Esta, evidencia apenas os 20 descritores e identificadores mais freqüentes. A freqüência significa o número de ocorrência das palavras nos campos de descritor e identificador. Apenas com a contagem de palavras podemos obter informações interessantes, sobretudo se compararmos sua ocorrência no tempo para observarmos seu comportamento. Palavras referentes a tecnologias ou temas específicos que têm sua freqüência aumentada num período determinado, demonstram que estão captando a atenção, que se está patenteando sobre o assunto, enfim, que há um interesse de investigação sobre elas. Ao contrário, as que diminuem de freqüência ao longo do tempo, podem representar temas que estão chegando ao limite de seu interesse.

A palavra dengue, por exemplo, com 61 ocorrências, entre descritores e identificadores, conduz no tempo às ações sanitárias para erradicação do mosquito *Aedes*. Segundo Donalísio (1999), "*...há noventa anos vem travando-se, nas Américas, uma luta contra o mosquito transmissor do dengue, da febre amarela e de outras viroses. Esforços de erradicação foram freqüentes, com grande mobilização de recursos, em extensas campanhas nacionais, em várias regiões do continente.*" Nas campanhas de erradicação do *Aedes aegypti*, o Brasil foi um dos melhores exemplos, tiveram êxito no país desde as primeiras décadas do século XX, em epidemias nos Estados de São Paulo, Rio de Janeiro e nas cidades de Salvador, Recife, entre outras. Em 1933 várias cidades costeiras do país estiveram novamente livres do *Aedes aegypti*, a custa de insistentes programas de eliminação de criadouros de larvas. No Brasil, o vetor foi considerado erradicado de 1955 a 1967 e de 1972 até 1976. Embora o dengue não se

tenha constituído, de fato, em problema de saúde pública brasileira até a década de 80, o controle do vetor da febre amarela influenciou diretamente as possibilidades de ocorrência de epidemias. (Donalísio, 1999)

A re-introdução do *Aedes* no Brasil, na década de 80, elevou o risco de infecção pelas moléstias veiculadas por este vetor. Com isso, segundo dados da Fundação Nacional da Saúde do Ministério da Saúde, o dengue volta à região norte em 1982, em 1986-1987 na região nordeste e sudeste, envolvendo os estados do Ceará, Alagoas, Pernambuco e Rio de Janeiro, quando apresentaram mais de 130.000 casos, nos anos de 1990 e 1991, outros 140.000 casos e a partir de 1995 em todas as regiões do país.

Gráfico 9 - Ocorrência da palavra "Dengue", nas publicações brasileiras, registradas na base de dados da "web of science", no campo da saúde, no período de 1990 a 2002.



O gráfico 9 mostra a evolução dessa palavra, presente no campo de descritores e identificadores, ao longo do período estudado e em dois períodos delimitados. No primeiro momento, de 1990 a 1996, temos uma baixa

ocorrência da palavra "dengue", 18,03%. Já de 1997 a 2002 um visível aumento pode ser observado com 81,97% num total de 61 ocorrências, em todo o período estudado.

Estes dados coincidem com a história da evolução do dengue no país, na década de 90, quando os pesquisadores, motivados pelo recrudescimento dessa moléstia, impulsionaram as investigações e buscaram investimentos, para a produção de novos conhecimentos científicos e tecnológicos de modo a controlar a propagação da epidemia e eliminar a doença.

6.2.2. ANÁLISE DO CRUZAMENTO ENTRE DESCRITORES

As análises de co-ocorrência são as bases para a elaboração dos mapas tecnológicos e de possíveis definições de "*clusters*" ou associações. (Cruz, Escorsa e Maspons, 2003) Estas podem ser simples ou múltiplas, simétricas ou assimétricas. As simples permitem o cruzamento de duas informações de natureza similar como, autor-autor, autor-ano, autor-país. As co-ocorrências múltiplas requerem um terceiro campo de análises como autor-descritor-ano. As simétricas analisam dados de um mesmo campo, como autor-autor ou descritor-descritor. Já as assimétricas são as que se prestam a analisar campos diferentes como autor-descritor, descritor-identificador, por exemplo.

A análise, do cruzamento entre descritores, resultou primeiramente em uma matriz, com resultados das

contagens que permitiram organizar, depurar e transformar os dados através de operações matemáticas. A ACP - Análise dos Componentes Principais é um método que se utiliza para dados quantitativos. É indicado quando se deseja trabalhar com valores absolutos das variáveis e permite destacar os valores de maior magnitude. Quando os dados originais das matrizes apresentam-se dispersos é necessário ordená-las por blocos absolutos, o que permite uma maior proximidade dos dados da tabela. Partimos, portanto, de uma matriz de co-ocorrência, onde nas linhas e colunas foram analisados os descritores. As linhas representam os indivíduos e as colunas as variáveis. Cada linha nos dá o valor de todas as variáveis para o indivíduo desta linha.

Os fragmentos da matriz simples e simétrica de co-ocorrência de descritores evidenciados a seguir, representam os "clusters" que aparecem melhor delimitados no "zoom" da matriz de ocorrência. A figura 1, por exemplo, mostra a palavra de maior freqüência dentre os descritores, que é "epidemiology" e seu respectivo "cluster". Originalmente os resultados podem vir acompanhados de zoom e da tabela de elementos que compõe o "cluster".

Figura 1 - Fragmento da matriz de co-ocorrência de descritores, segundo 1% das palavras de maior frequência, contidas no campo de descritores, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

	EPIDEMI	RISK_FA	PREVALE	SOCIOEC	MORTALI	CAUSES	HIV/AID	MYCOBAC	ADOLESC	DENTAL	NUTRITI	OBSITY	INSULIN	HYPERTE
epidemi	350	43	37	23	15	1	15	11	6	18	5	11	4	9
risk_fa	43	112	17	12	13	2	6	2	5	2	3	5	3	3
prevale	37	17	89	6	1	1	3	4	4	2	4	4	3	2
socioec	23	12	6	58	3	2	3	1	4	3	7	1		
mortali	15	13	1	3	124	30	2	1	4		2		5	1
causes	1	2		2	30	36	1		3		1		3	
hiv/aid	15	6	3	3	2	1	220	18	4		1			
mycobac	11	2		1	1		18	109						
adolesc	6	5	4	4	4	3	4		57	3	5	2		1
dental	18	2	4	3					3	91			1	
nutriti	5	3	2	7	2	1	1		5		124	9	1	
obesity	11	5	4	1					2		9	63	13	5
insulin	4	3	3		5	3				1	1	13	280	17
hyperte	9	3	2		1				1			5	17	160
blood_p	2	1	2						2		1	1	9	18
pregnan	5	3	3	3	4	2	2		2		3	2	3	1
nitric								1				1	5	16
rat											3	1	16	4
child	6	4	4	9										
smoking	11	5	4	2	1									
alcohol	9	2	3											
diet	5	2		1										
childre	6	1	3	1	2									
morbidi	8	2	5	2	9									
prevent				1	3									
schisto														
calcium														
angiot														
bradyki														
inflamm					1									
cytokin														
neutrop														
glutami														
lymphoc														
macroph1														
leishma														
trypano			1			1								
chagas	3				1									
antibod	2		1											

O software "Tétralogie", utilizado para a realização destas análises, dispõe de uma função "zoom" que nos dá uma síntese gráfica de uma matriz de ocorrência. Nesta síntese se apresentam as magnitudes das ocorrências como pontos. Os "clusters" se concentram ao longo de uma diagonal e cada acumulação de pontos nesta diagonal representa um "cluster". Na informação gráfica do "zoom", as palavras com maiores frequências aparecem, no início da diagonal. Pode se dar diferentes formas de ordenação à informação para conseguir melhores resultados nas análises de co-ocorrências. Entre estas formas está a ordenação por conexão ou ligação. A informação gráfica

fornecida pelo "zoom" mostrará a qualidade e estrutura das ordenações conseguidas, conforme a figura 2.

Figura 2 - Zoom da matriz de co-ocorrência de Descritores

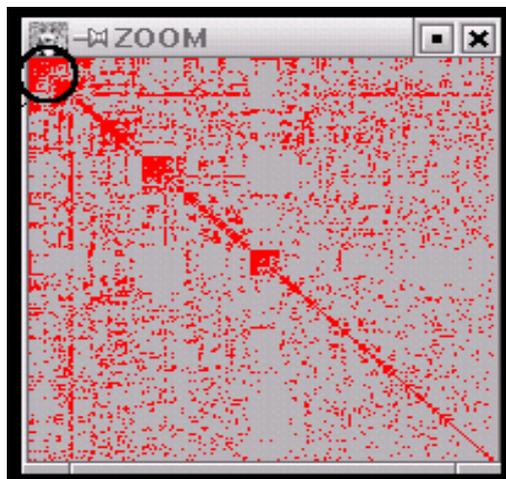


Figura 3 - Fragmento da matriz de co-ocorrência descritores, segundo 1% das palavras de maior frequência, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

	EPIDEMI	RISK_FA	PREVALE	SOCIOEC	MORTALI	CAUSES_	HIV/AI
epidemi	350	43	37	23	15	1	15
risk_fa	43	112	17	12	13	2	6
prevale	37	17	89	6	1		3
socioec	23	12	6	58	3	2	3
mortali	15	13	1	3	124	30	2
causes_	1	2		2	30	36	1
hiv/aid	15	6	3	3	2	1	220
mycobac	11	2		1	1		18
adolesc	6	5	4	4	4	3	4
dental	18	2	4	3			
nutriti	5	3	2	7	2	1	1
obesity	11	5	4	1			
insulin	4	3	3		5	3	
hyperte	9	3	2		1		
blood_p	2	1	2				
pregnan	5	3	3	3	4	2	2

A figura 3 representa um fragmento da matriz de co-ocorrência de descritores e mostra algumas palavras do cluster e o grau de proximidade ou força de associação entre elas. Por exemplo, a palavra "epidemiology" co-ocorre com "risk factor" 43 vezes do total de sua aparição no campo de descritores. Na figura 4 descrevemos o total de palavras que compõe este grupo. Algumas palavras demonstram o campo de estudo como HIV/AIDS, mas quando associada a epidemiology ou "risk factor" ou "prevalence" podemos inferir com que finalidade. Sua forte relação com a palavra "mycobacterium" também sugere a importância da investigação e da pesquisa quando da associação dessas duas infecções, pois o advento da AIDS modificou o perfil epidemiológico da Tuberculose.

Figura 4 - Grupo de descritores, liderado pela palavra "epidemiology", das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

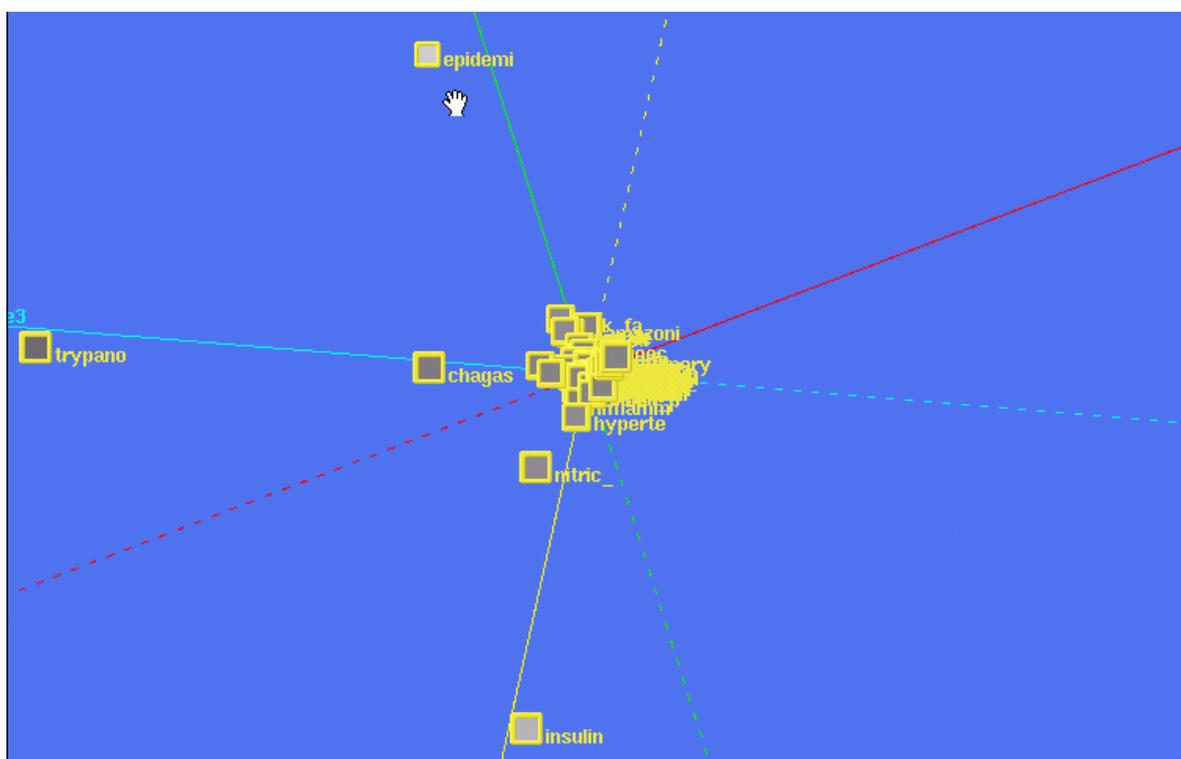
epidemiology	obesity
risk factors	insulin/diabetes
prevalence studies	hypertension
socioeconomic factors	blood pressure
mortality	pregnancy
causes of death	nitric oxide
hiv/aids	rat
mycobacterium/tuberculosis	child
adolescent	smoking
dental	alcoholism/alcohol
nutrition/malnutrition	diet

Para construção dos mapas, foi utilizado método estatístico integrado, isto é, aplicou-se à matriz original, de números absolutos, a análise de componentes principais, para fazer emergir as palavras mais fortes. Esta análise permite identificar os eixos principais de

uma nuvem de pontos e a melhor representação desta nuvem num triedro formado pelos primeiros eixos.

A figura 5 evidencia algumas das palavras mais freqüentes, lideradas por "epidemiology", porém as demais formam uma nuvem de pontos tornando impossível visualizá-las.

Figura 5 - Mapa segundo matriz de co-ocorrência de descritores, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.



Na figura 6 temos outro fragmento da matriz de descritores, onde delimitamos novo "cluster", como evidencia o "zoom" no mesmo quadro. Este é composto por 14 palavras, lideradas pela palavra "anxiety", conforme descritas na figura 7.

Figura 6 - Fragmento da matriz de descritores, liderado pela palavra "anxiety", das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

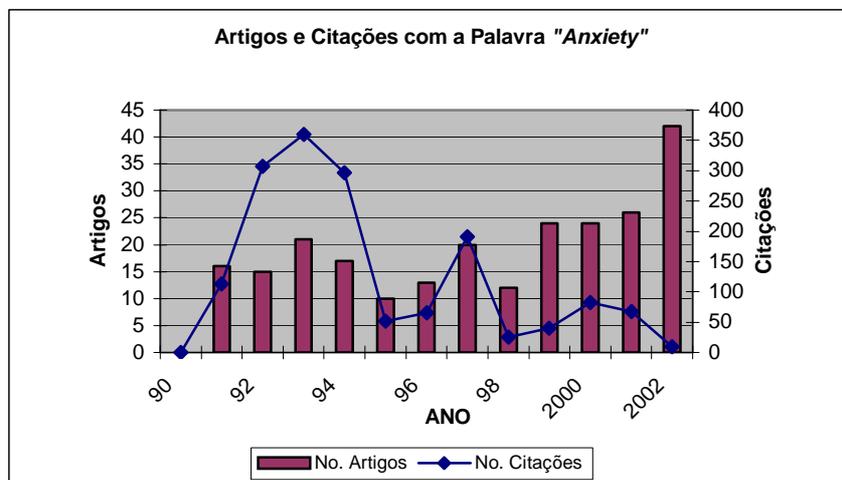
DIAGNOS	EOSINOP	ASTHMA	PROSTAGAN	ANXIETY	ELEVATE	MEMORY	BENZODI	HIPPOCA	INHIBIT	AMYGDAL	SEROTON	LEARN
	4	46										
			39									
				120	28	10	18	2	2	11	10	5
				28	44	2	4	1		3	1	1
				10	2	100	9	12	26	6	2	20
				18	4	9	58	1	2	6	3	1
				2	1	12	1	96	10	12	6	5
				2		26	2	10	41	6		2
				11	3	6	6	12	6	46	7	1
				10	1	2	3	6		7	75	1
				5	1	20	1	5	2	1	1	32
				4	1	2	1	3	1	1	4	
				2	2			1		4	2	
				10	4		5		1	9	5	
					1	4				1	4	

Figura 7 - Descritores liderados pela palavra "anxiety", das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

120 anxiety	44 elevated plus-maze
101 memory	45 periaqueductal gray
97 dopamine	41 inhibitory avoidance
96 hippocampus	37 fear
76 serotonin	32 learning
60 benzodiazepine	25 aversion
50 amygdala	25 superoxide

Ao acompanharmos a evolução desta palavra no período estudado, considerando as fronteiras impostas pelo escopo deste estudo, encontramos uma rede de 45 instituições colaboradoras. Observamos conforme o gráfico 10, que publicações que contém "anxiety" como descritor ou identificador de assunto, tiveram uma clara tendência de crescimento, embora com alguma oscilação em 1995, 1996 e 1998. Já as citações alcançam seu número máximo no período de 1991 a 1994, recuperando-se em 1998 e novo declínio até 2002. No entanto, é de se esperar que publicações recentes tenham um número menor de citações.

Gráfico 10 - Artigos e citações segundo a palavra "anxiety", registrado nos campos descritor e identificador, das publicações brasileiras, na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.



Na tabela 20, destacamos as instituições mais envolvidas com o termo "anxiety" e temos como líder a Universidade de São Paulo (USP) com 53,1% do total de 177 publicações e 68,93% do total das 1.926 citações no período. Optou-se por subdividir a USP em campi São Paulo e campi Ribeirão Preto devido à diferença de produção e número de citações da USP-Ribeirão Preto, que responde sozinha por 37,84% das publicações e 48,34% das citações. As colaborações internacionais com esta universidade se dão principalmente com as Universidades de Hull e Manchester e com o Guys Hospital na Inglaterra, com os laboratórios LNBC, CTR Neuroch. E SYTHELABO Recherche na França, as Universidades de Dusseldorf na Alemanha, de Mendoza na Argentina e a Hebraica de Jerusalém entre outras.

Tabela 20 - Artigos e citações que contém a palavra "anxiety", segundo Universidades, das publicações brasileiras, na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

Universidades	Artigos	% Artigos	Citações	% Citações
USP-Ribeirão Preto	67	37,84	931	48,34
USP-São Paulo	27	15,26	246	12,77
UFSta.Catarina	22	12,44	139	7,22
UNIFESP	20	11,3	47	2,44
U.Maringá	9	5,08	195	10,14
UNESP	8	4,52	49	2,54
UFRSul	8	4,52	142	7,37
UNICAMP	6	3,39	26	1,35
U.F. Espírito Santo	6	3,39	28	1,45
UFRio Janeiro	2	1,13	101	5,24
UFParaná	2	1,13	22	1,14
Total	177	100	1926	100

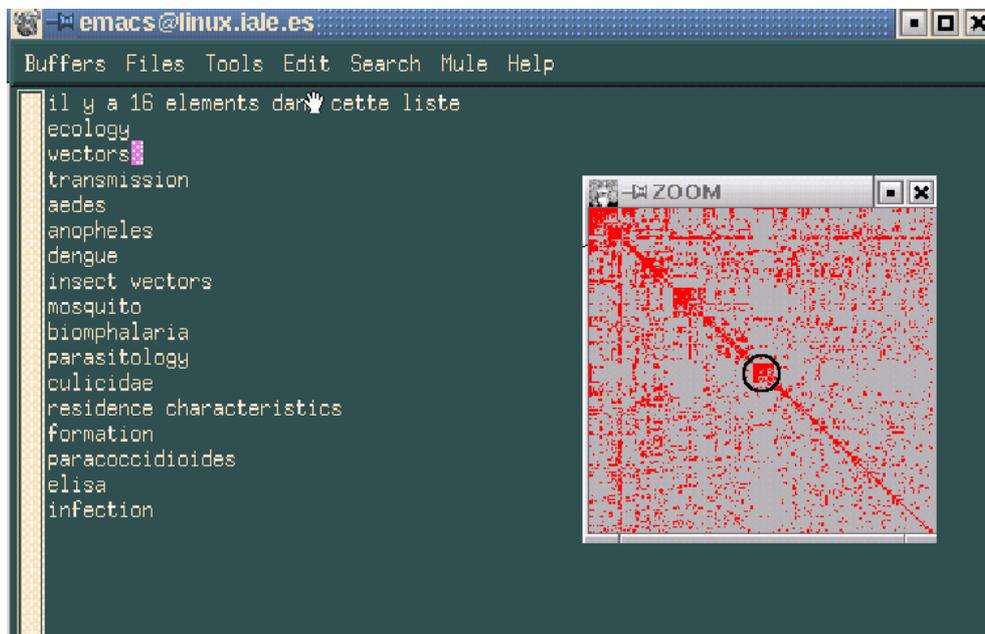
Outras instituições de destaque são a Universidade Federal de São Paulo e a Universidade Federal de Santa Catarina. Esta última têm como colaboradores a Universidade de "Leeds" na Inglaterra, mais recentemente a Universidade de Ferrara na Itália, mas sendo o "Institut National de la Santé et de la Recherche Médicale, -INSERM- da França, seu principal colaborador no exterior.

Pode-se dizer que tanto a USP quanto a UFSC, apresentam grupos de pesquisa consolidados nesse campo pois além da produtividade e impacto de seus trabalhos, ambas mantêm uma atividade contínua ao longo do período.

Figura 8 - Fragmento da matriz de co-ocorrência de descritores, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

ECOLOGY	VECTORS	TRANSMI	AEDES	ANOPHEL	DENGUE	INSECT_	MOSQUIT	BIOMPHA	PARASIT
				2				2	
			1						
100	56	12	20	13	5	12	9	2	1
56	62	10	17	11	5	8	8	2	1
12	10	49	9	3	6	5	1	3	4
20	17	9	52	1	13	8	7		
13	11	3	1	45		3	6		2
5	5	6	13		24	5	2		
12	8	5	8	3	5	21	1		3
9	8	1	7	6	2	1	31		
2	2	3						48	9
1	1	4		2		3		9	48
11	8	2	4		1	9	2		1
2	2	1	2		1			2	
								1	1

Figura 9 - Lista de descritores, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.



A figura 8 é um fragmento de uma matriz de co-ocorrência de descritores, simples e simétrica, onde está evidenciado o "cluster ecology" e a figura 9 é o resultado da decomposição dessa matriz na sua síntese gráfica, representado pelo "zoom" e na tábua que mostra as palavras do "cluster" bem delimitado e que está liderado pela palavra ecologia, contém 16 elementos, a saber: ***Ecology, vectors, transmission, aedes, anopheles, dengue, insect, vectors, mosquito, biomphalaria, parasitology, culicidae, residence characteristics, formation, paracoccidioides, Elisa e infection***". Este grupo de palavras foi escolhido para representar os exemplos de mapas tecnológicos e para evidenciar a potencialidade das análises de co-ocorrência com as demais variáveis que estão demonstradas a seguir.

Quando os dados são da mesma natureza, como o cruzamento de descritores com descritores, utiliza-se a métrica euclidiana. Nos mapas da figura 10 e 11, utilizou-se análise dos componentes principais reduzidos com a classificação por análise hierárquica com métrica euclidiana. A análise dos componentes principais reduzidos (ACPR) parte de normalizar os dados e equilibrar os caracteres analisados. A redução dos dados permite considerar as variáveis de menor valor absoluto. Nestes mapas, são evidenciados distintos grupos de palavras que possuem maior afinidade entre si e estão representados por cores ou apresentam a lista de palavras. A figura 10 mostra o mapa com 4 grupos diferenciados pela cor segundo a proximidade das palavras, o que permite diferenciar "clusters" bem definidos.

Figura 10 - Mapa dos "clusters", segundo matriz de co-ocorrência de descritores, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

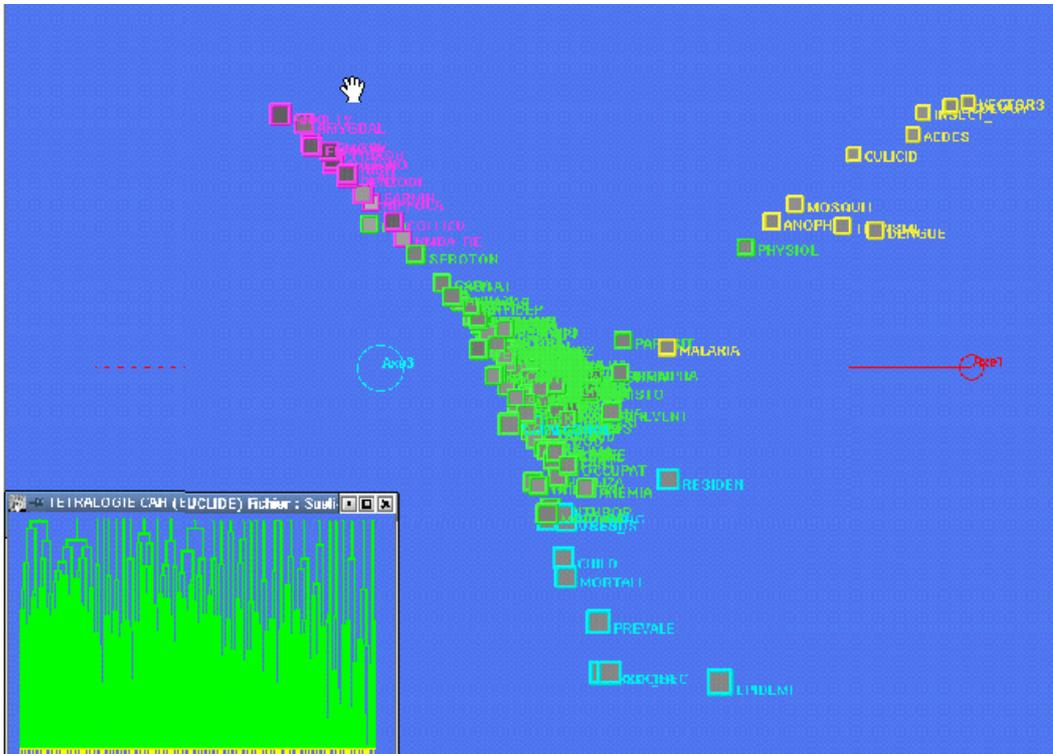


Figura 11 - Mapa segundo matriz de co-ocorrência de descritores, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.



Já o mapa da figura 11, evidencia o grupo que aparece na matriz da figura 8 que será estudado quanto as suas relações com autores, identificadores e instituições.

6.2.3. ANÁLISE DOS IDENTIFICADORES

Os identificadores somaram 40.160 palavras diferentes no total de 38.349 registros em todo o banco de dados e correspondem às palavras-chave presentes no campo de identificadores de cada artigo. Abaixo, a tabela 21 apresenta os 20 primeiros identificadores por ordem decrescente de frequência, antes e após o tratamento com o arquivo de sinônimos. Palavras como "*protein*" e "*rats*", por exemplo, tiveram sua frequência praticamente duplicada após a correção pelos sinônimos.

Tabela 21 - Identificadores, liderado pela palavra "expression" e "protein", das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

<i>FREQUÊNCIA DE IDENTIFICADORES</i>		<i>FREQUÊNCIA DE IDENTIFICADORES APÓS SINÔNIMOS</i>	
741	EXPRESSION	835	PROTEIN
581	CELLS	765	RAT
547	IDENTIFICATION	741	EXPRESSION
518	DISEASE	631	INFECTION
509	INFECTION	582	CELLS
438	GENE	570	MICE
433	PROTEIN	562	DISEASE
413	RAT	547	IDENTIFICATION
404	MICE	468	RECEPTOR
389	PROTEINS	438	MUTATIONS
374	DNA	438	GENE
366	CHILDREN	383	CHILDREN
360	BINDING	378	BINDING
358	INHIBITION	374	DNA
357	BRAIN	361	INHIBITION
342	GROWTH	357	BRAIN
340	RATS	342	GROWTH
315	PURIFICATION	336	POPULATION
297	BRAZIL	323	CANCER
282	ACTIVATION	315	PURIFICATION

A tabela 22 demonstra o percentual dos identificadores segundo sua frequência e observamos que 63,54% destes aparecem apenas uma vez como identificador de assunto de um artigo, um percentual menor se comparado com os descritores. Somente 36,46% dos 40.160 identificadores, têm frequência >2. Ainda assim, apenas 2,96% das palavras têm uma frequência de ocorrência de até 10 vezes, no período estudado.

Tabela 22 - Percentual dos Identificadores segundo sua frequência de ocorrência, nas publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

<i>Freqüência</i>	<i>Descritores</i>	<i>%</i>
1	25.518	63,54
>1	14.642	36,46
Total	40.160	100

Para avaliar as relações entre identificadores e descritores, foi realizado um cruzamento com o campo dos descritores e identificadores. Utilizamos o total de indicadores e apenas descritores constantes da figura 39 para fazer emergir a co-ocorrência de palavras entre eles. Ambos arquivos foram antes tratados com seus respectivos filtros de sinônimos.

A visualização dos resultados é dada primeiramente por uma matriz de co-ocorrência simples e assimétrica, pois trabalhamos campos diferentes, e em arquivos do editor de texto, "ultra edit".

A melhor maneira para visualizar os mesmos resultados de forma gráfica são os mapas tecnológicos. Na figura 12 podemos observar em cores diferentes os descritores e identificadores, formando 5 sub-grupos de palavras com maior afinidade entre si. Podemos selecionar os grupos para obtermos a lista de palavras correspondentes a cada grupo. Feita esta seleção,

obtivemos os resultados que podem ser visualizados em editor de texto.

Figura 12 - Grupo 4 de descritores, das publicações brasileiras, liderado pela palavra ecologia, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.



Os resultados são emitidos por meio de uma lista de palavras em 2 colunas. A coluna a direita mostra os identificadores em letra minúscula. A coluna da esquerda, que aparecem em letras maiúsculas, evidencia os descritores, ou seja, a relação existente entre um ou mais descritores, neste caso apenas os relacionados na figura 8, com todos os identificadores que tem proximidades com essas palavras. Isto significa que o descritor em evidência co-ocorre com os diferentes identificadores apresentados, num mesmo artigo. Os 5 sub-grupos originados desta análise, com 1 ou mais descritores liderando cada grupo de palavras associadas,

estão demonstrados nas figuras 13, 14, 15, 16 e 17. Um único descritor pode estar ligado a vários termos que estejam ocorrendo juntos em identificadores, o que pode significar que o âmbito do trabalho é maior, mais amplo do que fixou o "thesaurus".

Figura 13 - Cruzamento de descritores e identificadores, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

<i>Descritores</i>	<i>Identificadores</i>
Sub-Grupo 1	receptor expression
	antigen antibody strains population lymphocyte mammalian-cells modulation form chagas-disease susceptibility trypomastigotes
PARASITOLOGY	phosphate suppression
BIOMPHALARIA	complexes hemiptera soil genome parasite hymenoptera components stage aspartate receptors markers mitochondrial-dna hypersensitivity fungus pentatomidae ribosomal-rna mutant arbitrary primers schistosoma-mansoni hosts schistosoma pcr vesicular-arbuscular mycorrhizas seedlings biomphalaria-glabrata nucleotide-sequences cellulase glabrata hemocytes

Figura 14 - Cruzamento de descritores e identificadores, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

<i>Descritores</i>	<i>Identificadores</i>
Sub-Grupo 2	mice brain growth mutations activation diagnosis serum macrophages tumor-necrosis-factor tissue inflammation/inflammatory
PARACOCCIDIOIDES	ethanol immune-responses matrix subsets mycobacterium-tuberculosis host p.brasiliensis cytokine production in-vitro ketoconazole granuloma inbred mouse strains plating efficiency natural-resistance

Figura 15 - Cruzamento de descritores e identificadores, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

<i>Descritores</i>	<i>Identificadores</i>
Sub-Grupo 3	children cancer trends helminth communities
RESIDENCE CHARACTERISTICS	density competition model stability survival richness age vector income undernutrition population community health competition culicidae diptera america iron-deficiency supplementation cancer mortality

Figura 16 - Cruzamento de descritores e identificadores, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

<i>Descritores</i>	<i>Identificadores</i>
	monoclonal-antibodies
	brazil
Sub-Grupo 4	virus
	united-states
	pathogenesis
	diptera
	transmission
ECOLOGY	sao-paulo state
VECTORS	organization
MOSQUITO	immunosorbent-assay
AEDES	culicidae
ANOPHELES	biology
DENGUE	mosquito
CULICIDAE	psychodidae
	vector
	fixation
	fever
	ribeira valley
	project
	pyrethroids
	field-evaluation
	hemorrhagic-disease
	irrigation

Ao estudarmos por exemplo, a figura 48, os resultados das relações entre descritores e identificadores, temos que o grupo de palavras liderado por "ecology" co-ocorre com "**ribeira valley, irrigation, mosquito, sao-paulo state, psychodidae, culicidae, hemorrhagic-disease, vector**", entre outras e podemos concluir, por exemplo, que a linha de pesquisa está voltada para insetos vetores e que o grupo tem investigado algumas regiões do estado, ricas em material para seus estudos, onde as condições são favoráveis ao

desenvolvimento desses insetos e à manutenção da transmissão das doenças por eles veiculados.

Figura 17 - Cruzamento de descritores e identificadores, das publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

<i>Descritores</i>	<i>Identificadores</i>
Sub-Grupo 5	necrotic tissue decaying stems stenocereus gummosus opuntia stricta drosophila ecology identification classification
INSECT VECTORS	philogeny lice culicidae culicidae trypanosoma cruzi mosquito afrotropical anopheles infectivity dynamics sporozoites parasites diptera anthropic environment productivity virus skuse

6.2.4. ANÁLISE DOS AUTORES

O produto inicial da contagem realizada para o campo de autores continha 148.971 registros, contando a totalidade dos autores presentes no banco de dados e que estão registrados nos campos AU1, AU2, AU3... até AU6. Neste caso, além das duplicações, em alguns registros as informações presentes no campo endereço, devido sua formatação original na base de dados, "invadiram" o campo de autores. Após a retirada de dados não pertinentes e tratamento morfológico, este campo livre de duplicações e interferências conta com 73.644 autores, ou seja, a contagem inicial somava 49,43% registros indesejáveis.

Para exemplificar estudos com autores, tomamos por referência as palavras lideradas pelo grupo "ecology" que estão evidenciadas no fragmento de matriz da figura 8 e na figura 9. Foi realizada uma contagem de autores, com até 3 publicações no período, utilizando-se como filtro este grupo de palavras. Após selecionar os autores pertencentes a esse grupo, relacionados no anexo 4, o cruzamento entre autores resultou numa matriz, cujo fragmento está representado na figura 18. Neste, a função "zoom" evidencia a síntese gráfica da matriz, onde "clusters" aparecem bem delimitados. É possível consultar o conteúdo de cada "cluster" e dessa forma visualizar a que grupo se refere. Do "cluster" assinalado com a seta, figura 18, emerge um grupo de autores, doravante chamado "cluster ecology", aqui representado pelo autor "Forattini".

Figura 18 - Autores do "cluster ecology" segundo publicações brasileiras, no campo da saúde, registradas na base de dados da "web of Knowledge", no período de 1990 a 2002.

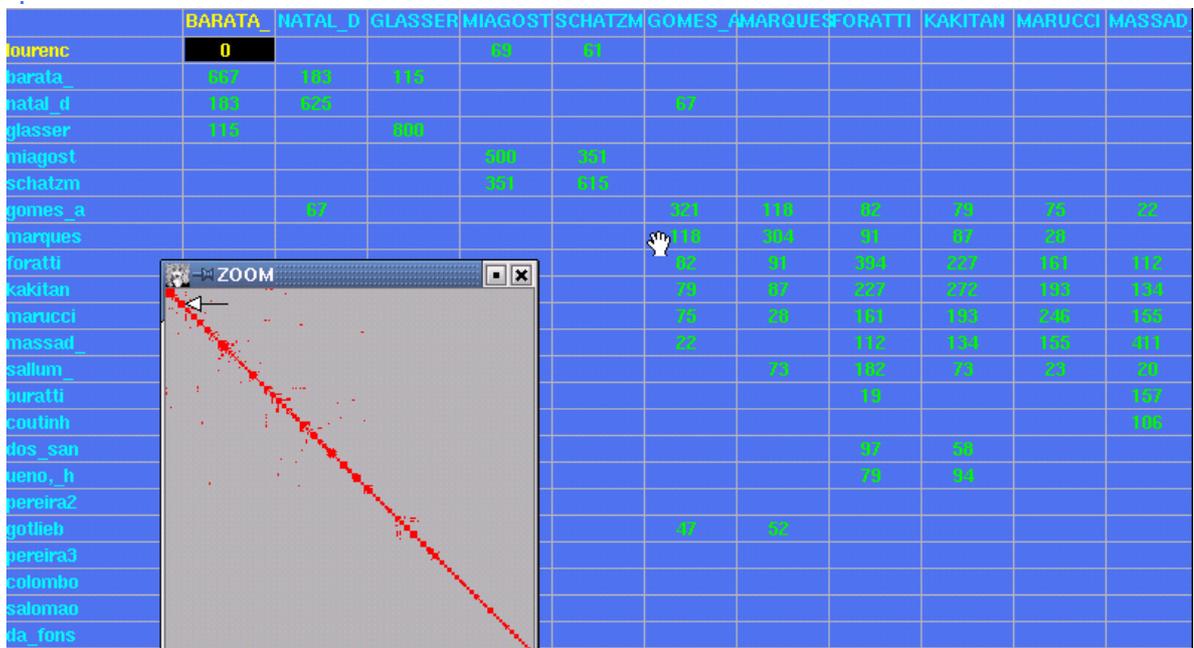
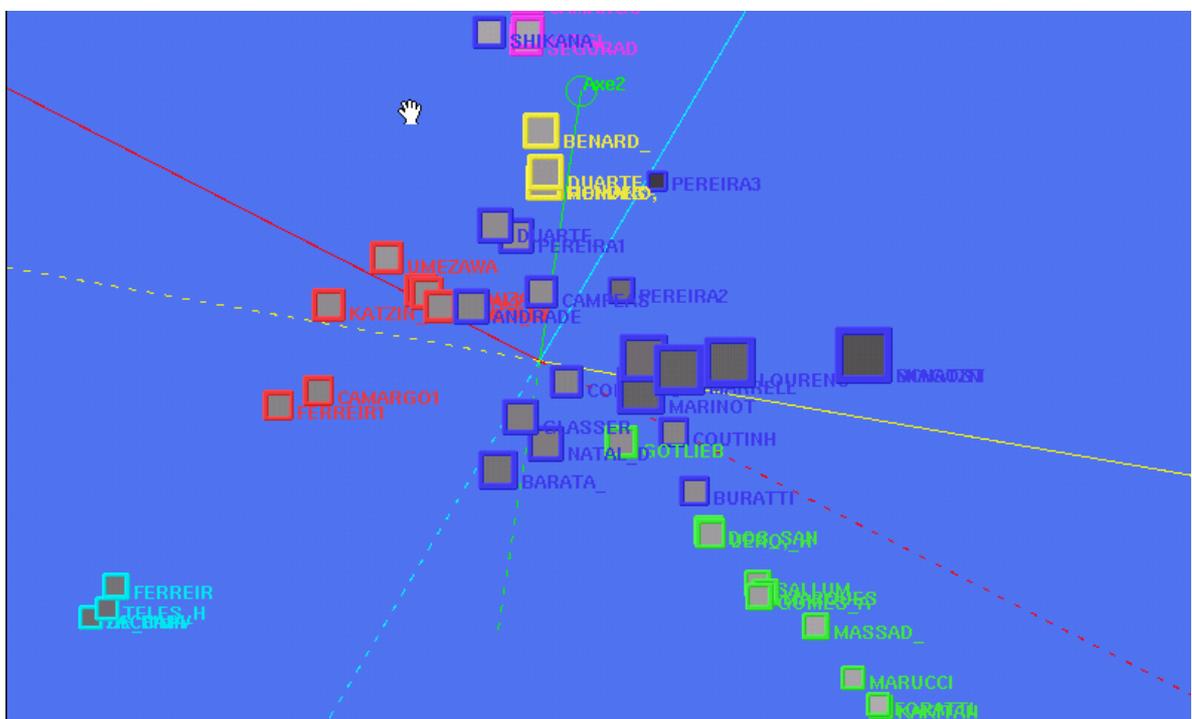


Figura 19 - Mapa de autores do "cluster" ecology, segundo descritores, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.



Na figura 19, podemos ver no mapa tecnológico, com identificação colorida, os grupos de autores que compõe o "cluster ecology". Nota-se no mapa uma interface entre os "clusters" liderados por "Forattini" e "Burattini", devido a eventuais colaborações e na figura 20 temos um mapa de autores onde distinguimos, no detalhe ao lado, a descrição do grupo liderado pelo autor "Forattini".

Figura 20 - Mapa de autores do "cluster" ecology, segundo descritores, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

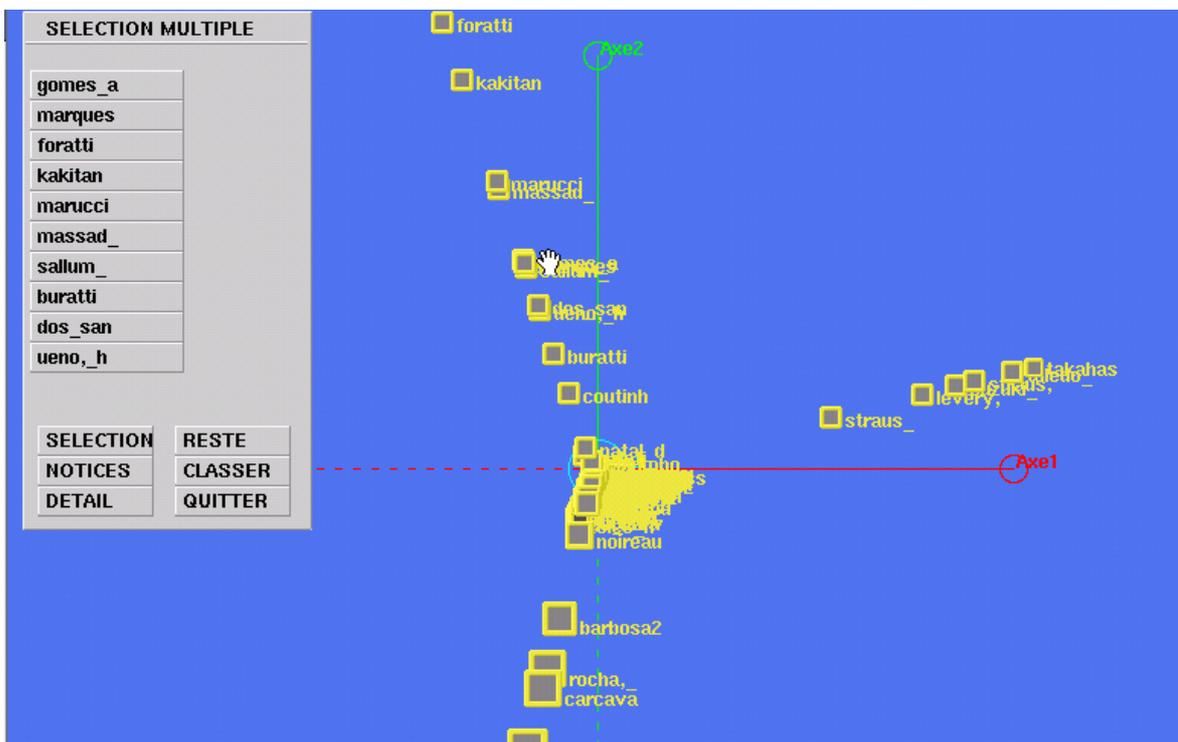
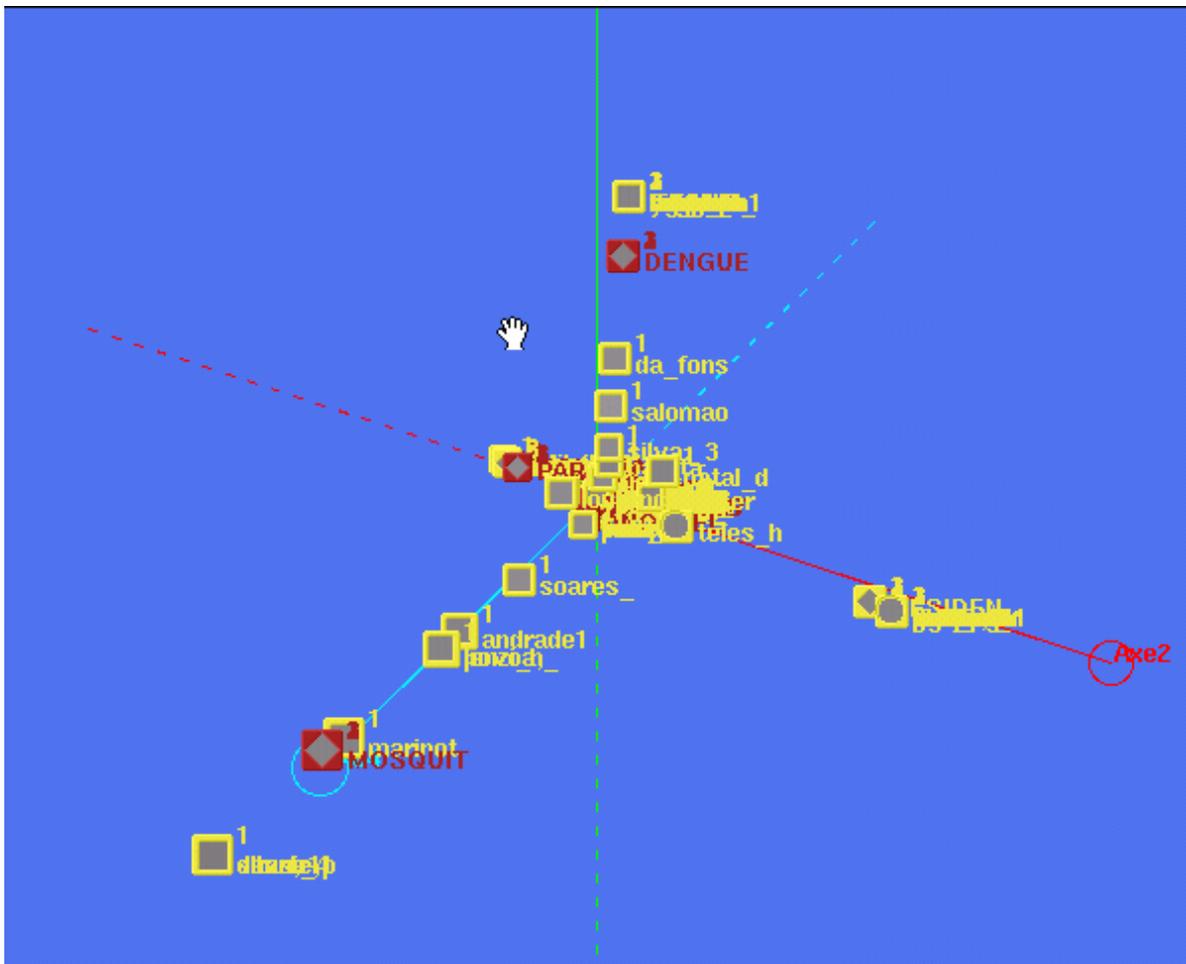
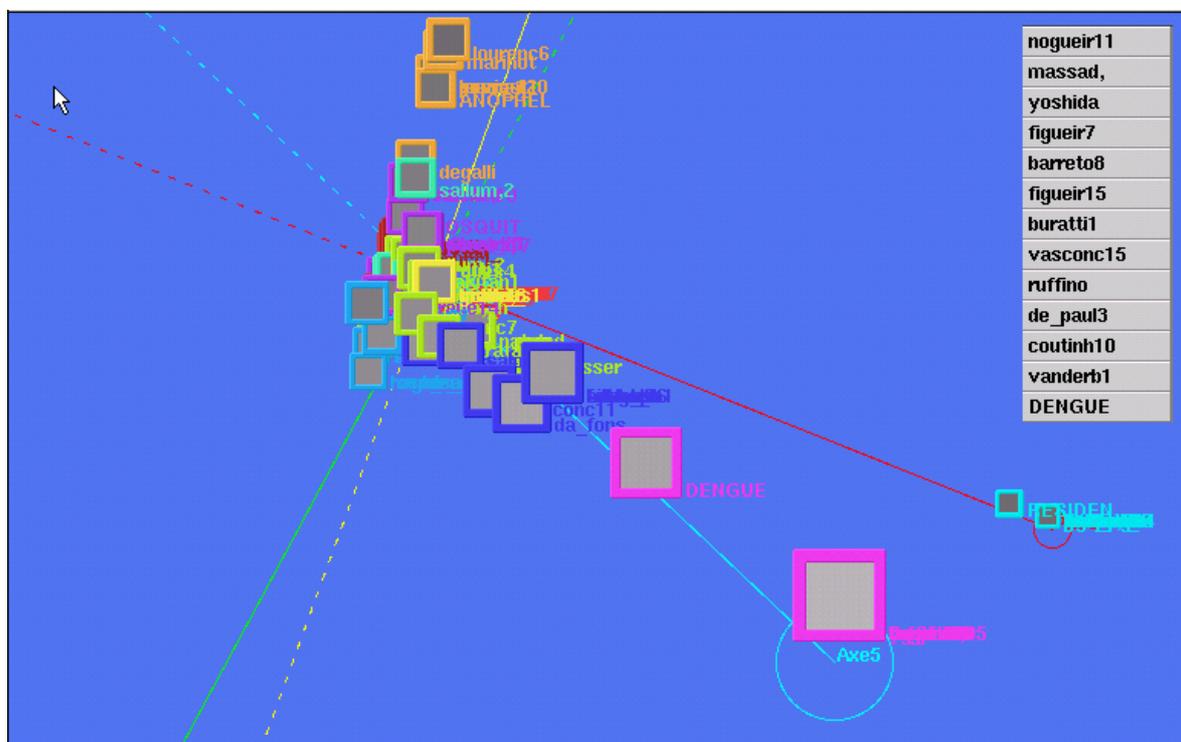


Figura 21 - Mapa tecnológico de autores e descritores do "cluster ecology", segundo publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.



Na figura 21 e 22 os mapas tecnológicos representam uma matriz simples assimétrica de co-ocorrência dos campos descritores e autores, onde podemos verificar a proximidade de grupos de palavras descritoras de temas e seus respectivos autores, na figura 22 o detalhe indica autores que tem maior afinidade com palavra dengue.

Figura 22 - Mapa tecnológico de autores e descritores do "cluster ecology", segundo publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

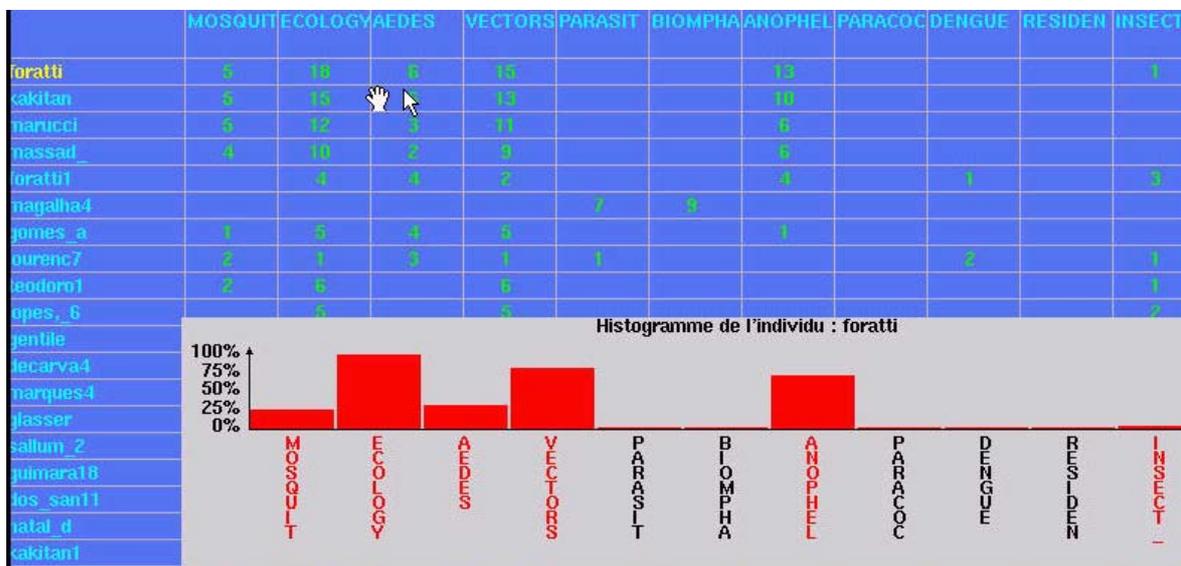


Na figura 23 mostramos a evolução das colaborações entre os autores do grupo "forattini".

Figura 23 - Evolução do Grupo de autores relacionados ao autor "Forattini" nos períodos 1990-1994, 1995-1998 e 1999-2002, segundo publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde.

1990 - 1994	1995 - 1998	1999 - 2002
**** Classe numero 12 ****	**** Classe numero 11 ****	**** Classe numero 14 ****
marques gram	zerbini caf	massad e
kakitani i	jorge mhpd	forattini op
forattini op	gotlieb sld	marques gram
sallum mam	marques gram	kakitani i
flores dc	gomes ad	marucci d
marques ccd	sallum mam	santos jlf
gomes ad	forattini op	gomes ad
santos jlf	kakitani i	sallum mam
marucci d	marucci d	galati eab
massad e	massad e	de brito m
	burattini mn	wilkerson rc
	saldiva phn	bergo es
	martins ma	de Brito M
	vaz fac	ueno hm
	paiva pso	kobayashi km
	lichtenfels ajfc	dos santos rl
	schultz r	fernandez z
	zanetta dmt	flores dc
	sakae rs	burattini mn
	lorenzi g	
	dolhnikoff m	
	conceicao gms	
	bohm gm	
	duarte mis	
	pereira pm	
	shikanaiyasuda ma	
	caramelli b	
	santos rtm	
	neto va	
	alves vaf	
	sartori amc	
	benard g	
	duarte ajs	
	hong ma	
	sato mn	
	campeas ae	
	barbosa ad	
	wakamatsu a	
	grumach as	
	bellinatipires r	
	salgado mm	
	ueda m	
	flores dc	
	de Rezende L	

Figura 24 - Matriz de descritores e autores do "cluster ecology", segundo as publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.



Na figura 24 temos a representação dessa matriz e um histograma com os resultados da relação dessas palavras com o autor "Forattini", que em 100% de seus artigos publicados, no período estudado, utilizou a palavra "ecology", que evidencia a linha de pesquisa desse grupo de autores.

O estudo das colaborações entre instituições, embora importantes, nem sempre é fácil de ser realizado, devido à qualidade da informação nas bases de dados. Para avaliar o campo de filiação do autor foi realizada uma curva de carga, assim designada porque reflete os registros de informação nova. Neste caso, figura 25, observamos uma curva ascendente que tende ao infinito, ou seja, o número de informações novas não se estabiliza. Esta curva, por exemplo, que expressa o que está registrado nos campos C01 até C06, que significam a

filiação profissional do autor, demonstra a urgente necessidade de uniformização do registro de informações como nome de universidades, institutos de pesquisa e outras instituições. As diferentes formas de registrar uma mesma informação, assim como já demonstramos, inclusive com autores, chega a triplicar dados idênticos de um mesmo campo.

Figura 25 - Curva de carga para a contagem do campo de endereços, segundo as publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.

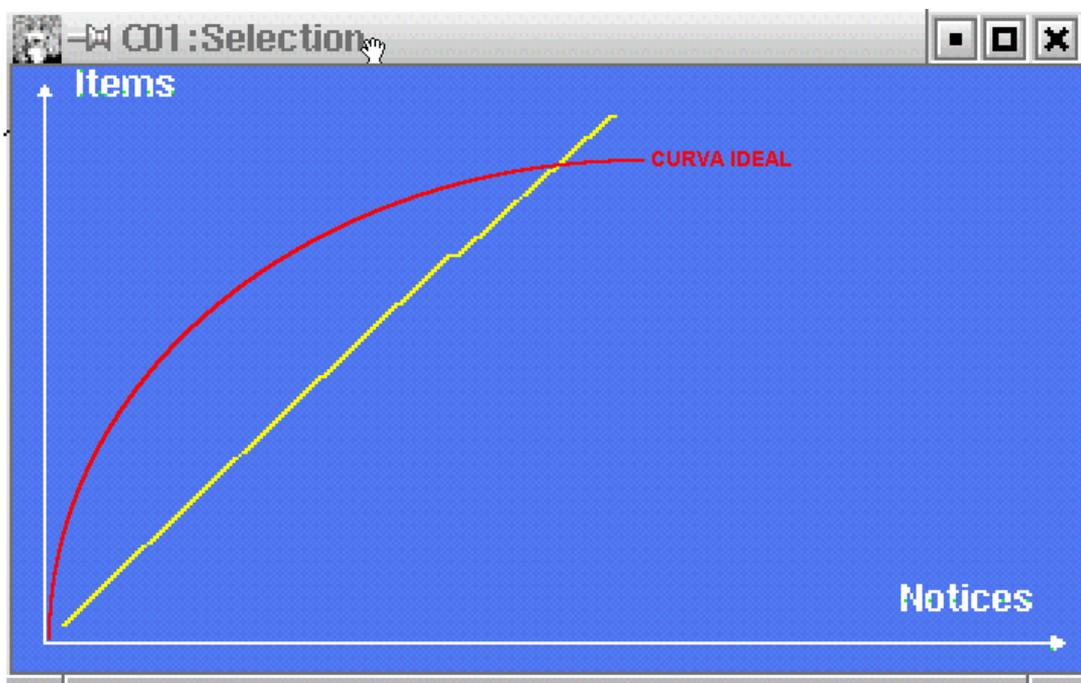
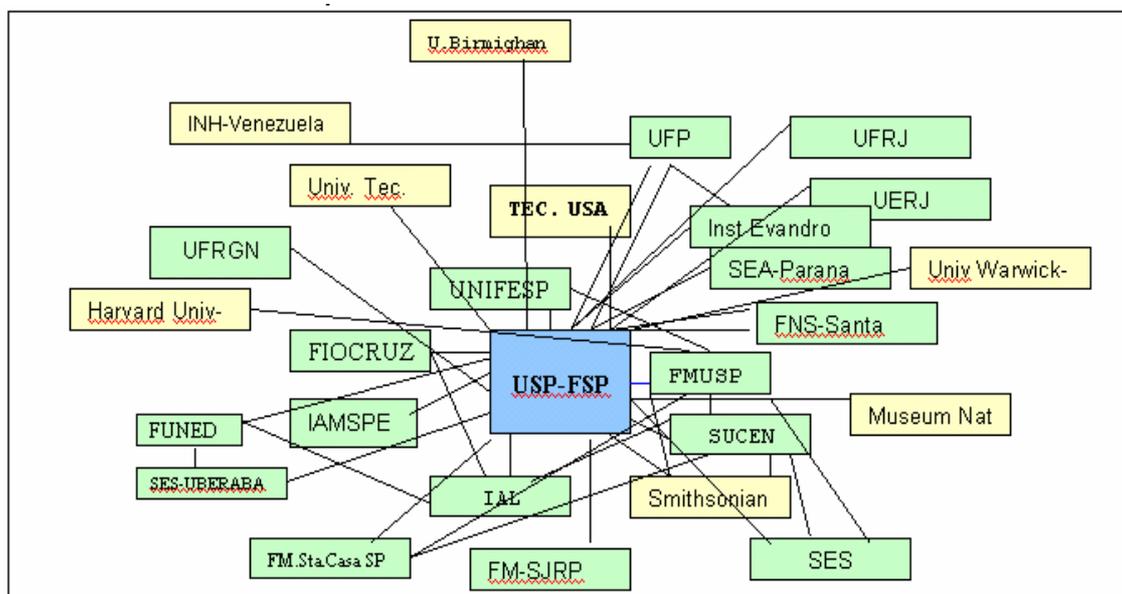


Figura 26 - Rede de colaboradores do "cluster ecology", segundo as publicações brasileiras, registradas na base de dados da "web of Knowledge", no campo da saúde, no período de 1990 a 2002.



A figura 26 mostra a rede de instituições parceiras do autor "Forattini" e seus principais colaboradores. A análise da dinâmica de colaboração do grupo, realizada em três períodos, 1990-1993, 1994-1997 e 1998-2002, mostra que 10 membros do grupo já no período inicial, cooperam em menor ou maior intensidade em todos os 3 períodos, figura 23. Este é um grupo consolidado, que está filiado principalmente à Faculdade de Saúde Pública, Faculdade de Medicina da USP e a SUCEN. Tomando-se o autor "Forattini" como líder do grupo, este publicou 52 artigos e recebeu 157 citações, uma média de 3,1 citações por artigo. Este grupo tem padrão de comportamento semelhante aos "clusters" 1 e 2, com alta média de citações por artigo publicado e, o que pode sugerir linhas de pesquisa emergentes, maior impacto e visibilidade das suas publicações, uma vez que atualmente

é consenso que número de citações não se traduz, obrigatoriamente, em qualidade da investigação científica. Como os "clusters" 1 e 2, é um campo estável e consolidado que talvez represente o padrão no campo das ciências da saúde.

7 - CONSIDERAÇÕES FINAIS

A necessidade de equacionar problemas sociais e impulsionar o crescimento econômico coloca a Ciência e Tecnologia (C&T) no centro das estratégias e políticas de desenvolvimento sustentado em países em desenvolvimento, como o Brasil.

É necessário fomentar esta área nas esferas pública e privada, e mais, fazê-lo de forma ordenada e firme, orientada por uma visão de futuro, articulada com as necessidades e demandas dos cidadãos.

A idéia de que a pesquisa em saúde é uma ferramenta importante para a melhoria da situação de saúde das populações, bem como, para a tomada de decisões na definição de políticas e no planejamento em saúde, vem se fortalecendo desde 1980. Isso tem contribuído para a melhoria das ações de promoção, proteção, recuperação e reabilitação da saúde e a diminuição das desigualdades sociais. É do conhecimento também, que a exemplo do que ocorre em vários países, o setor saúde representa o maior componente de toda a produção científica e tecnológica.

É importante avaliar o rendimento da atividade científica, seu impacto e benefícios na sociedade, com a finalidade primordial de adequar a distribuição de recursos. Para esse fim tem-se utilizado indicadores de insumos destinados a medir recursos humanos e gastos; indicadores de inovação tecnológica; indicadores de impacto da ciência e tecnologia no desenvolvimento social e indicadores bibliométricos.

Indicadores bibliométricos são complementares, mas não substituem meios tradicionais de avaliação de pesquisa. As publicações não são o único tipo de produção que pode ser medido. Na verdade, documentos internos, publicações científicas em apresentações ou conferências são atividades expressivas da dinâmica da produção científica, entre todos os tipos de organizações que tem interesse no desenvolvimento científico e tecnológico de alguma área do conhecimento.

Tomando-se em conta que indicadores bibliométricos são de um lado pouco relevantes, se simplesmente praticarmos a contagem de artigos produzidos; e de outro, pouco objetivos, se realizarmos somente análises de avaliação por pares, figura entre estes dois extremos um grande número de técnicas bibliométricas capazes de avaliar a produção científica e tecnológica de diferentes campos do conhecimento de forma satisfatória. No entanto deve estar sempre presente nas avaliações deste tipo, um olhar crítico de que os indicadores devem ser empregados com cautela e preferencialmente para grupos homogêneos de pesquisadores e que trabalhem em áreas afins.

Ao proceder a avaliação da produção científica do campo da saúde, no Brasil, de 1990 a 2002, de acordo com as publicações registradas nas bases de dados do Institut for Scientific Information, Web of Knowledge, os resultados encontrados evidenciam que foram publicados 38.349 artigos e a produção científica em saúde no país, cresceu a uma taxa anual média de 14,98% nesse período. Destes, 31.191, ou seja, 87% foram publicados no

exterior, o que dá maior visibilidade à ciência brasileira, na área da saúde. O número de autores envolvidos nessa produção somou 73.644, a colaboração entre autores concentra, até 10 autores.

O número médio de autores por artigo cresceu de 3,7 autores para 5,9 aproximadamente, no período estudado, ou seja aumentou o engajamento entre os cientistas brasileiros, bem como a busca de melhor desempenho, ou seja, a necessidade de conjugar esforços financeiros e competência técnica e científica, conseqüentemente aumentando o grau de parcerias, rumo a maior produtividade e competitividade, que por sua vez facilitam a obtenção de recursos para o financiamento de pesquisas.

Ficou comprovado o princípio da Lei de Lotka, que afirma que a proporção daqueles que fazem uma única contribuição em um determinado campo científico, é de mais ou menos 60%, pois encontramos um valor muito próximo dessa afirmação, 62,17%.

Análises recentes mostraram que o número de artigos sobre pesquisa genômica publicados por brasileiros, em periódicos científicos internacionais indexados, cresceu 72,4% entre 1998 e 2003 e que o número de citações também é crescente. Pode-se inferir, no caso brasileiro, que algumas áreas do conhecimento tiveram peso específico no desempenho em alta. A pesquisa genômica certamente é uma delas.

Os resultados observados no nosso trabalho vão ao encontro dessas afirmações, pois a área "Molecular

Biology and Genetics" pertence ao "cluster" 2 do nosso estudo. Este "cluster" possui uma alta produção, com uma média de 1596,76 publicações e foi o grupo que mais citações recebeu. Tem um número médio de 5,57 citações por artigo publicado e, portanto pode sugerir campos emergentes ou campos de maior impacto das suas publicações. Atualmente é consenso que número de citações não se traduz, necessariamente, em qualidade da investigação científica, mas indica o impacto, maior visibilidade e reconhecimento pelos pares.

Do total das 3.066 revistas pesquisadas neste trabalho, 24 revistas representam as que contêm o maior número de artigos publicados, 12.921, ou seja 540 artigos por revista e são responsáveis por 33,7% do total das publicações. A este grupo chamamos "grupo hard core" de Bradford, que segundo sua "Lei de Dispersão", permite, mediante a medição da produtividade das revistas, estabelecer o núcleo e as áreas de dispersão sobre um determinado assunto em um mesmo conjunto de revistas, portanto 1/3 das publicações concentram-se em um grupo de revistas. Revistas como "*Medical Research, Organs and Systems, Dentistry Oral surgery Medicine, Medical Research General Topics e Genetics and Molecular Biology*", pertencem a este grupo. O prestígio das fontes bibliográficas, onde se publicam os resultados das investigações científicas, representa uma medida da influência que podem exercer os artigos nelas publicados.

Considerações críticas sobre o nosso bom desempenho também devem ser levadas em conta. É necessário lembrar que esses estudos foram realizados em bases de dados que, apesar da abrangência, utilizam

critérios para indexar suas revistas que privilegiam os temas, prioridades e cultura de países e grupos líderes no desenvolvimento e na produção científica e tecnológica. Ao considerarmos o crescimento da ciência brasileira temos que incluir as publicações em congressos, livros e em revistas não indexadas. Não podemos excluir do computo dessa produção, temas prioritários de pesquisa que embora pertençam a temas científicos universais, estejam voltadas para problemas nacionais, tais como, as que dizem respeito as nossas doenças tropicais, endêmicas e nossa biodiversidade, mas que raramente encontram espaço nas revistas internacionais.

Há que se valorizar a bibliometria e seus indicadores, não só pelas possibilidades de análise e conhecimento do comportamento da ciência, mas por permitir a reflexão e a discussão induzidas por seus resultados.

Deve-se impedir, por meio dessas reflexões, que indicadores quando usados, por exemplo, para comparar diferentes países, possam induzir uma avaliação simplista de que a ciência no Brasil evolui bem e encontra-se bem assistida em suas políticas. A informação sobre o crescimento da produção científica brasileira e as comparações com outros países devem vir acompanhadas de outros indicadores de competência, no uso do conhecimento para gerar inovações, que é o crescimento do número de patentes.

Quando se trata de responder sobre o papel que a ciência e a tecnologia assumem no desenvolvimento de uma sociedade, os indicadores bibliométricos são importantes

ferramentas de análise. No entanto, as suas características e os critérios para sua construção, devem ser contemplados por diferentes olhares.

Um segundo grupo de análises foram realizados neste trabalho e trata da co-ocorrência de palavras ou "co-word analysis". Assim foram detectadas palavras que caracterizam o conteúdo dos trabalhos sobre um determinado tema, procedeu-se à contagem da aparição destas e posterior estudo de sua proximidade e significância. Os conceitos de proximidade ou distância foram representados graficamente e constituíram a base para a elaboração dos "*mapas tecnológicos*".

Na impossibilidade de apresentar a totalidade dos agrupamentos de palavras e todas suas inter-relações, as análises de co-ocorrências de palavras foram centradas em alguns exemplos significativos, de forma a evidenciar as possibilidades e potencialidades desse método para "mapear" diferentes áreas do conhecimento, sem, necessariamente, caracterizar o estado da arte da produção científica brasileira, na área da saúde, no período estudado.

As palavras que foram estudadas estão contidas no campo descritores da base de dados e foram relacionadas com outras variáveis como autores, identificadores, ano de publicação do artigo ou filiação da publicação. Os descritores encontrados somaram 39.455 diferentes palavras, que correspondem às palavras-chave presentes publicações registradas. Um ponto crítico para a qualidade dessas análises é a etapa de "**Tratamento**

Morfológico". Esta é uma função que permite a equiparação de palavras ou a criação de sinônimos para as palavras. Nesta etapa não puderam ser incorporados, ao software de análise, modelos lingüísticos, semânticos, que prevêm análises dos diferentes sentidos de uma palavra e sintáticos que relacionam as palavras por seu significado. Portanto, o tratamento dado aos descritores imprimiu um certo grau de subjetividade à criação dos sinônimos.

Definitivamente, a fase preparatória para proceder as análises de co-word, deve contar com um "thesaurus" que garanta aos documentalistas e pesquisadores a padronização da linguagem documentária, bem como, a análise de peritos da área objeto de estudo. Considero estes dois procedimentos complementares e a ausência de um ou outro prejudica a qualidade do estudo, principalmente em grandes volumes de informações.

Observamos que 72,03% dos descritores, aparecem uma única vez como descritor de assunto de um artigo. Apenas 27,97% dos 39.455 descritores, têm freqüência >2. No entanto, se tomarmos as palavras que se repetem até o mínimo 10 vezes, somente 1142 (2,89%) ocorrem numa freqüência igual ou maior que 10. (Lei de Zipf)

Apenas com a contagem de palavras pudemos obter informações interessantes, sobretudo se compararmos sua ocorrência no tempo para observarmos seu comportamento. Palavras referentes a tecnologias ou temas específicos que têm sua freqüência aumentada num período determinado demonstram que estão captando a atenção, que se está

patenteando sobre o assunto, enfim, que há um interesse de investigação sobre elas. Ao contrário, as que diminuem de frequência ao longo do tempo, podem representar temas que estão chegando ao limite de seu interesse. Por outro lado dependendo do enfoque dado ao estudo, patentes, por exemplo, as palavras de menor frequência e que apareceram nos últimos anos podem ser as mais importantes para as análises por estarem apontando para tecnologias ou produto novos.

As análises de co-ocorrência são as bases para a elaboração dos mapas tecnológicos e de possíveis definições de "clusters" ou associações. A análise, do cruzamento entre descritores e identificadores, por exemplo, resultou primeiramente em uma matriz, com resultados das contagens que permitiram organizar, depurar e transformar os dados através de operações matemáticas. São inúmeras as possibilidades de análises e uma vasta gama de resultados puderam ser obtidos com essas análises.

Em que pese os critérios e as limitações impostas a este estudo e as possíveis divergências entre uma ou outra métrica, que melhor se adequaria a este tipo de análise, a utilização da métrica euclidiana utilizada para estas análises privilegiou as palavras mais fortes e as co-ocorrências de maior afinidade.

Ao estudarmos os resultados das relações entre Descritores e Identificadores, temos que a palavra "ecology" co-ocorre com "ribeira valley, irrigation, mosquito, sao-paulo state, psychodidae, culicidae, hemorrhagic-disease, vector", entre outras e podemos

concluir, por exemplo, que esta linha de pesquisa está voltada para insetos vetores e que o grupo tem investigado algumas regiões do estado, ricas em material para seus estudos, onde as condições são favoráveis ao desenvolvimento desses insetos e à manutenção da transmissão das doenças por eles veiculados.

Para exemplificar estudos com autores, tomamos por referência as palavras lideradas pelo grupo "ecology". Foi realizada uma contagem de autores, com até 3 publicações no período, utilizando-se como filtro este grupo de palavras.

Após selecionar os autores pertencentes a esse grupo, o cruzamento entre autores resultou numa matriz que evidencia "clusters" bem delimitados. É possível consultar o conteúdo de cada "cluster" e dessa forma visualizar a que grupo se refere. Do "cluster" assinalado emergiu um grupo de autores, liderado pelo autor "Forattini".

Nos mapas tecnológicos, elaborados a partir dessas matrizes surgiram os autores que compõe o "cluster ecology". Ficou evidenciada nos mapas a interface entre os "clusters", como, por exemplo, os liderados por "Forattini" e Burattini, devido a eventuais colaborações.

Na análise de co-ocorrência dos campos de descritores e autores, pudemos verificar a proximidade de grupos de palavras descritoras de temas e seus respectivos autores. Resultados da relação dessas palavras com o autor "Forattini", por exemplo, mostra que em 100% de seus artigos publicados, no período estudado,

utilizou a palavra "ecology", que evidencia a linha de pesquisa desse grupo de autores.

A análise da dinâmica de colaboração do grupo, realizada em três períodos, 1990-1993, 1994-1997 e 1998-2002, mostra a rede de instituições parceiras do autor "Forattini" e que 10 membros do grupo já no período inicial, cooperam em menor ou maior intensidade em todos os 3 períodos. Este é um grupo consolidado, que está filiado, principalmente, à Faculdade de Saúde Pública, Faculdade de Medicina da USP e a SUCEN, mas que tem colaboração com diversas instituições nacionais e internacionais.

Tomando-se o autor "Forattini" como referência, o grupo publicou 52 artigos e recebeu 157 citações, uma média de 3,1 citações por artigo. Este grupo tem padrão de comportamento que revela alta média de citações por artigo publicado, impacto e visibilidade das suas publicações. Como é um campo estável e consolidado que talvez represente o padrão no campo das ciências da saúde.

As técnicas de "*co-word analysis*" demonstraram incrementar os resultados observados, com as técnicas bibliométricas, já tradicionalmente utilizadas, de maneira a poder qualificar as instituições, as disciplinas e os profissionais produtores do conhecimento técnico e científico, nas diferentes áreas que compõe este campo.

A representação visual de grandes volumes de dados proporciona em pouco tempo um panorama completo que

se recorda facilmente. Além disso, há uma redução da informação, o que é muito importante. Atualmente encontramos uma grande quantidade de "ruídos" na enorme quantidade de dados disponíveis. É crucial filtrar os dados significativos. Os mapas agregam os dados de uma forma que nenhum especialista é capaz de fazer, dada a impossibilidade de assimilar a quantidade de informação disponível.

Esta abordagem para a área de avaliação em Ciência e Tecnologia amplia as análises bibliométricas tradicionais e poderia incrementar as atividades de planejamento e definição de ações estratégicas indicando, para onde e como se poderiam dirigir esforços para desenvolvimento do campo estudado. Constitui uma importante ferramenta de gestão de Ciência e Tecnologia, no planejamento científico de qualquer instituição, instância de governo ou país, para conseguir uma rentabilidade máxima nos investimentos neste campo.

Consolidar no meio científico a utilização dessa qualidade de estudos quantitativos e qualitativos, pode apontar para a utilização das bases de dados, também como um instrumento para desenvolver atividades de análise, enfatizando as possibilidades que este tipo de avaliação pode possibilitar, para elevar o nível das informações, assim como para explorar as técnicas bibliométricas no gerenciamento de políticas e de tomadas de decisão.

8 - BIBLIOGRAFIA

- Almeida PIR, Klein L, Henrique JSJ. **Ciência e Tecnologia no Brasil. Ministério das Relações Exteriores.**
http://www.mre.gov.br/portugues/politica_externa/temas_agenda/ciencia_tecnologia/ciencia_tecbra.asp (acesso 7/12/04)
- Almind T, Ingwersen P. Informetric analyses on the World Wide Web: methodological approaches to webmetrics. **Journal of Documentation** 1997;53: 404-26.
- Arlington, VA. National Science Board, **Science & Engineering Indicators - 1998**, National Science Foundation.
- Basevi THMM. Tendências na aplicação de formatos, sistemas cooperativos e redes de intercâmbio: uma visão infométrica. **Departamento de Ciência da Informação e Documentação** 1993. Universidade de Brasília DF(Dissertação de Mestrado).
- Black N. A National Strategy for research and development: lessons from England. **Annu. Rev. Public Health** 1997;18:485-505.
- Bonalume Neto R. Desigualdade afeta produção científica. **Folha de São Paulo**, Ciência 15/07/2004.
- Bordons, M.; Zulueta M. A. *Evaluación de la actividad científica através de indicadores bibliométricos.* **Revista Española de Cardiología**, - Madri, v. 52, n. 10, p. 790-800, out. 1999. Disponível em:
http://www.revespcardiol.org/cgi-bin/wdbcgi.exe/cardio/cardioeng.mrevista_cardio.fulltext?pid=190 . (acesso 7 jun. 2004).
- Bruschini C, Ardaillon D, Ubehaum SG. Tesouro para estudos de gênero e sobre mulheres. **Fundação Carlos Chagas** São Paulo 2001.
- Callon M, Courtial JP et al. La Scientométrie, que sais-je?, **Presses Universitaires de France** 1993; Paris.
- Callon M, Courtial JP., Laville F. et al??. Co-word analysis as a tool for describing a network interactions between basic and technological research: the case of polymer chemistry. **Scientometrics** 1991; 22:155-205, 1991.

Callon M, Law J, Rip A. **Mapping the dynamics of science and technology: sociology of science in the real world.** Eds Basingstoke Hants: MacMillan 1986.

Características del Software *Tétralogie* Ver. 7, **Herramientas para la Inteligencia Tecnológica**, (mimeo), 1997.

Carneiro DM. **C&T em prol da cidadania**
<http://revista.fapemig.br/15/opinia.html> (acesso 20/07/04)

Cavalcanti CR. Indexação & tesauro metodologia & técnicas. **ABDF** 1978; Brasília.

Cavalcanti, Cordelia R. Indexação & Tesauro Metodologia & Técnicas. Brasília, **ABDF**, 1978.

CNPq -**Introdução ao Plano Institucional do CNPq**, novembro de 1998. <http://www.cnpq.br> (acesso 15/12/03)

Cole FJ, Eales NB. Part I: a Statistical Analysis of the Literature, Science Progress. In: **The History of Comparative Anatomy**. 1917.

Cruz E, Escorsa P, Maspons R, Guixé J et al. La Vigilancia Tecnológica en el Sector de los Nuevos Materiales. **Publicación del CYTED**, Ed. IALE Tecnologia, España 2003.

Documento Base da 2ª Conferência Nacional de Ciência, Tecnologia e Inovação em Saúde. **Situação Atual de Ciência e Tecnologia no Brasil** 2004; Brasília.

Donalísio MR. O Dengue no Espaço Habitado. **FUNCRAF - HUCITEC** 1999 São Paulo.

Dousset B. **Tétralogie** <http://atlas.irit.fr> (acesso 15/12/04)

El Estado de la Ciencia 2003 - <http://www.ricyt.edu.ar> - (acesso 15/01/05).

Escorsa P, Maspons R. La Ciencimetrica y sus Aplicaciones. In: eds? **Da vigilancia tecnológica a inteligencia competitiva**. Prentice Hall, Pearson Education, S.A; 2001.

Fundação Carlos Chagas - **O que é um Tesauro?** -
<http://www.fcc.org.br/pesquisa/rgenero/teg/tesauro.html> (acesso 10/08/04)

Gross PL, Gross EM. College Libraries and Chemical Education. **Science** 1927;66.

Guimarães R. Pesquisa no Brasil: A Reforma Tardia. **São Paulo Perspec** 2002; 16(4):41-47.

Gusmão, R. -**Indicadores FAPESP de CT&I: um diagnóstico atualizado dos esforços de Ciência, Tecnologia e Inovação no Estado de São Paulo**, apresentado em RICYT VI Taller de Indicadores de Ciencia y Tecnologia Iberoamericano e Interamericano Medir el conocimiento para la transformación social. Buenos Aires, 15-17 septiembre de 2004

Holm-Nielsen L, Crawford M, Saliba MA. Institutional and entrepreneurial leadership in the Brazilian science and technology Sector. **Setting a new agenda, World Bank Discussion Papers** 325.

Indicateurs de sciences et de technologies, <http://www.obs-ost.fr/fr/accueil.php> - (Acesso 15/01/05).

King D. The scientific impact of nations. **Nature** 2004; 430:. <http://www.nature.com/nature>.

King J. A review of bibliometric and other science indicators and the role in research evaluation. **Journal of Information Science** 1987; 13: 261-276.

Kostoff RN. Research impact quantification. **R&D Management** 1994; 24(3): 207-218.

Leta J, De Meeis L. A Profile of science in Brazil", **Scientometrics** 1996; 25(1): 33-44.

Leydesdorff L. The search of epistemic networks. **Social studies in Science** 1991; 21(1): 75-110.

Luukkonen-Gronow T. Scientific research evaluation: a review of methods and various contexts of their application. **R&D Management** 1987;17(3):207-214.

Marcovitch J et al. **Política e gestão em ciência e tecnologia**, Livraria Pioneira ed., São Paulo, 1986.

Marques F. Indicadores: uma prova de qualidade. **Boletim FAPESP** Ed. 10; 08/04.

Maspons R. La cienciometria como instrumento de evaluación, utilidade práctica de los indicadores. **VIII Seminário de Avaliação em C&T. Núcleo de Política e Gestão de Ciência e Tecnologia, FEA/USP** 22 a 26 de abril de 1996; 1-9.

Mugnaini, R. , Jannuzzi, P.M. , Quoniam, L. M. - Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal. - **Ci. Inf., Brasília**, v. 33, n. 2, p. 123-131, maio/ago. 2004.

NORMAS de VANCOUVER - Comissão Internacional de Editores de Revistas Médicas. Normas para apresentação de artigos propostos para publicação em revistas médicas. **Rev Port Clin Geral** 1997;14:159-74 - Tradução de "Uniform requirements for manuscripts submitted to biomedical journals".

Pellegrini Filho, A. e Goldbaum, M. - Production of scientific articles on health in six Latin American countries, 1973-1992. **Rev. Panam. Salud Publica** v.2 n.2 Washington Ago. 1997.

Pereira, J R C - **Análise de Dados Qualitativos - Estratégias Metodológicas para as Ciências da Saúde, Humanas e Sociais**. São Paulo, EDUSP -1999.

Pereira, J R C; Baião, M S; Fisher, A L Avaliação de Projetos em Institutos de Pesquisa - **Revista de Administração - USP** - 1996 Vol 31 No. 04, Pag. 77 a 92.

Pinto VB, Lima MC. Diálogos com os professores da UFC através das citações. In: **Encontros Universitários, Anais 2002** . Fortaleza, 04 a 06 de dezembro de 2002. (Anais Eletrônicos)
http://www.decigi.ufpr.br/anais_enebd/documentos/oral/ENEEDMIRIAN.doc.

Polanco X. Scientometrics analysis of the cognitive sciences in Pascal". **Inisit Info** 1993:7.

Pritchard A. Statistical bibliography or bibliometrics? **Journal of Publication** 1969; 25.

Quoniam L. Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil. **Ci. Inf. Brasília**, 2001; 30:2.

Robredo, J e da Cunha, MB - Aplicação de Técnicas Infométricas para identificar a abrangência do léxico básico que caracteriza os processos de indexação e recuperação da informação. **Ci. Inf., Brasília**, v.27, n.1, p.11-27, jan/abr 1998.

Rodrigues, T. **Quando os Indicadores de C&T deixam de indicar e passam a determinar**.
<http://www.comciencia.br/reportagens/2004/11/11.shtml> (acesso 15/01/05).

Rostaing H. La bibliométrie et ses techniques. **Sciences de la Société**. Toulouse y Centre de Recherche Rétrospective de Marseille (CRRM) 1995.

Rousseau, R. **Timeline of Bibliometrics**.
http://users.pandora.be/ronald.rousseau/html/timeline_of_bibliometrics.html (acesso 24/03/03).

Sancho R. Indicadores bibliométricos utilizados em la evaluación de la ciencia y la tecnología. Revisión Bibliográfica. **Revista Española de Documentación Científica** 1990; 13.

SPINAK, E. **Indicadores Cienciométricos**. Seminário de Avaliação da Produção Científica. Projeto SciELO, BIREME - São Paulo - 4 a 6 de março de 1998.

Tundisi JG. **Ciência no Brasil**.
http://www.mre.gov.br/portugues/politica_externa/temas_agenda/ciencia_tecnologia/ciencia_tec_bra.asp (acesso 10/08/04).

Turnock BJ. From measuring to improving public health practice. **Annu. Rev. Public Health** 1997; 18:261-82.

IBICT, Uma Radiografia da Ciência Brasileira - **Scientific American Brasil** (on-line) http://listas.ibict.br/pipermail/bib_virtual/2004-May/000159.html (acesso 07/07/04).

UNESCO - **Setor de Ciências e Meio Ambiente da UNESCO**
http://www.unesco.org.br/programas/meioamb/desenv_inst.asp (acesso 07/07/04).

Urbizagástegui R. A Lei de Lotka na bibliometria brasileira. **Ci Inf Brasília** 2002; 31(2): 14-20.

Van Raan AFJ. **Advanced bibliometric methods to assess research performance and scientific development: basic principles and recent practical applications**.ed;1993.

Vogt, C. e Gusmão, R. - **FAPESP lança busca de informações científicas**, www.fapesp.br/indicadores - (acesso 15/01/05).

Whitaker J. Creativity and conformity in science: titles, key-words and co-word analysis. **Social Studies in Science**;V.19, 1989:473-496.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)