



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE ESTUDOS SOCIAIS APLICADOS
DEPARTAMENTO DE BIBLIOTECONOMIA

**ANÁLISE COMPARATIVA ENTRE INDEXAÇÃO AUTOMÁTICA
E MANUAL DA LITERATURA BRASILEIRA DE
CIÊNCIA DA INFORMAÇÃO**

SIMONE BORGES BASTOS

**Dissertação de Mestrado apresentada
para obtenção do Título de Mestre em
Biblioteconomia e Documentação**

Professor Orientador: JAIME ROBREDO

BRASÍLIA

1984



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE ESTUDOS SOCIAIS APLICADOS
DEPARTAMENTO DE BIBLIOTECONOMIA

ANÁLISE COMPARATIVA ENTRE INDEXAÇÃO AUTOMÁTICA
E MANUAL DA LITERATURA BRASILEIRA DE
CIÊNCIA DA INFORMAÇÃO

SIMONE BORGES BASTOS

Dissertação de Mestrado apresentada
para obtenção do Título de Mestre em
Biblioteconomia e Documentação

Professor Orientador: JAIME ROBREDO

BRASÍLIA

1984

A meus pais e Luciano

AGRADECIMENTOS

Ao Professor Jaime Robredo, por sua dedicada orientação e pelo incentivo manifestado constantemente na vida acadêmica.

Aos Professores Murilo Bastos da Cunha, Kira Tarapanoff e em especial à Professora Cordélia Cavalcanti, pelas pertinentes sugestões apresentadas ao projeto de dissertação e pelo apoio durante o curso.

À Cecília, Lina e, especialmente, à Lena Vânia, pela colaboração prestada junto ao IBICT, o que permitiu concretizar esta dissertação.

Ao Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), pelo apoio concedido.

A Carlos Augusto S. Branco, pela elaboração das figuras.

À Tânia Mara de A. Campos, pelo tratamento estatístico dos dados.

A Adelce Pinto de Queiroz pela revisão datilográfica e de português.

À Vania Olguns do N. Martins, pela datilografia.

Ao pessoal da Secretaria do Departamento de Biblioteconomia, pela solicitude constante.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo suporte financeiro.

Finalmente, agradeço a todas as pessoas que, de maneira direta ou indireta, tornaram possível a realização deste trabalho.

" Algumas gerações nasceram para criar, outras para manter uma civilização. Nossa geração, lançada pela mudança histórica, foi compelida, por força das circunstâncias à criatividade. Se começarmos agora, nós e nossos filhos poderemos tomar parte da excitante reconstituição, não apenas de novas estruturas, mas da própria civilização. "

Alvin Toffler

RESUMO

Foi realizado um estudo comparativo da qualidade de indexação manual e automática, na área da Ciência da Informação, num conjunto de documentos em língua portuguesa, constituído por artigos publicados no periódico Ciência da Informação.

Na indexação automática utilizaram-se os títulos e os resumos e, na indexação manual os textos dos artigos.

A qualidade dos vocabulários foi avaliada em função do grau de coincidência dos termos em ambos os vocabulários, e de suas respectivas frequências de aparecimento.

A qualidade de recuperação da informação em linha, nas bases de dados, formadas com os mesmos registos dos referidos artigos indexados automaticamente e manualmente, foi medida através dos respectivos índices de precisão da recuperação.

A indexação automática apresenta menor redundância no vocabulário, e permite maior precisão na recuperação, especialmente quando se aplica a truncagem dos descritores.

ABSTRACT

Comparative analysis of automatic and manual indexing of the Brazilian literature on information science.

A comparative study of the quality of manual and automatic indexing was carried out, in the area of information science, on a population of documents written in Portuguese, integrated by the papers published in the Brazilian journal Ciência da Informação.

In the case of automatic indexing, the titles and the abstracts of the papers were considered, and in that of manual indexing the full text of the papers.

The evaluation of the quality of the vocabularies was made in function of the coincidence of the terms in both vocabularies, as well as in function of the respective frequencies of occurrence of the terms.

The quality of the on-line information retrieval, in the data bases integrated by the records of the same papers, indexed either manually or automatically, was established by using the precision measurements.

Automatic indexing leads to a lower redundancy in the vocabulary and a higher precision in the retrieval process, specially when roots or truncated descriptors are used.

LISTA DE SIGLAS

- BIB/DIALOGO - Sistema de Informações Bibliográficas em Modo de Dialogo
- CANDE - Command and Edit
- CIN - Centro de Informações Nucleares
- CINFORM - Base de dados bibliográfica em Ciência da Informação
- CLRU - Cambridge Language Research Unit
- CNRS - Centre National de la Recherche Scientifique
- DIRECT - Information Retrieval System of Philips Research Laboratories
- IBBD - Instituto Brasileiro de Bibliografia e Documentação
- IBICT - Instituto Brasileiro de Informação em Ciência e Tecnologia
- INDEXCI - Base de dados bibliográfica em Ciência da Informação indexada automaticamente
- INIS - International Nuclear Information System
- INSPEC - Information Service for Physics, Electrotechnology and Control
- KWIC - Key word in context
- KWOC - Key word out of context
- MAI - Machine Aided Indexing
- SLC - II - Language Translator Package
- SMART - Automatic Document Analysis and Retrieval System
- SPIRIT - Système Syntaxique et Probabiliste d'Indexation et de Recherche d'Information
- SPSS - Statistical Package for the Social Sciences
- UnB - Universidade de Brasília
- UNESCO - Organização das Nações Unidas para a Educação, Ciência e Cultura
- UNISIST - Sistema Mundial de Informação Científica

LISTA DE FIGURAS

- Figura 1 Processo de indexação manual
- Figura 2 Processo de indexação automática
- Figura 3 Índice KWIC dos títulos do periódico Ciência da Informação
- Figura 4 Palavras significativas eliminadas e palavras vazias não eliminados
- Figura 5 Referências bibliográficas
- Figura 6 Índice de assunto
- Figura 7 Índice de autores pessoais
- Figura 8 Índice de entidades
- Figura 9 Fragmentos da listagens com descritores e respectivas frequências (ordem alfabética e ordem numérica)
- Figura 10 Fragmento da listagem de associações dos descritores com frequência maior do que 3
- Figura 11 Registro indexado manualmente e registro hifenado preparado para indexação automática.
- Figura 12 Fragmento da tabela WORDFIXED
- Figura 13 Fragmento da tabela WORDROOTS
- Figura 14 Conceituação do subsistema de indexação automática a partir dos títulos e resumos

- Figura 15 Representação de estratégias de buscas e respectivos resultados
- Figura 16 Busca bibliográfica realizada na base de dados INDEXCI
- Figura 17 Busca bibliográfica realizada na base de dados CIFORM
- Figura 18 Descritores coincidentes e não-coincidentes entre as bases de dados INDEXCI e CIFORM e o índice KWIC
- Figura 19 Raízes e truncagens coincidentes e não coincidentes entre as bases de dados INDEXCI, CIFORM e o índice KWIC
- Figura 20 Descritores em função da frequência nas bases de dados INDEXCI, CIFORM e raízes e truncagens de INDEXCI
- Figura 21 Representação em escala semi-logaritmica da variação do número de ocorrências dos descritores em função do número de descritores identificados, nas bases de dados CIFORM e INDEXCI
- Figura 22 Índices de precisão e ruído apresentados nas bases de dados INDEXCI e CIFORM

LISTA DE TABELAS

Tabela 1	Descritores coincidentes
Tabela 2	Descritores não-coincidentes
Tabela 3	Descritores com frequência ≥ 4
Tabela 4	Raízes e truncamentos coincidentes
Tabela 5	Raízes e truncamentos não-coincidentes
Tabela 6	Lei de Bradford aplicada à INDEXCI
Tabela 7	Lei de Bradford aplicada à CIFORM
Tabela 8	Lei de Bradford aplicada à lista de raízes e palavras truncadas de INDEXCI
Tabela 9	Índices de precisão

LISTA DE ANEXOS

- Anexo 1 Descrição das funções dos programas utilizados
- Anexo 2 Descritores obtidos por indexação manual e automática
- Anexo 3 Raízes e palavras truncadas dos descritores
- Anexo 4 Buscas bibliográficas
- Anexo 5 Índices médios de precisão na base de dados INDEXCI
- Anexo 6 Índices médios de precisão na base de dados CIFORM

SUMÁRIO

1. INTRODUÇÃO	1
2. OBJETIVOS E JUSTIFICATIVA	3
2.1. Objetivo geral	3
2.1.1. Objetivos específicos	4
2.2. Justificativa	4
3. REVISÃO DA LITERATURA	7
3.1. Fundamentos gerais de indexação	7
3.2. Revisão da literatura estrangeira	15
3.2.1. Estudos comparativos entre indexação automática e manual	15
3.2.2. Estudos comparativos entre o título e outras fontes para indexação automática	18
3.2.3. Métodos de indexação automática	22
3.2.3.1. Método de frequência ou análise estatística	23
3.2.3.2. Método de atribuição de peso	26
3.2.3.3. Método probabilístico	28
3.2.3.4. Método matemático	29
3.2.3.5. Análise de "cluster"	30
3.2.3.6. Método de associação entre palavras	30
3.2.3.7. Experimentos em avaliação de recuperação da informação em linguagens de indexação	31
3.3. Revisão da literatura nacional	34
4. HIPÓTESES	39
4.1. Variáveis	40
4.2. Definições operacionais	40
4.2.1. Conceitos expressos nas hipóteses	40
4.2.2. Conceitos implicitamente relacionados com as hipóteses	41
5. METODOLOGIA	42
5.1. Descrição e registro dos documentos	44
5.2. Captura dos dados e validação dos registros	46
5.3. Processamento do arquivo de dados indexado manualmente	46
5.4. Formatação da base de dados a partir dos registros indexados manualmente	55
5.5. Preparação do arquivo de dados indexado automaticamente	55

5.6. Processamento do arquivo de dados indexado automaticamente	61
5.7. Formatação da base de dados a partir dos registros indexados automaticamente	61
5.8. Escolha das perguntas e montagem das estratégias de busca	61
5.9. Interrogação das duas bases de dados	66
5.10. Tabulação e análise dos resultados	66
5.11. Limitações da pesquisa	70
6. APRESENTAÇÃO DOS RESULTADOS	71
6.1. Resultados dos estudos comparativos entre os vocabulários obtidos por indexação automática e manual	71
6.2. Resultados dos estudos comparativos entre a qualidade de recuperação em linha das bases de dados INDEXCI e CIFORM	78
7. INTERPRETAÇÃO DOS RESULTADOS	83
7.1. Análise comparativa entre os vocabulários obtidos por indexação automática e manual	83
7.2. Análise comparativa da qualidade de recuperação em linha entre a base de dados indexada automaticamente e manualmente	93
8. CONCLUSÃO	99
8.1. Sugestões para estudos futuros	102
9. REFERÊNCIAS BIBLIOGRÁFICAS	103
10. ANEXOS	123
10.1. Anexo 1	124
10.2. Anexo 2	130
10.3. Anexo 3	155
10.4. Anexo 4	192
10.5. Anexo 5	201
10.6. Anexo 6	203

SUMÁRIO

1. INTRODUÇÃO	1
2. OBJETIVOS E JUSTIFICATIVA	3
2.1. Objetivo geral	3
2.1.1. Objetivos específicos	4
2.2. Justificativa	4
3. REVISÃO DA LITERATURA	7
3.1. Fundamentos gerais de indexação	7
3.2. Revisão da literatura estrangeira	15
3.2.1. Estudos comparativos entre indexação automática e manual	15
3.2.2. Estudos comparativos entre o título e outras fontes para indexação automática	18
3.2.3. Métodos de indexação automática	22
3.2.3.1. Método de frequência ou análise estatística	23
3.2.3.2. Método de atribuição de peso	26
3.2.3.3. Método probabilístico	28
3.2.3.4. Método matemático	29
3.2.3.5. Análise de "cluster"	30
3.2.3.6. Método de associação entre palavras	30
3.2.3.7. Experimentos em avaliação de recuperação da informação em linguagens de indexação	31
3.3. Revisão da literatura nacional	34
4. HIPÓTESES	39
4.1. Variáveis	40
4.2. Definições operacionais	40
4.2.1. Conceitos expressos nas hipóteses	40
4.2.2. Conceitos implicitamente relacionados com as hipóteses	41
5. METODOLOGIA	42
5.1. Descrição e registro dos documentos	44
5.2. Captura dos dados e validação dos registros	46
5.3. Processamento do arquivo de dados indexado manualmente	46
5.4. Formatação da base de dados a partir dos registros indexados manualmente	55
5.5. Preparação do arquivo de dados indexado automaticamente	55

5.6. Processamento do arquivo de dados indexado automaticamente	61
5.7. Formatação da base de dados a partir dos registros indexados automaticamente	61
5.8. Escolha das perguntas e montagem das estratégias de busca	61
5.9. Interrogação das duas bases de dados	66
5.10. Tabulação e análise dos resultados	66
5.11. Limitações da pesquisa	70
6. APRESENTAÇÃO DOS RESULTADOS	71
6.1. Resultados dos estudos comparativos entre os vocabulários obtidos por indexação automática e manual	71
6.2. Resultados dos estudos comparativos entre a qualidade de recuperação em linha das bases de dados INDEXCI e CIFORM	78
7. INTERPRETAÇÃO DOS RESULTADOS	83
7.1. Análise comparativa entre os vocabulários obtidos por indexação automática e manual	83
7.2. Análise comparativa da qualidade de recuperação em linha entre a base de dados indexada automaticamente e manualmente	93
8. CONCLUSÃO	99
8.1. Sugestões para estudos futuros	102
9. REFERÊNCIAS BIBLIOGRÁFICAS	103
10. ANEXOS	123
10.1. Anexo 1	124
10.2. Anexo 2	130
10.3. Anexo 3	155
10.4. Anexo 4	192
10.5. Anexo 5	201
10.6. Anexo 6	203

1. INTRODUÇÃO

A informação é considerada como um produto tão valioso quanto os recursos minerais e energéticos em um país. Para se medir o desenvolvimento de uma nação, basta verificar a quantidade e qualidade de informações que seus habitantes geram e consomem.

A informação é o conhecimento circulante, e a sua transferência é realizada por canais informais ou por comunicação pessoal e canais formais ou documentários. Os canais formais são compostos por documentos primários, que posteriormente darão origem a documentos secundários e terciários, estes últimos normalmente elaborados por organismos de informação, numa tentativa de controlar o grande volume de documentação produzida. De acordo com WEISMAN(123), um profissional que desejar manter-se atualizado levará 50 anos, lendo 24 horas por dia, para ler documentos publicados em um ano.

Devido a este problema, canais intermediários surgiram com o objetivo de tornar acessíveis informações adequadas às necessidades de cada um, evitando, em muitos casos, duplicações de pesquisas e investimentos inadequados na geração de novos conhecimentos.

Esses canais são, entre outros: bibliotecas, sistemas de informação, centros de tradução, centros referenciais e bases de dados. Dentre estes, os que mais comercializam seus produtos são as bases de dados. Pode-se dizer que elas são as responsáveis, em parte, pela transferência de ciência e tecnologia para muitos países.

A tarefa de tornar acessíveis as informações relevan-

tes requer uma série de atividades que compõem o que se denomina o ciclo documentário. Essas atividades compreendem, basicamente: seleção, aquisição, registro, descrição, física, análise de conteúdo, armazenamento, recuperação e disseminação da informação.

Dentre as diversas formas de análise de conteúdo, a indexação é a técnica que oferece uma melhor condensação do assunto do documento, através da atribuição de descritores, possibilitando aumentar a capacidade de armazenamento e o desempenho da recuperação.

A indexação é uma das operações significantes que compõe o ciclo documentário. Pode-se dizer que é uma atividade meio, que possibilita ao usuário o acesso adequado ao conteúdo dos documentos. É considerada um dos pontos de estrangulamento dos sistemas de informação.

A indexação pode ser realizada pelo homem (indexação manual) ou por programas de computador (indexação automática).

A aplicabilidade de uma técnica ou outra foi testada através de vários experimentos, em diversas áreas do conhecimento e em várias línguas estrangeiras. Ambas as técnicas foram consideradas eficientes. Em alguns casos há maior aceitação da indexação automática, em outros da manual, dependendo das características dos sistemas de informação, das línguas e das áreas do conhecimento em que foram aplicadas.

Estudos experimentais, em língua portuguesa e na área de Ciência da Informação, para testar comparativamente a aplicabilidade e desempenho da indexação automática em relação à manual não foram, até o momento, como demonstra a revisão de

literatura nacional, realizados no Brasil.

A eficiência do título e do resumo, como recursos informacionais para a indexação é um aspecto que foi amplamente discutido em diversos países. Na literatura estrangeira observa-se a presença de vários trabalhos comparativos para verificar a validade do título e do resumo, como elementos adequados a descrever o conteúdo dos documentos, em relação a descritores manuais, ao texto integral e outras unidades informacionais existentes nos documentos. Este tipo de estudo comparativo, em língua portuguesa, aplicado à área de Ciência da Informação e utilizando métodos estatísticos, ainda não foi desenvolvido.

Pretende-se analisar a qualidade da indexação automática em relação à manual, na área específica de Ciência da Informação, e para a língua portuguesa, aplicando-se metodologia que compara as duas linguagens documentárias e testa a capacidade de recuperação de ambas.

A área de Ciência da Informação, neste trabalho, refere-se a estudos sobre fundamentos gerais da informação, processos de comunicação da informação, sistemas e organismos da informação.

2. OBJETIVOS E JUSTIFICATIVA

2.1. Objetivo geral

Demonstrar que a indexação automática em títulos e resumos permite indexar, eficientemente, os documentos que irão constituir uma base de dados na área de Ciência da Informação.

2.1.1. Objetivos específicos

2.1.1.1. Formar uma pequena base de dados, na área de Ciência da Informação, com referências bibliográficas armazenadas e recuperáveis por computador.

2.1.1.2. Testar e analisar comparativamente a qualidade do vocabulário da indexação automática em títulos e resumos, e da indexação manual em texto integral.

2.1.1.3. Testar e analisar comparativamente a recuperação obtida com documentos indexados manual e automaticamente.

2.2. Justificativa

O enorme volume de documentos produzidos em todas as áreas do conhecimento torna cada dia mais difícil a tarefa de analisar e indexar os referidos documentos. Tal fato ocorre, principalmente, quando se deseja preparar, em prazos os mais curtos possíveis, bases de dados nacionais passíveis de interrogação via computador. Isto se torna um pouco mais grave quando se verifica que a necessidade de implantação de bases de dados nacionais, nas diversas áreas do conhecimento, deveria ser considerada uma das prioridades brasileiras na área de informação. Como profissionais desta área, deveríamos ser os primeiros a desenvolver técnicas que facilitassem e aprimorassem a criação dessas bases.

A escassez de pessoas especializadas em indexar rapidamente e com eficiência os documentos é uma realidade brasileira. A formação de profissionais nesta área não é muito adequada. Só a partir do novo currículo mínimo, aprovado em 1982,

É que os cursos de Biblioteconomia passaram a incluir, de uma maneira mais acentuada, a indexação como uma das técnicas a serem ministradas. Mesmo assim não se pode acreditar que, em áreas muito especializadas, como por exemplo a de energia nuclear, os bibliotecários possam indexar com perfeição, pois a formação acadêmica que receberam está, tradicionalmente, voltada mais para a área de ciências sociais.

A indexação manual é uma tarefa que requer conhecimento do assunto do documento, consistência técnica e desenvolvimento de linguagem de indexação própria a cada sistema de informação. É uma técnica que exige do profissional um tempo de dedicação razoável por documento. Todos os aspectos apresentados tornam, por sua vez, a tarefa de análise manual cada dia mais dispendiosa.

Assim, a indexação automática em títulos e resumos apresenta-se como uma opção rápida, eficiente e com baixo custo, a longo prazo, para analisar o conteúdo dos documentos.

É uma técnica que prescinde da presença do homem para a realização intelectual da atividade. O computador, programado especialmente, e alimentado com os títulos e resumos a serem indexados, pode efetuar a indexação em espaço de tempo bem inferior, além de permitir maior consistência.

Este estudo irá fornecer diretrizes que avaliarão a eficiência da indexação automática em títulos e resumos em relação à indexação manual, em língua portuguesa e na área de Ciência da Informação. Fornecerá insumos para se elaborar um vocabulário, baseado nos próprios termos dos especialistas, acompanhado da frequência respectiva de cada descritor, que poderá ser utilizado como um instrumento preliminar para a construção de

um tesouro na área de Ciência da Informação.

Será criada, também, uma base de dados, interrogável, com documentos do periódico Ciência da Informação, em sua maioria em língua portuguesa. Isto possibilitará a elaboração automática de um índice cumulativo da revista de 1972 a 1983.

3. REVISÃO DA LITERATURA

A revisão da literatura limita-se a abordar aspectos relacionados com a pesquisa a ser desenvolvida. Não aborda, dentro do assunto indexação automática, estudos teóricos, matemáticos, lingüísticos e históricos. Preocupa-se em descrever, principalmente, estudos estatísticos.

A revisão da literatura nacional, ao contrário da estrangeira, pretendeu-se ser exaustiva, pois o volume de documentos brasileiros existentes sobre indexação automática é significativamente menor do que os encontrados na literatura estrangeira.

Nesta revisão descrevem-se as várias pesquisas e seus resultados sobre análise comparativa entre indexação automática e manual, análise comparativa entre o uso do título, resumo e outras fontes para indexação, os vários métodos estatísticos de indexação automática, pesquisas em avaliação da recuperação da informação em linguagem de indexação, a evolução da técnica de indexação automática no Brasil e estudos e pesquisas nacionais nesta área ou relacionados a ela.

3.1. Fundamentos gerais de indexação

O Sistema Mundial de Informação Científica (UNISIST), em um de seus grupos de estudo, elaborou em um documento os princípios de indexação (119). Esses princípios estão voltados especificamente para a indexação manual e pode-se dizer que são os mesmos adotados por vários autores brasileiros (21, 22, 23) e estrangeiros (14, 34, 36, 48, 49, 51). Esse documento foi a primeira tentativa internacional de se normalizar o proces-

so de indexação.

De acordo com o UNISIST (119), a indexação é a operação que descreve e identifica o conteúdo de um documento, através de termos. Os conceitos dos documentos podem ser representados por termos selecionados através da linguagem natural ou por símbolos.

A indexação está diretamente relacionada à descrição física do documento e ambos constituem um registro bibliográfico, proporcionando ao usuário informações físicas e de conteúdo do documento. Os dados são organizados da forma mais acessível para a recuperação da informação.

A indexação pode ser realizada em documentos, no todo ou em partes, e na pergunta a um sistema de informação.

Durante a indexação manual os conceitos são extraídos por um processo de análise intelectual, que compreende basicamente três fases:

1. compreensão do conteúdo do documento através da leitura completa do texto ou do título, resumo e outras partes que compõem um documento. O UNISIST (119) não recomenda o uso somente do título e do resumo para indexar, pois nem sempre eles contêm os termos que identificam o conteúdo;

2. identificação dos conceitos estabelecendo o ambiente lógico e as diferenças entre os fenômenos, processos, as propriedades, as operações, os equipamentos e outras que venham a surgir;

3. seleção dos conceitos, observando alguns fatores como: exaustividade, especificidade e consistência.

O processo de indexação da pergunta requer do profissional de informação a tradução da pergunta do usuário para

a linguagem de indexação adotada. Normalmente se utilizam como instrumentos auxiliares tesouros, lista de cabeçalho de assunto, lista de termos controlados e outros.

Nesse processo é essencial a interação usuário-indexador-sistema.

Alguns sistemas utilizam, ainda, pesos e elos na recuperação e indexação.

Essas fases podem ser melhor visualizadas na figura 1.

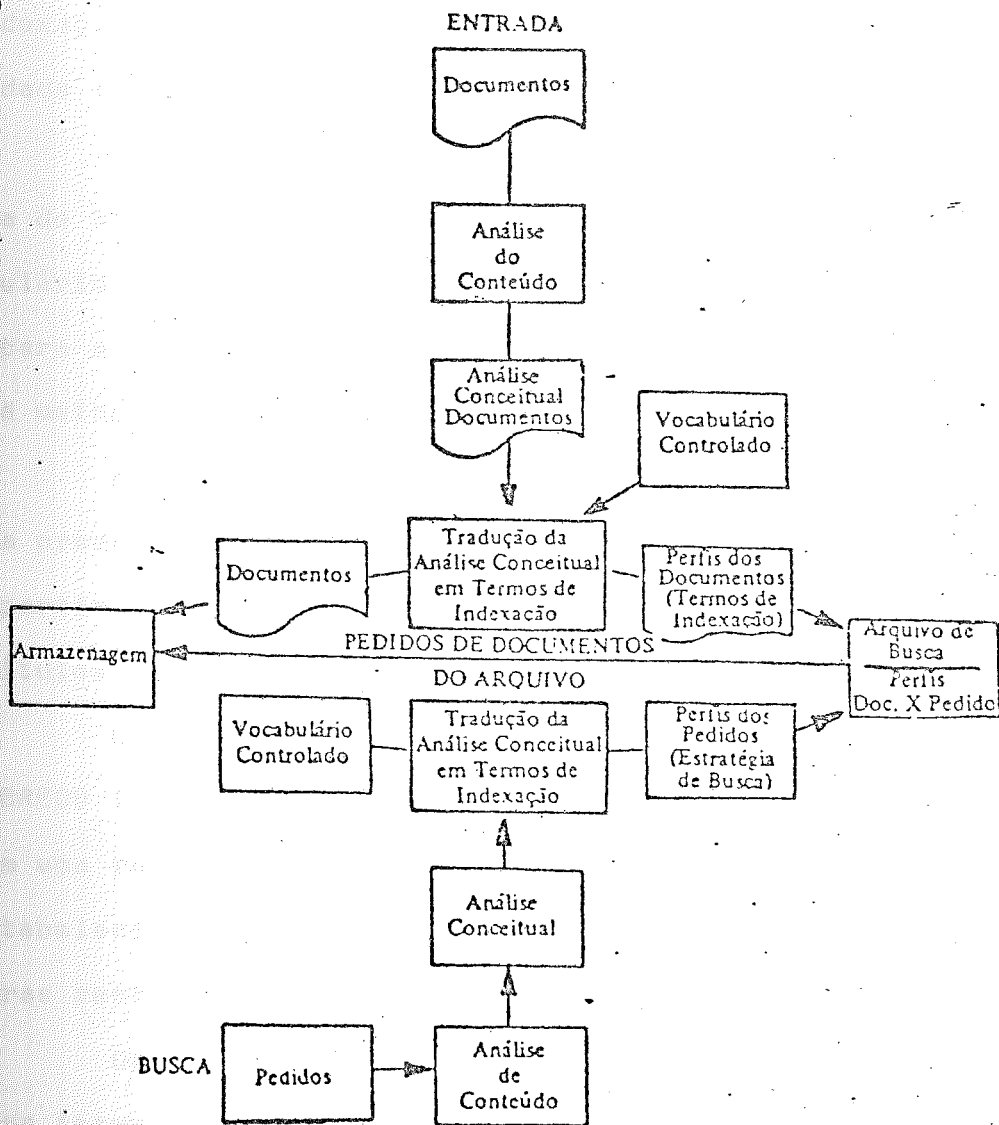


Figura 1 - Processo de indexação manual, segundo LANCASTER apud CESARINO & PINTO (23, p. 33)

A consistência, pode-se dizer, é o ponto de estrangulamento da indexação manual. Segundo PINHEIRO (74), ela reflete similaridades e diferenças de termos, atribuídos pela subjetividade dos indexadores na análise de conteúdo. Também está diretamente relacionada, quanto aos indexadores, no que se refere ao seu desempenho, à regularidade e imparcialidade de seleção dos conceitos, à experiência prévia da técnica, ao conhecimento do assunto dos documentos, a submissão às diretrizes da indexação adotadas pelo sistema e, quanto aos instrumentos de indexação utilizados pelos sistemas de informação, à sua qualidade (23,48,60).

O início da indexação automática deu-se ao final da década de 50, quando LUHN (62,63) desenvolveu a idéia de que o vocabulário existente em um documento deveria se constituir na base para a análise de conteúdo do referido documento, sendo esta a melhor maneira de recuperá-lo.

A primeira aplicação realizada por LUHN (62,63) foi para a produção do índice Key Word in Context (KWIC), elaborado a partir da rotação automática das palavras significativas dos títulos.

O processo de indexação automática baseia-se, segundo ROBREDO (86, p.238), "na comparação de cada palavra do texto com uma relação de palavras vazias de significado, previamente estabelecidas, que conduz, por eliminação, a considerar as palavras restantes do texto como palavras significativas".

De acordo com ROBREDO (86), esse processo pode identificar termos, pares de termos ou até frases significativas que expressem o conteúdo do documento, e pode-se dizer que é semelhante ao processo de leitura-memorização.

O processo de indexação automática encontra-se melhor visualizado e detalhado na figura 2.

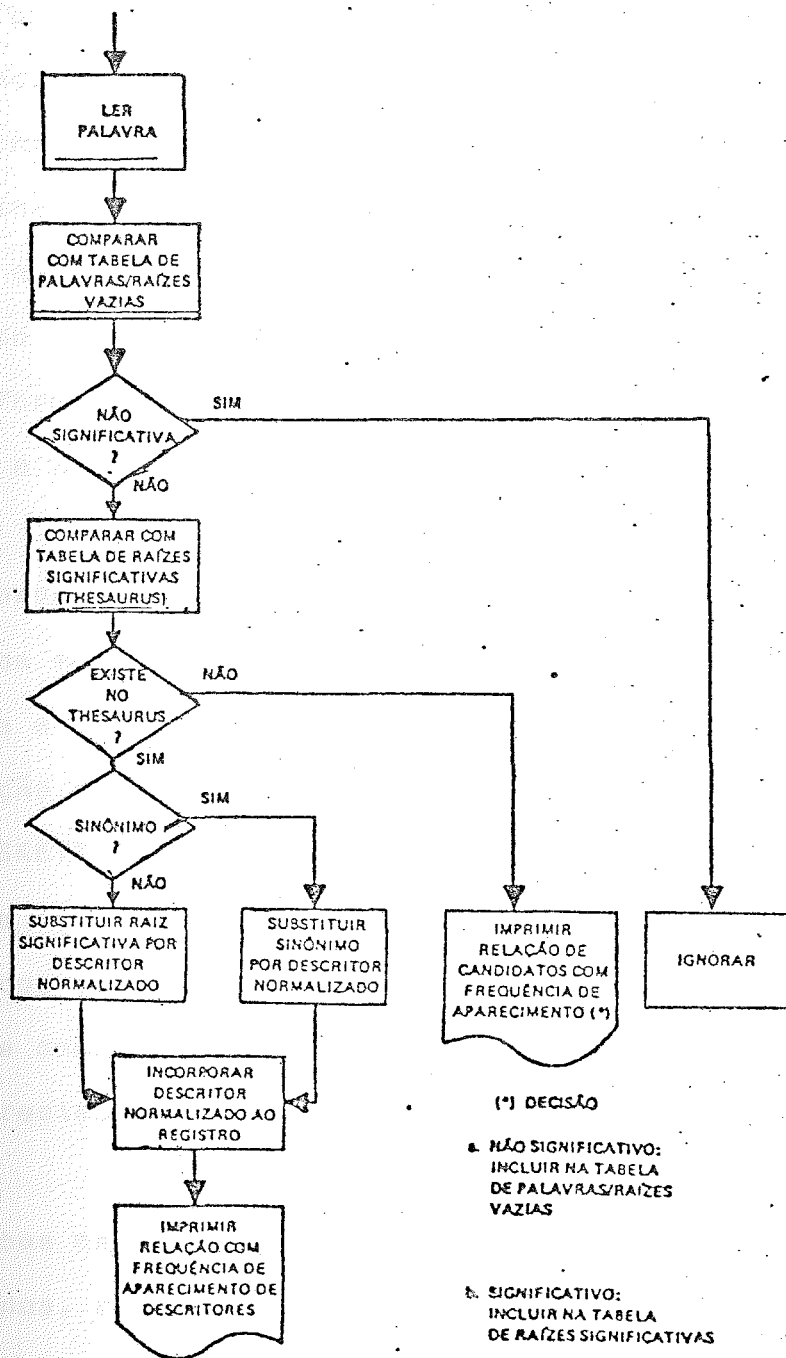


Figura 2 - Processo de indexação automática,

segundo ROBREDO (86, p. 247).

A indexação automática é um processo que pode utilizar métodos diferentes (ver item 3.2.3.), desenvolvidos para programas de computador. Essa operação, ainda segundo ROBREDO (86), é objetiva, pois utiliza sempre os mesmos programas para extração de termos significativos dos documentos. O aspecto de inconsistência devido à subjetividade da indexação manual é eliminado, possibilitando uma melhor recuperação.

Os aspectos de especificidade e exaustividade estão diretamente relacionados com as características das necessidades de informação delineadas pelos usuários dos sistemas de informação e pela decisão administrativa dos referidos sistemas.

O processo de indexação automática da pergunta é realizado pelos mesmos programas, assegurando assim a compatibilidade entre a linguagem utilizada na indexação e a utilizada na formulação da pergunta (83).

FOSKETT (34) define a exaustividade como sendo a extensão com que se analisa um documento, a fim de se estabelecer exatamente todos os assuntos que esse documento referencia, e a especificidade como a extensão em que um sistema de informação permite ser preciso ao se especificar o assunto de um documento.

Ambos os fatores, exaustividade e especificidade estão relacionados, respectivamente, à revocação e relevância, que são as medidas de qualidade da recuperação da informação.

A relevância ou precisão é definida por SARACEVIC (103) como a medida de contato efetivo entre a fonte e o destinatário. Pode ser quantificada, tal como a revocação, através das seguintes fórmulas matemáticas, mencionadas por ROBREDO (83,p.9):

Relevância = $\frac{a}{b}$, sendo que:

a = número de documentos pertinentes selecionados;

b = número total de documentos selecionados;

Revocação = $\frac{a}{a + c}$, sendo que:

a = número de documentos pertinentes selecionados;

c = número de documentos não pertinentes selecionados.

VICKERY (122, p. 77) menciona as seguintes funções das linguagens de indexação:

a) recuperar documentos com conteúdo semelhante;

b) recuperar documentos relevantes sobre um assunto específico;

c) recuperar documentos por grandes áreas de assunto;

d) possibilitar a conversão dos termos de indexação entre diferentes linguagens; e

e) auxiliar na escolha do termo adequado para a estratégia de busca.

CESARIANO & PINTO (23), posteriormente, abordaram ainda os seguintes aspectos:

a) representar o assunto de uma maneira consistente; e

b) permitir a compatibilidade entre a linguagem do indexador e a do pesquisador.

Segundo ROBREDO (83, p. 73, 79), a indexação pode ser realizada basicamente em três níveis, partindo do mais

O processo de indexação automática encontra-se melhor visualizado e detalhado na figura 2.

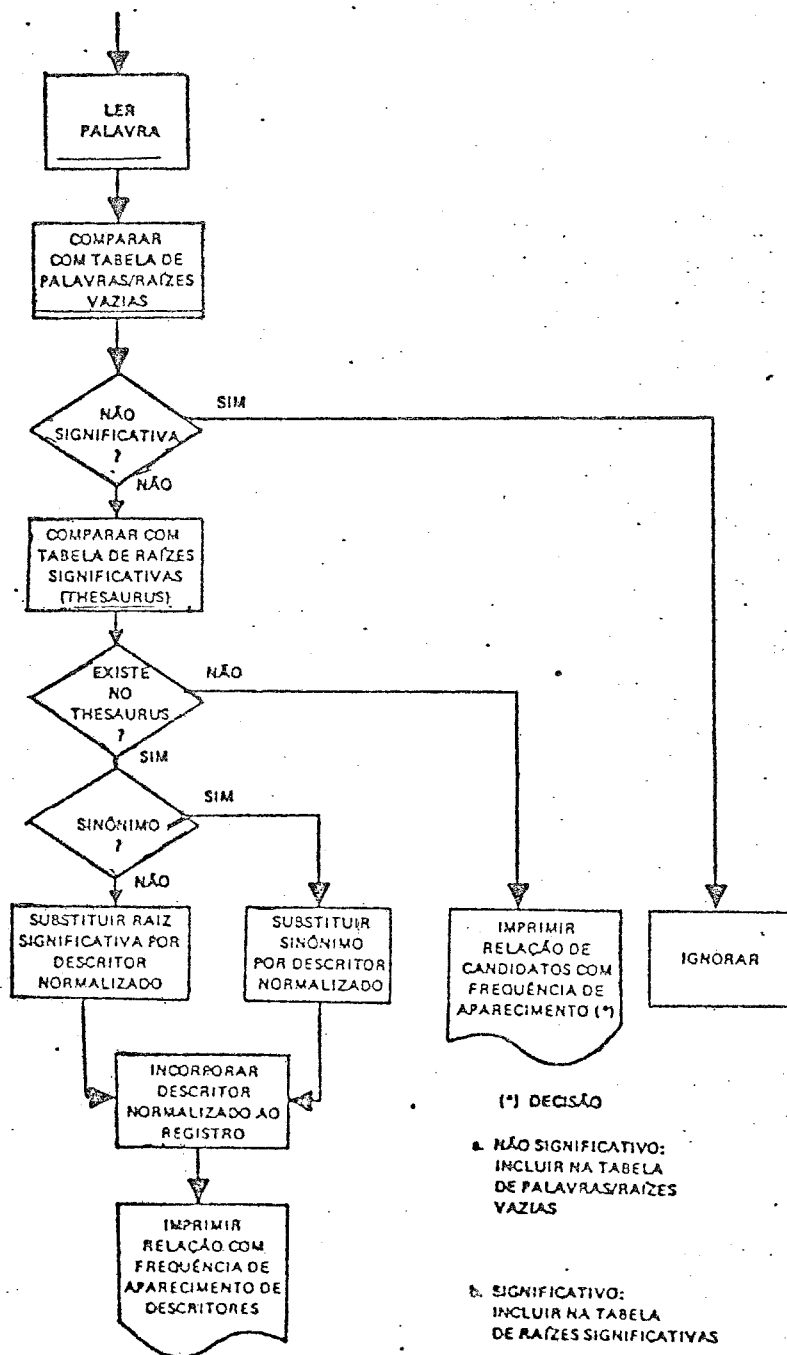


Figura 2 - Processo de indexação automática, segundo ROBREDO (86, p. 247).

A indexação automática é um processo que pode utilizar métodos diferentes (ver item 3.2.3.), desenvolvidos para programas de computador. Essa operação, ainda segundo ROBREDO (86), é objetiva, pois utiliza sempre os mesmos programas para extração de termos significativos dos documentos. O aspecto de inconsistência devido à subjetividade da indexação manual é eliminado, possibilitando uma melhor recuperação.

Os aspectos de especificidade e exaustividade estão diretamente relacionados com as características das necessidades de informação delineadas pelos usuários dos sistemas de informação e pela decisão administrativa dos referidos sistemas.

O processo de indexação automática da pergunta é realizado pelos mesmos programas, assegurando assim a compatibilidade entre a linguagem utilizada na indexação e a utilizada na formulação da pergunta (83).

FOSKETT (34) define a exaustividade como sendo a extensão com que se analisa um documento, a fim de se estabelecer exatamente todos os assuntos que esse documento referencia, e a especificidade como a extensão em que um sistema de informação permite ser preciso ao se especificar o assunto de um documento.

Ambos os fatores, exaustividade e especificidade estão relacionados, respectivamente, à revocação e relevância, que são as medidas de qualidade da recuperação da informação.

A relevância ou precisão é definida por SARACEVIC (103) como a medida de contato efetivo entre a fonte e o destinatário. Pode ser quantificada, tal como a revocação, através das seguintes fórmulas matemáticas, mencionadas por ROBREDO (83,p.9):

Relevância= $\frac{a}{b}$, sendo que:

a= número de documentos pertinentes selecionados;

b= número total de documentos selecionados;

Revocação= $\frac{a}{a + c}$, sendo que:

a= número de documentos pertinentes selecionados;

c= número de documentos não pertinentes selecionados.

VICKERY (122, p. 77) menciona as seguintes funções das linguagens de indexação:

a) recuperar documentos com conteúdo semelhante;

b) recuperar documentos relevantes sobre um assunto específico;

c) recuperar documentos por grandes áreas de assunto;

d) possibilitar a conversão dos termos de indexação entre diferentes linguagens; e

e) auxiliar na escolha do termo adequado para a estratégia de busca.

CESARIANO & PINTO (23), posteriormente, abordaram ainda os seguintes aspectos:

a) representar o assunto de uma maneira consistente; e

b) permitir a compatibilidade entre a linguagem do indexador e a do pesquisador.

Segundo ROBREDO (83, p.73, 79), a indexação pode ser realizada basicamente em três níveis, partindo do mais

geral para o específico:

- a) categorização - representa o assunto que predomina;
- b) superficial - representa os conceitos principais de forma geral;
- c) profunda - representa todos os conceitos fundamentais

Os termos de indexação podem ser expressos através de linguagem:

- a) natural ou livre, utilizando os mesmos termos do autor;
- b) controlada, adotando termos aceitos e definidos previamente;
- c) codificada, utilizando códigos previamente estabelecidos para expressar os conceitos significativos; e
- d) coordenada - as relações lógicas entre os termos, quando existentes, estabelecem-se através de:

- . equivalência ou sinonímia entre os termos;
- . subordinação ou hierarquia - parte do genérico para o específico, e vice-versa; e
- . coordenação ou associação - os conceitos estão relacionados à idéia de outro conceito.

Os termos podem, de acordo com VICKERY (122,p.72), estar inter-relacionados por subgrupo de assunto, eles, pesos ou expressos em pequenas frases da linguagem natural.

3.2. Revisão da Literatura Estrangeira

3.2.1. Estudos comparativos entre indexação automática e manual

A comparação entre os dois tipos de indexação, automática e manual, é realizada para se verificar as diferenças e semelhanças entre os termos selecionados por programas de um computador e pelo homem. De acordo com os resultados obtidos avalia-se a aplicabilidade de uma ou outra técnica.

Os testes de comparação podem ser divididos em testes de qualidade de indexação e de qualidade de recuperação.

A grande maioria de testes comparativos entre descritores atribuídos manual e automaticamente, segundo SALTON (97,98), chega a um resultado aproximado de 60% de compatibilidade entre uma linguagem e outra.

SALTON (96, p. 336), em um dos seus artigos, descreve uma fórmula matemática pela qual se obtém o "coeficiente de avaliação" entre dois vocabulários: $q = \frac{c}{a + m - c}$

q - valor comparativo entre dois vocabulários,

c - número de termos comuns,

a - número de termos atribuídos automaticamente,

m - número de termos atribuídos manualmente.

CARROL & ROELOFFS (20) realizaram estudos comparativos entre indexação manual e automática na área de Ciência da Informação, aplicando a análise de correlação estatística. Verificaram que os termos obtidos pelos indexadores foram se-

melhantes aos da indexação automática, mas levando-se em conta os custos de contratação, treinamento de mão-de-obra especializada e a inconsistência humana, a indexação automática é mais viável.

Vários testes comparativos de revocação e precisão foram realizados para verificar o desempenho da recuperação, através de termos atribuídos manual e automaticamente.

SALTON (97) e BOYCE & LOCKARD (16) realizaram suas experiências na área médica. O primeiro comparou os mesmos documentos indexados manualmente, utilizando vocabulário controlado, e indexados automaticamente, pelo programa SMART, utilizando termos livres do resumo. SALTON (97) verificou que, numa indexação automática somente com truncagem de palavras, a indexação manual torna-se mais efetiva cerca de 15% a 20%. Quando se utiliza um controle através de tesauros e dicionários, a eficiência da indexação automática é semelhante à da manual. Aplicando-se no momento da recuperação a técnica de realimentação de relevância na pergunta, a indexação manual é menos eficiente.

BOYCE & LOCKARD (16) aplicaram dois tipos de indexação manual e um tipo de indexação automática, em textos integrais. Seus resultados demonstram que a indexação automática foi mais consistente na revocação com perguntas gerais e específicas. A indexação manual obteve melhores resultados de precisão com perguntas utilizando termos específicos, e a automática com termos gerais. A indexação automática é tão eficiente quanto a manual, concluíram ao final.

VAN DER MEULEN & JANSSEN (120) avaliaram comparativamente a indexação automática do programa DIRECT, que uti-

liza títulos e resumos, e a indexação manual desenvolvida pelo INSPEC. Criaram duas bases de dados, com os mesmos documentos, indexadas pelas duas técnicas, e realizaram duas perguntas-teste. Verificaram, ao contrário dos resultados mencionados anteriormente, que a indexação manual apresenta melhores índices de revocação e precisão, cerca de 20% em relação à automática. Os autores justificaram esse resultado devido ao pequeno número de perguntas-teste efetuadas.

KLINGBIEL & RINKER (53) compararam a eficiência da indexação manual e da automática realizadas em títulos e resumos pelo programa Machine-Aided Indexing (MAI). A indexação manual e a automática obtiveram os mesmos índices de revocação, mas a manual mostrou-se inferior na precisão.

BARNES, COSTANTINI & PERSCHKE (4) compararam a indexação manual e a automática em títulos e resumos do sistema SLC-II. O teste foi realizado em 5.000 documentos do INIS Atomindex. A indexação automática mostrou-se mais eficiente na recuperação do que a manual, e os termos existentes em uma estavam compatíveis com os termos da outra. A revocação da indexação automática apresentou um índice de 90%.

A maioria dos testes revela que a indexação automática produz resultados de recuperação, no mínimo, equivalentes aos obtidos pela manual. Nos casos em que são aplicadas técnicas mais sofisticadas, na recuperação, a indexação automática mostra-se, segundo SALTON (97,98), ainda mais eficiente.

LANCASTER (58) aponta alguns dos problemas relacionados à indexação manual:

- . falta de especificidade ou coordenação falsa entre termos no vocabulário;

- . perguntas muito exaustivas ou muito específicas na formulação da pesquisa;
- . insuficiência de exaustividade, ou exaustividade em excesso, ou ainda omissão de termos importantes na indexação dos documentos; e
- . falta de interação do usuário no sistema.

WESSEL (125) acrescenta um outro problema. Na teoria, o indexador deveria produzir uma indexação superior à realizada por programas de computadores, mas na prática a indexação manual apresenta muitas inconsistências, para produzir efetivos instrumentos de recuperação da informação.

3.2.2. Estudos comparativos entre o título e outras fontes para indexação automática

Desde a criação do índice KWIC por LUHN (62,63), a indexação automática em títulos tem sido questionada a nível de qualidade da indexação e de qualidade de recuperação da informação.

Vários estudos foram realizados para testar a validade do título como fonte, para extração automática de palavras significativas, como, por exemplo, pesquisa sobre o crescimento de palavras significativas em títulos, análises comparativas qualitativas entre a indexação e recuperação por títulos, resumos e texto integral e utilização de indexação por título para índices de serviços de alerta.

Um dos primeiros estudos foi realizado por MAIZZEL (66), em 1960, testando o conteúdo dos títulos dos artigos do Physics Abstracts. Ele concluiu que 63% dos títulos continham in-

formações suficientes para indexação.

Estudos semelhantes foram realizados por vários autores, comparando palavras de títulos, extraídas automaticamente, e descritores obtidos através de indexação manual, existentes nos índices de assunto de publicações secundárias.

MONTGOMERY & SWANSON (69), em análise realizada, encontraram um índice de 86% dos títulos do Index Medicus com conteúdo suficientemente significativo para serem utilizados em indexação automática.

RUHL (93), em pesquisa na área de Química, encontrou 57% de títulos contendo todos os descritores existentes, do índice de assunto do Chemical Abstracts, e somente 12% dos títulos não continham três ou mais palavras significativas.

KRAFT (55), em títulos de Direito, encontrou 64% de palavras significativas, e somente, 10% não continham nenhuma palavra existente no índice do Index to Legal Periodicals.

Os títulos tornaram-se mais significativos, segundo TOCATLIAN (117), nos anos posteriores ao surgimento do KWIC. O crescimento de informações significativas em títulos estaria relacionado à preocupação dos autores em torná-los mais relevantes, para serem utilizados em índices tipo KWIC. BIRD & KNIGHT (8) sugerem ainda outra justificativa - a necessidade de os autores tornarem seus títulos mais precisos, para poderem se sobressair em relação à explosão bibliográfica existente.

Ao examinar o crescimento de palavras em títulos, GHOSH (39) verificou um aumento significativo entre 1933 e 1972, quando 80% dos documentos poderiam ser recuperados por pesquisa somente no título.

BUXTON & MEADOWS (19) compararam títulos, nas áreas

de ciências exatas e ciências sociais, e constataram um significativo aumento de palavras substantivas entre 1947 e 1973. Para eles, o aumento de substantivos nos títulos pressupõe aumento de palavras relacionadas com o conteúdo dos documentos. Verificaram que os títulos de Química e Botânica possuíam maior valor para a recuperação do que os das áreas de Física, Medicina e História. Na área de ciências sociais é que se encontram títulos com menor valor para a recuperação.

BLOOMFIELD (9) avaliou comparativamente a qualidade da recuperação da indexação manual, da indexação automática em títulos e do KWIC. Concluiu, entre outros resultados, que o uso do resumo para enriquecer a indexação por título gera número elevado de descritores irrelevantes para a recuperação.

SVENONIUS (113) afirma que a indexação, somente por título, apresenta uma precisão maior do que por resumo ou texto integral. De acordo com a autora, o bom desempenho da recuperação não está relacionado com a quantidade de descritores atribuídos a um documento, mas sim à qualidade dos mesmos.

SALTON (97,94), ao contrário desses autores, não aconselha o uso somente de títulos para indexação, pois verificou que é menos eficiente para expressar o conteúdo, do que o uso, também, de resumos.

BARKER, VEAL & WYATT (3), ao compararem a eficiência e o custo de busca bibliográfica em títulos, resumos e descritores na área de Química, verificaram que o resumo e os descritores, obtidos manualmente, aumentam, respectivamente, em 68% e 35% a revocação, mas diminuem em 23% e 10% a precisão. Os títulos são relativamente mais precisos e menos exaustivos. Uma busca em linha utilizando títulos enriquecidos por resumo

ou descritores aumenta cerca de 20% o tempo de uso do computador e, conseqüentemente, cresce o custo de impressão das referências recuperadas, que é inversamente proporcional ao índice de revocação. O título é o mais indicado, pela sua maior precisão e menor custo final ao usuário.

Em 1975 FEINBERG (30) escreveu um livro tratando de estudos comparativos entre o índice KWIC^e o índice Key Word out of Context (KWOC), entre outros índices de palavras permutadas e índices elaborados por indexação automática em títulos. Afirma, como SVENONIUS (113), que o número de descritores atribuídos a um documento não está relacionado à qualidade de indexação, e que um grande número de descritores pode, inclusive, prejudicar a recuperação. Segundo FEINBERG (30), uma das vantagens de indexação a partir do título é a precisão.

GARFIELD (38) e NEUFELD et alii (70) descreveram a aplicação de indexação automática em títulos para elaboração de índices de assunto em boletins de alerta. Segundo os autores essa técnica foi a que ofereceu maior rapidez e precisão.

KWOK (56), partindo do pressuposto de que o título não possui palavras estatisticamente suficientes para indexação automática, propõe como fator de enriquecimento o uso de títulos citados nas referências. Após análises comparativas entre o uso de títulos citados com títulos e resumos e somente com títulos, verificou que o enriquecimento com títulos citados oferece uma representação de conteúdo mais compacta, uniforme e possibilita estabelecer, de forma adequada, relações associativas entre os descritores.

GARFIELD (37), propõe para a indexação automática,

o uso de títulos das citações existentes nos documentos. As citações, segundo o autor, são ilustrações ou complementações do que se deseja informar. São formas de se estabelecerem relações entre trabalhos que possuem pontos em comum e, portanto, são ótimas fontes de indexação, melhores do que os títulos. É a técnica aplicada na elaboração do Índice de assunto do Citation Index.

Um dos mais recentes trabalhos desenvolvidos para testar comparativamente o desempenho da recuperação da informação, através de palavras extraídas automaticamente em títulos, resumos, textos integrais e outras fontes, foi realizado por CLEVELAND, CLEVELAND & WISE (24).

Os autores desenvolveram esta pesquisa baseados no alto custo e na impossibilidade prática da indexação automática em texto integral. Verificaram, após testes de comparação entre oito combinações de indexação automática, que os índices de revocação e precisão apresentados em documentos indexados em fontes como, resumos e títulos são semelhantes aos obtidos por indexação em texto integral.

3.2.3. Métodos de indexação automática

A indexação automática é uma operação que identifica, através de programas de computador, palavras ou expressões significativas dos documentos, para descrever de forma condensada o seu conteúdo.

As palavras significativas são selecionadas automaticamente, através de metodologias específicas, adotadas de acordo com as políticas de indexação e recuperação da informa-

ção, desenvolvimento de software e capacidade de hardware dos sistemas de informação.

As políticas de indexação e recuperação variam, respectivamente, de acordo com a exaustividade e precisão da análise de conteúdo e com os índices de revocação e precisão do resultado da pesquisa. Ambas as políticas dependem, diretamente, das necessidades de informação, caracterizadas pelos diversos tipos de usuários a que um sistema de informação atende.

Esta parte da revisão da literatura trata dos diversos métodos de indexação automática, sem no entanto se deter nos aspectos históricos de cada um. Esses aspectos encontram-se suficientemente analisados no artigo de revisão publicado por BATTY (6).

3.2.3.1. Método de freqüência ou análise estatística

O método de freqüência de palavras foi o primeiro a surgir. Foi proposto por LUHN (62,63) em 1957 e 1958. O autor demonstrou em seus trabalhos que a freqüência de uma palavra em documentos está diretamente relacionada com a capacidade dessa palavra em representar o conteúdo do documento, a nível de indexação e de recuperação da informação. As palavras mais adequadas para a indexação serão as que possuem média freqüência.

O método de freqüência trata da contagem automática do aparecimento da palavra, que pode estar localizada, segundo CLEVELAND, CLEVELAND & WISE (24), no título, resumo, título das referências citadas, texto e em diversas combinações entre estas unidades, como, por exemplo, em título e resumo.

GARFIELD (37) acrescenta ainda, a localização das

palavras significativas através da freqüência no título das citações.

A contagem automática do termo é realizada através da ocorrência e / ou co-ocorrência da palavra. A freqüência pode ser estabelecida, de acordo com SOERGEL (106) e SPARCK JONES (109), através de:

a. ocorrência total da palavra no documento - a palavra é contada todas as vezes que aparece, fazendo-se o somatório das vezes em que co-ocorre, posteriormente;

b. ocorrência única da palavra no documento - conta-se somente uma vez a palavra, independentemente do número de vezes que ela aparece;

c. ocorrência da palavra na coleção - a contagem é realizada somando-se seu aparecimento na coleção.

SOERGEL (106) diferencia, ainda, contagem de conceito. A contagem de conceito é a somatória das freqüências de ocorrência de todas as palavras que determinam aquele conceito. A freqüência de ocorrência de palavras será utilizada para desenvolver a estrutura terminológica, e a do conceito para desenvolver a estrutura classificatória.

A freqüência pode ser realizada, também, em palavras truncadas ou em raízes de palavras. Alguns sistemas utilizam esse tipo de freqüência para diminuir o ruído, evitando o aparecimento de mesmas palavras com diferentes desinências gramaticais (72,92,96,104).

O método de freqüência possui outras aplicações, além da indexação automática.

ROSENBERG (91) utilizou a freqüência de co-ocorrência de palavras como forma de aumentar o desempenho da indexa-

ção manual, fornecendo ao indexador uma lista de descritores candidatos extraídos automaticamente pela análise estatística combinada com a análise de associação entre palavras. Esses descritores, acompanhados de suas respectivas frequências, serão utilizados para indexar novos documentos.

HENZLER (47) aplicou a análise estatística em um estudo quantitativo comparando o vocabulário livre e o controlado e concluiu que a linguagem controlada fornece uma maior perda de informação do autor para o usuário. O ideal seria combinar as duas linguagens de indexação.

A maior aplicação desse método, o de frequência, é para realizar a seleção automática de descritores. A lei de distribuição de palavras em um texto, a lei de Zipf, surge como uma das técnicas que complementam a escolha do descritor. Vários autores aplicaram-na em seus experimentos.

SVENONIUS (113) aplicou a lei de Zipf para verificar qual a frequência que melhor se adapta à seleção automática de descritores. Os resultados encontrados demonstraram que palavras específicas, as de baixa frequência, proporcionam maior precisão na recuperação; em contrapartida, as palavras de média frequência proporcionam maior revocação.

SCHUEGRAF & HEAPS (104), PAO (72) e ROWBOTTOM & WILLET (92) trabalharam com a lei de Zipf em raízes de palavras. Os dois primeiros autores propõem o uso de radicais de palavras para otimizar os custos da recuperação e, principalmente, de armazenamento da informação. Propõem, também, um algoritmo para fragmentar automaticamente palavras equidistantes.

ROWBOTTOM & WILLET (92) não aconselham a extração de palavras aplicando a lei de Zipf em pequenos textos, tais

como títulos e resumos, pois a indexação não será suficientemente exaustiva e precisa.

3.2.3.2. Método de atribuição de peso

O método de atribuição de peso aos descritores, segundo SALTON (94), é uma forma de atribuir-lhes valores semânticos para torná-los mais precisos, sem no entanto diminuir sua capacidade de revocação. É baseado na frequência de cada descritor.

LUHM (62,63) foi, novamente, o precursor deste método. Ele propôs um modelo relacionando diretamente a frequência de uma palavra ou raiz de palavra ao valor dessa palavra para expressar o conteúdo dos documentos, ou seja, quanto maior a frequência, maior peso a palavra receberá.

O peso pode ser atribuído, de acordo com PARKER (73); SALTON, WU & YU (99); SALTON & YANG (100); SPARCK JONES (111) e SOERGEL (106), por:

- a) frequência total ou frequência única - a palavra recebe o mesmo valor do número de sua frequência;
- b) fonte - se a palavra se encontrar em um documento reconhecido como relevante, receberá um peso maior do que outra existente em um documento menos relevante;
- c) por fonte e usuário - o usuário é quem julgará se o documento recuperado é relevante ou não. Se for, os descritores utilizados na estratégia de busca terão, posteriormente, seu valor aumentado; e
- d) frequência na coleção.

A indexação automática com pesos, proposta por

SPARCK JONES (111), denominada por SALTON & YANG (100) de "frequência inversa do documento", trabalha com a especificidade da palavra. As palavras de baixa frequência são as mais específicas e recebem maior peso. As palavras de alta frequência são os responsáveis pelo ruído da recuperação da informação. Segundo SPARCK JONES (111), a extração de um número grande de palavras por documento aumenta a quantidade de frequência, mas não a qualidade dos novos descritores.

Nesse mesmo trabalho SPARCK JONES (111) realizou estudos comparativos qualitativos sobre o desempenho da recuperação através da atribuição de pesos por frequência de ocorrência e co-ocorrência de palavras em documentos e ocorrência de palavras na coleção. Verificou-se que o peso atribuído em relação à coleção é o mais problemático.

A análise discriminatória de documentos é uma das variações do método de atribuição de peso. SALTON & YANG (100) SALTON, WU & YU (99) aplicaram esta análise na frequência de palavras na coleção para aperfeiçoar a revocação e precisão da recuperação. A melhor palavra será aquela que possuir capacidade de discriminação entre os vários documentos semelhantes de uma coleção.

Nesta técnica as palavras que possuem média frequência são as mais indicadas para a indexação do documento ou da pergunta. As palavras de alta e baixa frequência são, respectivamente, raras e gerais em termos de ocorrência e possuem um baixo poder de discriminação. O valor da palavra dependerá da maior ou menor distância que provocar entre os documentos da coleção. Esse valor é calculado através de uma fórmula matemática específica.

DILLON & FEDERHART (27) aplicaram em sua pesquisa um outro tipo de análise discriminatória para selecionar raízes de palavras relativamente freqüentes. Trabalharam, ao contrário dos outros autores anteriormente citados, só com raízes que possuíam alta freqüência de ocorrência na coleção. As raízes foram analisadas de acordo com os vários significados semânticos e aplicou-se, posteriormente, uma função discriminatória para detectar, caracterizar e classificar os grupos semelhantes e diferentes.

SALTON & YANG (100) e SALTON, WU & YU (99) analisaram a teoria da relevância do usuário como método de atribuição de peso. Esta é uma técnica que aplica a freqüência de ocorrência de palavras no documento e na coleção. Requer uma realimentação constante, pois utiliza o julgamento da relevância do usuário para atribuição de pesos.

ROBERTSON & SPARCK JONES (80), YU & SALTON (126) e HARTER & VAN RIJSBERGEN (45) propuseram fórmulas matemáticas para atribuição de pesos baseados na teoria de relevância. PARKER (73) desenvolveu um modelo matemático aplicado à atribuição de peso pelo usuário no momento da pergunta, utilizando palavras extraídas por indexação manual.

3.2.3.3. Método probabilístico

O método probabilístico aplicado por CARROL & ROELOFFS (20), BOOKSTEIN & SWANSON (11) e HARTER (46) aplica a freqüência de co-ocorrência em palavras truncadas automaticamente. As palavras truncadas são extraídas através de um critério estatístico de distribuição binominal, denominado distribuição de Poisson. Aquelas palavras cuja freqüência de dis-

tribuição for descrita pela função de Poisson serão não-significantes.

CARROL & ROELOFFS (20) aplicaram esta técnica para comparar a indexação automática e a manual utilizando artigos da área de Ciência da Informação. BOOKSTEIN & SWANSON (11) aplicaram, além desta técnica, a análise de "cluster", concluindo que as palavras significativas tendem a se aproximarem mais do que as não-significativas e concentram-se mais nas áreas de "cluster".

HARTER (46) introduziu a noção de relevância, aperfeiçoando o modelo proposto por BOOKSTEIN & SWANSON(11) e criou a distribuição de Poisson - 2 para analisar com maior profundidade palavras técnico-científicas. As palavras são tratadas em dois níveis -significantes e significantes especializadas.

HARTER (46), baseado em modelo matemático por ele elaborado, definiu um algoritmo para medir a "indexabilidade" de uma palavra como reflexo do significado relativo das palavras.

3.2.3.4. Método matemático.

O método matemático é baseado na identificação da frequência de co-ocorrência em pares de palavras em documentos, através de algoritmo.

TANIMOTO (115), em 1958, foi o primeiro a propor este método, mas para a classificação automática. Esta idéia foi aplicada à indexação automática, segundo demonstra BATTY (6), a partir dos anos 60.

STEINACKER (112), em 1974, propôs um algoritmo para

detectar frases ou grupos de palavras significativas. O algoritmo produz "cortes no texto" ao localizá-los e, posteriormente, ordena-os alfabeticamente construindo um índice rotado das várias combinações entre as palavras de um mesmo corte. Entre as aplicações dessa técnica, podem ser citadas: criação de dicionários, elaboração de tesouros e desenvolvimento, controle e manutenção de enciclopédias.

3.2.3.5 Análise de "cluster"

A análise de "cluster" foi introduzida, segundo BATTY (6), no início dos anos 60 pelo Cambridge Language Research Unit (CLRU), para a classificação de documentos e elaboração de esquemas de classificação.

A técnica se baseia, de acordo com SALTON (96), no reconhecimento automático, em um grupo de documentos, dos subgrupos de assunto que mais se assemelham entre si e entre outros subgrupos.

SPARCK JONES (110) realizou um projeto de indexação automática e recuperação baseado na análise de "cluster". É aplicada em "pattern of term", termos simples, pares de termos isolados em correlação matricial termo a termo.

3.2.3.6. Método de associação entre Palavras

O método de associação entre palavras, mencionado por SALTON (96), utiliza a frequência de ocorrência e co-ocorrência de palavras ou pares de palavras para identificar o conteúdo dos documentos. As palavras isoladas e as que se co-associam são identificadas em sentenças. Se as co-associações

das mesmas palavras co-ocorrerem com determinada freqüência, então serão consideradas "descritores associados".

O modelo associativo proposto por GIULIANO, JONES & CURTICE (50) parte do princípio que todas as palavras significativas estão relacionadas linearmente. A primeira associação entre as palavras é denominada relação contínua, a segunda representa relações de sinonímia.

LESK (61) comparou os resultados de recuperação em documentos indexados por análise de freqüência e associação, e verificou que o método associativo aumenta o desempenho da recuperação, além de poder ser utilizado em construção e normalização de terminologia para tesauros.

3.2.3.7. Experimentos em avaliação de recuperação da informação em linguagens de indexação

Pesquisas sobre medidas de desempenho de recuperação em linguagens documentárias começaram a se desenvolver, de acordo com REGAZZI (77), após a Segunda Guerra Mundial, devido a necessidade de selecionar a informação útil no caos documentário instalado pela explosão bibliográfica.

A comunidade de pesquisadores realizou vários estudos para descobrir a fórmula ideal de se medir a eficiência de sistemas de recuperação da informação e os instrumentos utilizados para a identificação do conteúdo dos documentos. BOURNE (15) levantou, em sua revisão, as várias formas encontradas para quantificar o desempenho da recuperação em linguagens de indexação. Encontrou, como fatores mais citados, as medidas de revocação e precisão ou relevância.

Várias instituições avaliaram seus sistemas de recuperação e indexação utilizando as duas medidas. BOURNE(15) apresenta um quadro histórico resumindo os projetos experimentais encontrados na literatura a partir de 1954. Dentre estes destacam-se os projetos Cranfield e SMART, por serem os mais citados na literatura.

O projeto Cranfield, como menciona BLOOMFIELD (10), subdivide-se em I e II, e ambos foram desenvolvidos sob a direção de C. W. Cleverdon, no College of Aeronautics, Inglaterra.

O Cranfield I, iniciado em 1957, tinha como objetivos testar e comparar a capacidade de recuperação de quatro sistemas de classificação: Classificação Decimal Universal, Alphabetic Subject Index, uma classificação facetada e o Uniterm System of Coordinate Indexing. CLEVERDON mediu a eficiência de cada linguagem através dos índices de revocação e relevância.

Várias críticas foram dirigidas ao projeto (10,26), tais como: algumas variáveis no julgamento de relevância não foram suficientemente controladas e o número mínimo de palavras, entre 20 e 60, para descrever de forma adequada o conteúdo dos documentos, tal como proposto por CLEVERDON (25), não é adequado para armazenagem e indexação manual.

Em continuação às experiências do Cranfield I, conforme menciona BLOOMFIELD (10) e SIMMONS (105), CLEVERDON criou, em 1967, o Cranfield II para testar três linguagens de indexação. O primeiro tipo de indexação utilizou termos livres, o outro, termos controlados e o terceiro, conceitos simples.

Os testes foram, novamente, baseados nas medidas de revocação e precisão. Partiu da linguagem mais simples para a mais sofisticada, utilizando dicionários de sinonímia, associa-

ções entre conceitos e hierarquia de termos. Os resultados apresentados, segundo descreve SALTON (97,98), demonstraram que a indexação com termos simples e livres é mais eficiente do que as mais sofisticadas.

CLEVERDON (25), através dessas experiências, propôs uma indexação, idealmente exaustiva, com 33 descritores por documento. Novamente, de acordo com BLOOMFIELD (10) foi questionada a viabilidade de um número tão grande de descritores.

Segundo REGAZZI (77), o maior mérito dos projetos Cranfield foi abrir as pesquisas na área de avaliação de recuperação da informação para diversas linguagens documentárias, além de ter definido claramente como aumentar a revocação e precisão dos sistemas de informação.

O projeto SMART foi desenvolvido, a partir de 1965, por G. Salton (105). Foi elaborado para realizar avaliações de várias linguagens de indexação, em termos de revocação e precisão. Contém, também, um conjunto de programas para realizar indexação automática em textos integrais.

Vários testes de avaliação de linguagem foram realizados no projeto SMART (96,p.339), e possibilitaram algumas conclusões:

- a) o uso de termos com peso é normalmente mais efetivo do que termos sem peso;
- b) o uso de dicionários de sinonímia é melhor do que o controle por palavras truncadas;
- c) o uso de títulos é normalmente menos efetivo para a análise de conteúdo do que o resumo; e
- d) a mais importante das conclusões, segundo SALTON (105) e que foi, também, encontrada no Cranfield; as linguagens de indexa-

ção mais sofisticadas são menos eficientes do que as que utilizam termos livres e simples.

3.3. Revisão da Literatura Nacional

A literatura brasileira sobre indexação automática, comparada à estrangeira, é significativamente menor. O que, provavelmente, reflete o pouco desenvolvimento desta técnica no País.

A indexação automática de documentos, no Brasil, iniciou-se, praticamente, segundo BRAGA (17), com a utilização do programa KWIC para elaborar os índices das bibliografias especializadas que o Instituto Brasileiro de Bibliografia e Documentação (IBBD), atual Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), publicava.

Os índices de assunto, conforme menciona OLIVEIRA (71), eram elaborados por palavras-chave permutadas retiradas, automaticamente dos títulos das obras, com inclusão, se necessário, de termos preestabelecidos para enriquecer os títulos.

A primeira experiência da utilização do programa KWIC de acordo com ZAHER & DUARTE (127) e ZAHER et alii (128), foi para editar a Bibliografia Brasileira de Física, em 1968. Foi realizada por um grupo de especialistas do IBBB e do Centro Brasileiro de Pesquisa Físicas. Pode-se dizer que este fato marca o início da indexação automática no Brasil.

Após as experiências do antigo IBBB com o programa KWIC, surgiram trabalhos que o questionavam ou o aplaudiam (54, 127) e relatos de experiências de sua utilização, para a indexação e recuperação automática da informação em bibliotecas (64).

Esses trabalhos estão concentrados entre 1960 e 1970. Após 1970 encontram-se algumas pesquisas que não tratam especificamente de indexação automática, mas que servem como apoio ao desenvolvimento dessa técnica.

Estudos sobre a utilização de títulos e / ou resumos para indexação automática foram realizados por SOUZA (108) e BRAGA (17). A primeira autora analisou títulos de artigos de periódicos estrangeiros em Ciência da Informação e Biblioteconomia, simulando manualmente a técnica do KWIC. Constatou um crescente aumento do número de palavras significativas nos títulos dos artigos dos periódicos analisados, entre 1980 e 1970, anos posteriores à criação do KWIC e concluiu que existe uma tendência de aumentar as palavras significativas nos títulos.

BRAGA (17) propõe, em sua dissertação a utilização do resumo como fator de enriquecimento do título. Os títulos apesar de serem pontos de acesso ao conteúdo de documentos, são insuficientes para uma perfeita indexação e recuperação da informação. A autora concluiu, que a proporção de palavras significativas do resumo é da ordem de doze para cada uma existente no título. Seu estudo foi realizado em títulos de periódicos científicos na área de Química. Ela aplicou técnica do KWIC simulado, e para comparar as palavras dos títulos e dos resumos elaborou tabelas de frequência de palavras.

A análise de frequência de palavras utilizando leis bibliométricas foi, encontrada pela primeira vez na literatura brasileira em 1973. MAIA (65) aplicou a primeira e segunda lei de Zipf esta última na forma enunciada por BOOTH

(13), assim como a fórmula de transição de Goffman (41), como fatores de análise da informação em língua portuguesa. As duas leis de Zipf estabelecem relações entre a ordem de série de uma palavra e a freqüência de seu aparecimento em texto suficientemente longo.

BOOTH (13) enuncia a segunda lei, através de uma fórmula matemática ($r.f=c$). Estabelece que, quando as palavras de um texto qualquer são ordenadas numa tabela, em ordem decrescente de freqüência de aparecimento, o produto da ordem na série (r) da palavra por sua freqüência (f) é uma constante (c).

A fórmula de transição de Goffman (41) determina as ordens de série nas quais devem-se encontrar as palavras significativas de um texto em língua inglesa. A autora concluiu que as leis são aplicáveis à língua portuguesa, apresentando para o português um valor diferente da constante "C" da língua inglesa.

Outros estudos bibliométricos foram realizados utilizando as leis de Zipf, sendo alguns na área de Linguística documentária (79), e outros utilizando a lei de Bradford (74).

ROBREDO (87) utilizou a lei de Bradford, formulada por BROOKES (18), como instrumento de controle terminológico estabelecendo: "descritores de escopo" - termos de alta freqüência e baixa especificidade que caracterizam áreas do conhecimento; "descritores de facetas" - termos de média freqüência e especificidade, caracterizam subáreas de interesse e os "descritores pontuais", os de baixa freqüência e alta especificidade, caracterizando um número limitado de documen-

tos.

O método de frequência de palavras em títulos e ou resumos, para determinação de descritores e construção de núcleos de termos, foi aplicado por ROBREDO em vários trabalhos (7, 85, 86, 89, 90).

O Centro de Informações Nucleares (CIN), segundo BARREIRO (5), aplica a técnica de frequência e uso de descritores para a seleção adequada do descritor em indexação manual e atualização do tesauro INIS.

A análise estatística e estudos de co-ocorrência de frases e palavras significativas foram utilizados como metodologia por QUEIROZ (75), para elaboração automática de resumos, e por TORRES FILHO (118), na elaboração de índices automáticos de livros técnicos. Este autor propõe um algoritmo de frequência.

A indexação automática, utilizando raízes vazias e raízes significativas foi proposta por ROBREDO (83, 85), para aumentar a rapidez de processamento e a precisão da recuperação da informação.

FREUND (35) descreve em artigo técnica semelhante. As raízes são extraídas automaticamente por análise estrutural, através de um algoritmo específico. A técnica é utilizada, somente, para a recuperação em linha, ou seja, para a indexação da pergunta e montagem da estratégia busca. O autor fez, também, uma breve comparação entre a análise estrutural e o truncamento arbitrário de palavras.

Na literatura brasileira encontram-se alguns trabalhos que aplicam a lingüística computacional para a indexação automática em textos integrais. Citamos HALLER (42, 43) e ANDREWSKY & RUAS (1). O primeiro autor desenvolveu seu progr-

ma na Universidade de Brasília. Esse programa consta de análises morfológica e sintática das palavras do texto, para extrair os descritores. Possui vários dicionários, entre eles, o de frequência de palavras vazias e significativas e o de raízes.

ANDREWSKY & RUAS (1) utilizam métodos lingüísticos e estatísticos de atribuição de pesos para as palavras significativas. O programa, em andamento, é uma adaptação para a língua portuguesa do Sytème Syntaxique et Probabiliste d' Indexation et de Recherche d' Informations Textuelles (SPIRIT), desenvolvido pelo Centre National de la Recherche Scientifique (CNRS) para a língua francesa. Possui algoritmos de análise sintática e análise semântica, além de diversos dicionários.

4. HIPÓTESES

Primeira hipótese

A utilização de técnicas de indexação automática, quando aplicadas a títulos e resumos de um número suficientemente elevado de registros bibliográficos, permite identificar os termos significativos que caracterizam o conteúdo de uma base de dados formada por esses registros, com probabilidade, no mínimo, equivalente àquela obtida quando aplicadas técnicas de indexação manual.

Segunda hipótese

A qualidade de recuperação em linha de registros bibliográficos, em termos de precisão, a partir de uma base de dados indexada automaticamente é, no mínimo, equivalente à da recuperação em linha a partir de uma base de dados, contendo os mesmos registros, indexada manualmente.

4.1. Variáveis

- Termos escolhidos na indexação manual de registros bibliográficos, à vista do texto integral dos artigos e sua frequência na base de dados;
- Termos de indexação escolhidos automaticamente a partir dos títulos e resumos dos registros bibliográficos e sua frequência na base de dados;
- Termos coincidentes nos vocabulários de termos obtidos manual e automaticamente dos registros bibliográficos;
- Quantidade de registros bibliográficos recuperados e quantidade de registros relevantes recuperados, na busca na base de dados, indexados manualmente;
- Quantidade de registros bibliográficos recuperados e quantidade de registros relevantes recuperados, na busca na base de dados, indexados automaticamente;
- Estratégias de busca.

4.2. Definições operacionais

4.2.1. - Conceitos expressos nas hipóteses

- Número suficientemente elevado: quantidade, de registros bibliográficos que compõem uma base de dados, acima de 100.

- Termos significativos: palavras que expressam aspectos conceituais relativos à Ciência da Informação, que ocorrem nos vocabulários resultantes de indexação manual e auto-

mática.

Considera-se que a ocorrência simultânea de um termo nas duas bases de dados, reforça seu valor significativo.

- Precisão ou relevância: índice de medida de qualidade da recuperação, calculado através da quantidade de registros bibliográficos relevantes recuperados sobre a quantidade de registros recuperados.

4.2.2. Conceitos implicitamente relacionados com as hipóteses

- Frequência dos termos: número de vezes que um termo ocorre no arquivo de dados, indexado manual ou automaticamente;

- Frequência de associação dos termos: número de vezes que um termo se associa com um outro termo em registros bibliográficos diferentes;

- Registros bibliográficos relevantes recuperados: referências com resumos dos documentos considerados como relacionados com o(s) assunto (s) na(s) busca(s);

- Ruído: registros bibliográficos não relevantes recuperados. É inversamente proporcional à precisão;

- Documento (ou unidade documentária): cada um dos artigos que compõe o periódico Ciência da Informação;

5. METODOLOGIA

Dentre as várias áreas do conhecimento existentes, foi selecionada a literatura brasileira referente à Ciência da Informação, por ser relativamente nova, carente de estudos sistemáticos quanto à sua terminologia e por ser a área de atuação dos profissionais de informação, bibliotecários, cientistas da informação, documentalistas e tantos outros.

A presente pesquisa foi realizada utilizando, como universo do estudo, os artigos publicados no periódico Ciência da Informação, no período de 1972 a 1983, por considerar que esse periódico:

- cobre um leque de assuntos variados;
- cobre amplamente o território nacional;
- contém artigos de diversos autores;
- cobre um período retrospectivo de 12 anos;
- oferece aproximadamente 150 artigos, em grande maioria em língua portuguesa, o que representa uma quantidade suficiente de documentos para geração de uma base de dados e realização dos testes projetados.

A coleta de dados compreendeu as operações de registro e descrição bibliográfica, incluindo a preparação de um resumo e a escolha dos descritores por indexação manual dos registros.

As etapas do trabalho experimental foram as seguintes:

1. descrição bibliográfica de cada artigo do periódico Ciência da Informação, incluindo um resumo sucinto e a escolha manual dos descritores;

2. digitação em linha dos registros bibliográficos e formatação dos dados em forma processável;
3. elaboração de diversos instrumentos de controle referentes ao arquivo anterior;
4. formatação de uma base de dados, para interrogação em linha, a partir dos registros indexados manualmente;
5. cópia do arquivo anteriormente formatado, com eliminação dos descritores introduzidos manualmente, para gerar um novo arquivo e processamento deste para indexação automática dos registros;
6. elaboração de diversos instrumentos de controle, referentes ao arquivo anterior;
7. formatação de uma base de dados, para interrogação em linha, a partir dos registros indexados automaticamente;
8. escolha das perguntas e montagem das estratégias de busca para interrogação das duas bases de dados;
9. interrogação das duas bases de dados;
10. tabulação e análise comparativa dos resultados correspondentes às duas bases de dados.

Para a digitação, formatação, processamento, indexação automática, formatação das bases de dados e recuperação da informação foram utilizadas as facilidades oferecidas pelo sistema BIB/DIALOGO (84,88), implementado no Departamento de Biblioteconomia da UnB, para computadores Burroughs B6700 e terminais Burroughs, modelo TVA 800/10, com a linguagem de controle CANDE (Command and Edit)⁽⁺⁾. No anexo 1 encontra-se uma breve descrição das funções dos programas utilizados.

(+) BURROUGHS CORPORATION. B6700/B7700 Command and Edit (CANDE) language: information manual. S.L. Burroughs Corporation, 1972.

Descrevem-se, em continuação, com maiores detalhes, as etapas seguidas na pesquisa.

5.1. Descrição e registro dos documentos

O formato do registro bibliográfico é o utilizado pelo sistema BIB/DIALOGO, (++) o qual se baseia, com algumas simplificações, no formato sugerido pela UNESCO (+++).

O registro possui os seguintes campos, subdivididos, eventualmente, em subcampos:

- a) número do registro;
- b) autor(es) pessoal (is);
- c) entidade (autor corporativo ou filiação do autor pessoal);
- d) título do documento, seguido, eventualmente, de: elementos recuperáveis do título, tradução enriquecimento do título, notas;
- e) imprensa, compreendendo: título do periódico, volume, número, páginas inicial e final do artigo, data, elementos complementares (referências a tabelas, quadros, figuras, etc.) e notas;
- f) resumo indicativo (máximo de 240 caracteres);
- g) descritores (máximo de 25 descritores de até 32 caracteres).

(++) ROBREDO, J. BIB/BATCH, Manual de registro bibliográfico. Brasília, UnB. Departamento de Biblioteconomia, 1981.

(+++) UNESCO, UNISIST guide to standards for information handling. Paris, UNESCO, 1980. 1980. 304 p.

Os registros bibliográficos foram elaborados por uma estagiária de Biblioteconomia, do último semestre do curso ministrado no Departamento de Biblioteconomia da Universidade de Brasília, sem experiência prática anterior em indexação ou elaboração de resumos, exceto a adquirida durante o curso. A estagiária foi instruída a levantar e registrar os elementos físicos e de conteúdo dos documentos de acordo com padrões preestabelecidos, no sentido de garantir um mínimo de coerência e normalização.

A escolha dos descritores deveria ser a mais próxima possível da linguagem do autor do artigo, utilizando-se como recursos informacionais para a indexação o título, o resumo já existente nos artigos e o texto integral. A indexação foi realizada em linguagem livre, sem qualquer controle de vocabulário. Os únicos padrões estabelecidos foram no sentido de evitar o uso de adjetivos ou verbos como descritores, optando-se, em cada caso, pela forma substantiva e, no caso de descritores formados por várias palavras, provocar uma pequena repetição, quando os termos independentes possuem significado relevante próprio, como por exemplo:

- . Informação científica e tecnológica;
- . Informação científica;
- . Informação tecnológica;
- . Informação;
- . Ciência e Tecnologia;
- . Informação científica;
- . Ciência;
- . Informação tecnológica;
- . Tecnologia.

critores no arquivo de dados (ordenados alfabeticamente e em ordem crescente de frequência) e listagens de frequência de associação de todos os descritores com aqueles que ocorrem mais de quatro vezes.

Para obtenção do índice KWIC foi utilizada a tabela de palavras vazias de significado, já utilizada na obtenção de um índice KWIC com os títulos dos artigos publicados na Revista de Biblioteconomia de Brasília. Na figura 3 pode-se ver uma página do índice KWIC dos títulos do periódico Ciência da Informação.

O índice KWIC gerado apresentou-se satisfatório, mas com algumas palavras vazias e omissões de palavras significativas. Isto se deve ao fato de que algumas palavras que deveriam ser consideradas significativas para a área de Ciência da Informação estavam na tabela de palavras proibidas, o que determinou a sua eliminação. Inversamente, algumas palavras vazias de significado, que não deveriam aparecer no índice KWIC, principalmente as estrangeiras, apareceram por não terem sido incluídas na tabela de palavras vazias.

A figura 4 apresenta alguns exemplos de palavras significativas eliminadas e palavras vazias não eliminadas. Estes pequenos problemas encontrados demonstram a necessidade de aprimoramento constante de qualquer tabela de palavras, para torná-la mais adequada para seu uso numa área específica, no nosso caso na área de Ciência da Informação.

Na figura 5 encontra-se um exemplo da listagem das referências dos registros bibliográficos. As figuras 6, 7 e 8 são, respectivamente, exemplos dos índices de assuntos, de autores pessoais e de entidades.

A figura 9 mostra um fragmento da listagem de des-

DA BIBLIOGRAFIA A CIENCIA DA INFORMACAO; UM ARIAS ESTUDO REALIZADO NO INSTITUTO DE FISIQUA E CIENCIAS A CIENCIA DA INFORMACAO E AS CIENCIAS	C100001
HISTORICO E UMA POSICAO.	C100012
HUMANAS.	C100143
HUMANOS NA PRODUCAO E COMERCIALIZACAO DO LIVRO DO BRASIL PRO DUZ: ATIVIDADES DE PESQUISA EM QUIMICA DO BRASIL, RELATIVAS	C100035
IBICT SOBRE A ATIVIDADE PROFISSIONAL DOS EGRESSOS.	C100123
IBICT.	C100106
IBICT.	C100064
IDENTIFICACAO DA FRENTE DE PESQUISA ATRAVES DE FILIROS DE LUA II A RELACAO ENTRE O AUTOR FONTE E AUTORES CITADOS).	C100094
II THE RELATION BETWEEN SOURCE AUTHOR AND CITED AUTHOR POPULA II. O BANCO DE DADOS I.S.J. - ASCA - BRASIL, UM NOVO INSTRUME	C100050
III. C.I.E.L.H.C.I.A./BRASIL UM CATALOGO MULTIDIRECCAO SOBRE A IMPACTO DA AUTOMACAO NAS BIBLIOTECAS; UMA REVISAO.	C100048
IMPACTOS DA TECNOLOGIA DA INFORMACAO.	C100112
IMPACTOS DOS CURSOS DO IBICT SOBRE A ATIVIDADE PROFISSIONAL O	C100125
IMPLANTACAO DE UM NUCLEO REGIONAL DE INFORMACAO NO NORDESTE.	C100123
IMPLANTACAO DE UMA REDE REGIONAL; A REDE EUROPEIA DE INFORMA	C100090
IMPLANTACAO DO SISTEMA NACIONAL DE INFORMACAO CIENTIFICA E TE	C100062
IMPORTANCIA DA EUROPA.	C100022
IMPORTANCIA E PARASITOLOGIA (IMP) NA LITTERATURA CIENTIFICA INT	C100061
INDICACAO AUTOMATICA BASEADA EM METODOS LINGUISTICOS E SUA AP	C100116
INDICACAO DOS OCUROSITOS E DE RECUPERACAO DA INFORMACAO MEDIA	C100058
INDICACAO E RECUPERACAO DA INFORMACAO NOS SISTEMAS NACIONAIS	C100083
INDEXACAO.	C100130
INDICADOR DE QUALIDADE DA PRODUCAO CIENTIFICA; UMA REVISAO.	C100134
INDICATIVOS DE LINGUAG, DE PAIS, DE AUTORIDADE, DE EBAUOS E I	C100006
INDICE EM CADEIA.	C100034
INDICES ALFABETICOS DA CLASSIFICACAO DECIMAL UNIVERSAL.	C100026
INDICES DAS TABELAS DA CLASSIFICACAO DECIMAL UNIVERSAL.	C100025
INDUSTRIAL.	C100089
INFANTIL DA BIBLIOTECONOMIA.	C100019
INFLUENCIA NA PUBLICACAO DE INFORMACAOES.	C100108
INFORMACAO).	C100023
INFORMACAO).	C100047
INFORMACAO - EURONET.	C100062
INFORMACAO; FATOR DE DESENVOLVIMENTO.	C100085
INFORMACAO; FOMULACAO PRELIMINAR DO PROBLEMA.	C100156
INFORMACAO A SERVICO DE MUDANCAS CULTURAIS.	C100122
INFORMACAO AGRICOLA).	C100076
INFORMACAO AMBIENTAL.	C100113
INFORMACAO ATUALIZADAS).	C100030
INFORMACAO BIBLIOGRAFICA.	C100126
INFORMACAO CIENTIFICA E TECNICA NA FRANCA E SUA IMPORTANCIA N	C100061
INFORMACAO CIENTIFICA E TECNICA.	C100083
INFORMACAO CIENTIFICA E TECNOLÓGICA (SHICT).	C100022
INFORMACAO CIENTIFICA E TECNOLÓGICA NO BRASIL.	C100105
INFORMACAO CIENTIFICA E TECNOLÓGICA.	C100129
INFORMACAO CIENTIFICA E TECNOLÓGICA.	C100104
INFORMACAO CIENTIFICA E TECNOLÓGICA).	C100036
INFORMACAO CIENTIFICA NO MUNDO.	C100050
ANALISE DO BANCO DE DADOS DO O IMPACTOS DOS CURSOS DO	
YFORMACAO EM CIENCIA E TECNOLOGIA; O PLANO PRIORITARIO DO YSTITUTO BRASILEIRO DE INFORMACAO E CIENCIA E TECNOLOGIA	
MOR AND CITED AUTHOR POPULATIONS (ESTUDOS EM CIENCIA E TECNOL	
PRODUCAO CIENTIFICA BRASILEIRA, SEGUNDO OS DADOS DO I.S.J.	
PRODUCAO CIENTIFICA BRASILEIRA SEGUNDO OS DADOS DO I.S.J.	
PROGRESSIVOS PARA A O EXEMPLO DE	
ULTRIZES BASICAS PARA	
A INFORMACAO CIENTIFICA E TECNICA NA FRANCA E SUA O COMERCIALIZACAO TECNOLÓGICA E SUA	
PARTICIPACAO DOS PESQUISADORES DE MICROBIOLOGIA; MEDIDAS DE CONSISTENCIA DA	
UTILIZACAO DOS PROCESSOS DE FUNCOES DAS LINGUAGENS DE	
ELECCIOS DE LINGUISTICA PARA ESTUDOS DE VALIDADE DA ANALISE DE CITACAO CURIO	
SERVICOS DE INFORMACAO EM INSTITUTOS DE PESQUISA AUTOMACAOES	
CIENCIAS E TIPOLOGIAS E CIENCIA E TECNOLOGIA E SUA SCIENCE COMO A UMA TEORIA DE BIBLIOTECOMIA E CIENCIA DA	
INFORMATIO SYSTEMS (ESPACOS QUASE-RETICOS E SISTEMAS DE LO DE IMPLANTACAO DE UMA REDE REGIONAL; A REDE EUROPEIA DE	
TRANSFERENCIA DE	
BARREIRAS CULTURAIS A TRANSFERENCIA DE	
A INFORMACAO E A EVOLUCAO DOS CENTROS DE	
LUS USUARIOS DE LA INFORMACION AGRICOLA (OS USUARIOS DA POLITICA BRASILEIRA DE	
SIDERACAO DA CLASSIFICACAO EMPRENHATIVA PARA NECESSIDADES DE A MOKA COMO FONTE DE	
AO DA INFORMACAO NOS SISTEMAS NACIONAIS E INTERNACIONAIS DE DIRETRIZES BASICAS PARA IMPLANTACAO DO SISTEMA NACIONAL	
A	
CONTRIBUICAO AO DEBATE SOBRE POLITICA NACIONAL DE POLITICAS E PROGRAMAS NACIONAIS DE	
FIG AND TECHNICAL INFORMATION (NECESSIDADES DA SOCIEDADE EM SINIERE DA	

Figura 3 - Índice KWIC dos títulos do periódico Ciência da Informação

Palavras incluídas como vazias na tabela utilizada para o índice KWIC e que de fato são significativas na área de Ciência da Informação

SISTEMA (S)

BRASIL

Palavras não incluídas na tabela e que de fato são vazias na área de Ciência da Informação

ACABA	EXEMPLO(S)
ATIVIDADE	INFLUENCIA
QUEBRA-CABEÇAS	PROBLEMA
	•
	•
	ETC

Figura 4 - Palavras significativas eliminadas e palavras vazias não eliminadas

CI00040

BRAGA, G.H.

INSTITUTO BRASILEIRO DE BIBLIOGRAFIA E DOCUMENTAÇÃO,
 INFORMAÇÃO, CIÊNCIA, POLÍTICA CIENTÍFICA: O PENSAMENTO DE DE
 REK DE SOLLA PRICE.

CIÊNCIA DA INFORMAÇÃO, V.3, N.2, P.155-177, 1974.

ABORDA OS PROBLEMAS FUNDAMENTAIS DA CIÊNCIA E DA INFORMAÇÃO
 CIENTÍFICA, PROPOUNDO O LANÇAMENTO DE UM JORNAL CIENTÍFICO DI
 ÁRIO PARA FACILITAR O INTERCÂMBIO DE INFORMAÇÕES.

CI00041

CESAR, I.D.

INSTITUTO BRASILEIRO DE BIBLIOGRAFIA E DOCUMENTAÇÃO,

PROJETO CAPESQ.

CIÊNCIA DA INFORMAÇÃO, V.3, N.2, P.183-187, 1974, COMUNICAÇÃO.

PROJETO DE CADASTRO DE PESQUISAS, TENDO POR META O REGISTRO
 DAS PESQUISAS NO CAMPO DA CIÊNCIA E DA TECNOLOGIA. PERMITE A
 OS USUÁRIOS ACESSO AOS TRABALHOS CIENTÍFICOS EM FASE DE EXPERI
 MENTAÇÃO.

CI00042

CALDEIRA, P. DA T.

UNIVERSIDADE FEDERAL DE MINAS GERAIS,

PROCESSO DE CRESCIMENTO EPIDEMIOLÓGICO APLICADO A LITERATURA
 BRASILEIRA DE DOENÇA DE CHAGAS.

CIÊNCIA DA INFORMAÇÃO, V.4, N.1, P.5-16, 1975.

APLICA A TEORIA DE GOFFMAN NA ANÁLISE DA PRODUÇÃO DE TRABALH
 OS SOBRE DOENÇA DE CHAGAS NO PERÍODO DE 1909 A 1971.

CI00043

GUSMÃO, H.R.; CRISTOVÃO, R.T.; GARCIA, M.J. DE O.

UNIVERSIDADE FEDERAL FLUMINENSE; INSTITUTO BRASILEIRO DE BIB
 LIOGRAFIA E DOCUMENTAÇÃO.

A CIÊNCIA E O QUEBRA-CABECAS.

CIÊNCIA DA INFORMAÇÃO, V.4, N.1, P.17-22, 1975.

APLICAÇÃO DE TESTES E SUAS INTERPRETAÇÕES, SUGESTÕES DE UM M
 ETODO PARA TENTATIVA DE COMPROVAÇÃO DA ANALOGIA ENTRE O DESE
 NVOLVIMENTO DA CIÊNCIA E DO QUEBRA-CABECAS.

CI00044

PRICE, D. DE S.; GURSEY, S.

YALE UNIVERSITY,

STUDIES IN SCIENTOMETRICS I TRANSCIENCE AND CONTINUANCE IN SC
 IENTIFIC AUTHORSHIP (ESTUDOS EM CIENCIOMETRIA I TRANSCIÊNCIA
 E CONTINUIDADE DE AUTORIA CIENTÍFICA).

CIÊNCIA DA INFORMAÇÃO, V.4, N.1, P.27-40, 1975.

ESTUDA O FENÔMENO DE TRANSCIÊNCIA/PERMANÊNCIA NA FRENTE DE P
 ESQUISA PELA ANÁLISE DO FLUXO ANUAL DE AUTORES DE UMA POPULA
 ÇÃO AMOSTRADA.

JORNAIS	CI00029
JORNAL=CIENTIFICO	CI00040
JORNAL=DO=BRASIL	CI00039
KWIC	CI00093
LEI	CI00024
LEI	CI00141
LEI=DE=BRADFORD	CI00057
LEI=DE=BRADFORD	CI00141
LEI=DE=LOTKA	CI00082
LEI=DE=ZIPF	CI00024
LEI=DE=ZIPF	CI00039
LEIS	CI00017
LEIS	CI00081
LEITURES	CI00101
LEITURA	CI00130
LEVANTAMENTO=BIBLIOGRAFICO	CI00077
LINGUA	CI00006
LINGUA	CI00057
LINGUA=PORTUGUESA	CI00024
LINGUA=PORTUGUESA	CI00039
LINGUA=PORTUGUESA	CI00135
LINGUAGEM	CI00006
LINGUAGEM=DE=BUSCA	CI00031
LINGUAGEM=DOCUMENTARIA	CI00027
LINGUAGEM=JORNALISTICA	CI00039
LINGUAGEM=NATURAL	CI00025
LINGUAGEM=NATURAL	CI00130
LINGUAGEM=NATURAL	CI00135
LINGUAGENS=DE=INDEXACAO	CI00083
LINGUAGENS=DE=RECUPERACAO	CI00083
LINGUAGENS=NATURAIS	CI00027
LINGUAGENS=NATURAIS	CI00091
LINGUISTICA	CI00039
LINGUISTICA	CI00130
LINGUISTICA	CI00131
LINGUISTICAS	CI00027
LINGUISTICOS	CI00135
LISA	CI00093
LITERATURA	CI00017
LITERATURA	CI00038
LITERATURA	CI00042
LITERATURA	CI00046
LITERATURA	CI00052
LITERATURA	CI00057
LITERATURA	CI00081
LITERATURA=CIENTIFICA	CI00058
LITERATURA=CIENTIFICA	CI00094
LITERATURA=GEOLOGICA	CI00017
LIVRO	CI00035
LIVRO	CI00107
LIVROS	CI00005
LIVROS	CI00013

Figura 6 - Índice de assunto

ABREU, D. DE	CI00005
AGUIAR, A.C.	CI00106
AIRES, J.	CI00112
ALVES, C.M.	CI00080
ANDREWSKI, A.	CI00135
AQUINO, L. DE	CI00090
ARAUJO, E.E.R. DE	CI00096
ARAUJO, V.M.R.H. DE	CI00100
ARAUJO, V.M.R.H. DE	CI00115
BARANOW, U.G.	CI00131
BARBOSA, A.P.	CI00008
BARRETO, A. DE A.	CI00110
BAUZER, R.	CI00099
BOMENY, R.H.D.	CI00082
BOYCE, B.	CI00038
BRAGA, G.M.	CI00016
BRAGA, G.M.	CI00029
BRAGA, G.M.	CI00040
BRAGA, G.M.	CI00074
CALDEIRA, P. DA T.	CI00042
CARNEIRO, P.	CI00065
CARVALHO, A. DE O.	CI00049
CARVALHO, M. DE L.B. DE	CI00057
CARVALHO, M.B.P. DE	CI00049
CARVALHO, M.M. DE	CI00052
CASSIM, M.B.	CI00122
CESAR, I.D.	CI00041
CINTRA, A.M.M.	CI00124
CINTRA, A.M.M.	CI00130
COSTA NETO, C. ET ALII.	CI00117
CRISTOVAO, H.T.	CI00043
CRISTOVAO, H.T.	CI00094
CUNHA, L.G.C. DA	CI00020
CUNHA, L.G.C. DA	CI00070
CUNHA, M.L.M.	CI00053
CURVO FILHO, P.F.	CI00095
CURVO FILHO, P.F.	CI00139
DAHLBERG, I.	CI00091
DATTA, S.	CI00068
DAVIG, A.P.	CI00089
DIAS, A.C.	CI00019
DUSOULIER, N.	CI00059
DUSOULIER, N.	CI00060
DUSOULIER, N.	CI00061
DUSOULIER, N.	CI00062
DUSOULIER, N.	CI00063
EYRE, J.	CI00097
FIGUEIREDO, L.M. DE	CI00017
FIGUEIREDO, L.M. DE	CI00075
FIGUEIREDO, N.M. DE	CI00077
FIGUEIREDO, N.M. DE	CI00103
FIGUEIREDO, N.M. DE	CI00127

Figura 7 - Índice de autores pessoais

-INSTITUTO BRASILEIRO DE BIBLIOGRAFIA E DOCUMENTACAO	CI00054
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00073
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00074
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00075
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00090
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00094
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00099
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00110
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00111
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00114
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00115
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00122
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00123
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00124
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00127
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00130
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00133
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00140
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00141
-INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA	CI00143
-INSTITUTO DE BOTANICA	CI00108
-INSTITUTO DE BOTANICA	CI00120
-INSTITUTO DE DOCUMENTACAO DA FUNDACAO JOAQUIM NABUCCO	CI00107
-INSTITUTO DE ENERGIA ATOMICA DE SAO PAULO	CI00037
-INSTITUTO DE PESQUISAS ESPACIAIS	CI00031
-INSTITUTO DE PESQUISAS TECNOLOGICAS DO ESTADO DE SAO PAULO	CI00125
-INSTITUTO ESTADUAL DE PATRIMONIO CULTURAL/SEEC - RIO DE JANEIRO	CI00136
-INSTITUTO MILITAR DE ENGENHARIA	CI00080
-INSTITUTO NACIONAL DE METROLOGIA, NORMALIZACAO E QUALIDADE INDUSTRIAL	CI00126
-INSTITUTO NACIONAL DE TECNOLOGIA	CI00039
-INSTITUTO UNIVERSITARIO DE PESQUISAS DO RIO DE JANEIRO	CI00088
-INTERNATIONAL CLASSIFICATION VERLAG DOKUMENTATION	CI00091
-LOUGHBOROUGH UNIVERSITY OF TECHNOLOGY	CI00102
-PONTIFICIA UNIVERSIDADE CATOLICA DO RIO DE JANEIRO	CI00135
-PONTIFICIA UNIVERSIDADE CATOLICA DO RIO DE JANEIRO	CI00080
-SCHOOL OF LIBRARIANSHIP	CI00097
-SINDICATO NACIONAL DOS EDITORES DE LIVROS	CI00095
-SISTEMA NACIONAL DE INFORMACAO CIENTIFICA E TECNOLOGICA	CI00022
-THE POLYTECHNIC OF NORTH LONDON	CI00068
-UNIVERSIDADE DE BRASILIA	CI00077
-UNIVERSIDADE DE BRASILIA	CI00002
-UNIVERSIDADE DE BRASILIA	CI00010
-UNIVERSIDADE DE BRASILIA	CI00015
-UNIVERSIDADE DE BRASILIA	CI00078
-UNIVERSIDADE DE BRASILIA	CI00103
-UNIVERSIDADE DE BRASILIA	CI00118
-UNIVERSIDADE DE BRASILIA	CI00131
-UNIVERSIDADE DE ILLINOIS	CI00051
-UNIVERSIDADE DE SAO PAULO	CI00119
-UNIVERSIDADE DE SAO PAULO	CI00144

Figura 8 - Índice de entidades

0002	ACERVO BIBLIOGRAFICO	0002	TEORIA DA COMUNICACAO
0001	ACESSIBILIDADE DA INFORMACAO	0002	TRANSCENCIA
0001	ACESSO	0002	TRATAMENTO DA INFORMACAO
0005	ACESSO A INFORMACAO	0002	USUARIO POTENCIAL
0001	ACOPLEMENTO BIBLIOGRAFICO	0002	UTILIDADE DA INFORMACAO
0001	ADMINISTRACAO DE PROJETOS	0002	VOCABULARIO ESPECIALIZADO
0001	AGENCIA SOCIAL	0003	AUTORES CIENTIFICOS
0001	AGENCIAS BIBLIOGRAFICAS	0003	BIBLIOTECA UNIVERSITARIA
0001	AGENTE DE INFORMACAO	0003	CENTROS DE DOCUMENTACAO
0001	AGENTE SOCIAL	0003	COMUNICACAO FORMAL
0005	AGRICULTURA	0003	DEPOSITO LEGAL
0002	AMERICA LATINA	0003	DESCRIPTOR
0008	ANALISE BIBLIOMETRICA	0003	DOCUMENTOS
0013	ANALISE DA INFORMACAO	0003	EDITORES
0002	ANALISE DE CITACOES	0003	ESTUDO DE USUARIO
0002	ANALISE DO DOCUMENTO	0003	FREQUENCIA A BIBLIOTECA
0001	ANALISE OCCUMENTARIA	0003	LITERATURA CIENTIFICA
0001	ANALISE ESTRUTURAL	0003	PATENTES
0001	ANALISE LINGUISTICA	0003	PERIODICOS
0001	ANALISE SEMANTICA	0003	PRODUTIVIDADE DO SISTEMA
0001	ANALISE SINTATICA DOS TERMOS	0003	QUIMICA
0001	AQUISICAO	0003	RECURSOS DE INFORMACAO
0001	AQUISICAO PLANIFICADA	0003	RELEVANCIA
0001	ARIST	0003	SERVICOS DE BIBLIOTECA
0001	ARMAZENAGEM DA INFORMACAO	0003	SISTEMAS DE INFORMACAO
0002	ARMAZENAMENTO DA INFORMACAO	0003	TECNOLOGIA DA INFORMACAO
0004	ARQUIVO	0003	TEORIA DA INFORMACAO
0001	ARQUIVO DO ISI	0003	TERMINOLOGIA
0001	ARQUIVO PARTICULAR	0003	VOCABULARIO
0001	ARQUIVO PASCAL	0004	ARQUIVO
0001	ARQUIVO PRINCIPAL	0004	CATALOGACAO
0001	ASCA	0004	CITAcoes BIBLIOGRAFICAS

Figura 9 - Fragmentos das listagens com descritores e respectivas frequências (ordem alfabética e ordem numérica).

critores com suas respectivas freqüências, ordenadas alfabeticamente e em ordem crescente de freqüência de aparecimento no arquivo. A figura 10 mostra a estrutura da listagem que indica a freqüência de associações dos descritores que ocorrem quatro vezes ou mais.

5.4. Formatação da base de dados a partir dos registros indexados manualmente.

A base de dados formatada a partir dos registros bibliográficos indexados manualmente foi denominada CIFORM (Ciência da INFORMação).

5.5. Preparação do arquivo de dados indexado automaticamente.

A preparação do arquivo de dados indexados automaticamente compreende a realização das seguintes etapas:

1. cópia do arquivo indexado manualmente, contendo a descrição bibliográfica, o resumo e os descritores, de maneira a formar um novo arquivo idêntico, do qual foram eliminados os descritores;

2. hifenação de algumas palavras do título e / ou do resumo para provocar uma ligação entre elas;

3. criação / atualização das tabelas de palavras e raízes vazias de significado, necessárias para execução de programas de indexação automática;

4. indexação automática dos registros a partir dos títulos e dos resumos previamente hifenados.

A figura 11 mostra um registro do arquivo indexado

13.000	ANALISE DA INFORMAÇÃO
1.000	ACERVO BIBLIOGRAFICO
2.000	ACESSO A INFORMAÇÃO
1.000	AGRICULTURA
1.000	ANALISE DO DOCUMENTO
1.000	ANALISE DOCUMENTARIA
1.000	ARMAZENAGEM DA INFORMAÇÃO
1.000	ARQUIVO DO ISI
1.000	ARQUIVO PASCAL
1.000	ASSOCIAÇÃO DE DESCRITORES
8.000	AUTOMAÇÃO
1.000	AUTORES CIENTIFICOS
3.000	BANCO DE DADOS
2.000	BASES DE DADOS
1.000	BIBLIOGRAFIAS
2.000	BIBLIOTECA
1.000	BIBLIOTECAS
1.000	BIBLIOTECONOMIA
1.000	BINAGRI
3.000	BRASIL
1.000	BUSCA RETROSPECTIVA
1.000	CANAIS DE INFORMAÇÃO
1.000	CATALOGO MULTIINDEXADO
5.000	CENTRO DE DOCUMENTAÇÃO
1.000	CENTROS DE DOCUMENTAÇÃO
1.000	CIENCIA
2.000	CIENCIA DA INFORMAÇÃO
1.000	COLEGIO INVISIVEL
4.000	COMPUTADOR
4.000	COMUNICAÇÃO CIENTIFICA
1.000	COMUNICAÇÃO DA INFORMAÇÃO
1.000	COMUNICAÇÃO INFORMAL
1.000	COMUNICAÇÃO RURAL
1.000	COMUNICAÇÃO TECNICA
5.000	COMUNIDADE CIENTIFICA
1.000	CONCEITO DE BIBLIOTECA
1.000	CONSISTENCIA DA INDEXAÇÃO
1.000	CONTEUDO DO DOCUMENTO
1.000	CONTEUDO DOS DOCUMENTOS
1.000	CONTROLE TERMINOLOGICO
2.000	CRESCIMENTO DA LITERATURA
1.000	DESCRITOR
1.000	DESCRITORES
4.000	DISSEMINAÇÃO DA INFORMAÇÃO
4.000	DOCUMENTAÇÃO
3.000	DOCUMENTO
1.000	DOCUMENTO ORIGINAL
1.000	DOCUMENTOS
1.000	ENTRADA DE INFORMAÇÃO

Figura 10 - Fragmento da listagem de associações dos descritores com freqüência maior do que 3

C100141
 PINHEIRO, L.V.R.

INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA.

LEI DE BRADFORD : UMA REFORMULACAO CONCEITUAL.

CIENCIA DA INFORMACAO, V.12, N.2, P.59-80, 1983.

INTRODUZ-SE O CONCEITO DE PRODUTIVIDADE RELATIVA E COMPARA-S
 E COM O DE PRODUTIVIDADE ABSOLUTA, MOSTRANDO-SE AS ALTERACOE
 S OCORRIDAS. RESSALTA-SE AS CONSEQUENCIAS DAS DISTORCOES DA
 LEI, SOBRETUDO PARA A POLITICA DE AQUISICAO.

00C100141*

LEI DE BRADFORD

LITERATURA CIENTIFICA

PRODUTIVIDADE RELATIVA

PRODUTIVIDADE ABSOLUTA

PERIODICOS CIENTIFICOS

BIBLIOMETRIA

ANALISE BIBLIOMETRICA

PRODUTIVIDADE DE PERIODICOS

LEI DE ZIPP

LEI DA DISPERSAO DA LITERATURA

POLITICA DE AQUISICAO

AQUISICAO

C100141
 PINHEIRO, L.V.R.

INSTITUTO BRASILEIRO DE INFORMACAO EM CIENCIA E TECNOLOGIA.

LEI-DE-BRADFORD : UMA REFORMULACAO CONCEITUAL.

CIENCIA DA INFORMACAO, V.12, N.2, P.59-80, 1983.

INTRODUZ-SE O CONCEITO DE PRODUTIVIDADE-RELATIVA E COMPARA-S
 E COM O DE PRODUTIVIDADE-ABSOLUTA, MOSTRANDO-SE AS ALTERACOE
 S OCORRIDAS. RESSALTA-SE AS CONSEQUENCIAS DAS DISTORCOES DA
 LEI, SOBRETUDO PARA A POLITICA-DE-AQUISICAO.

00C100141*

Figura 11 - Registro indexado manualmente e
 registro hifenado preparado pa-
 ra indexação automática.

manualmente (acima) e o mesmo registro, após eliminação dos descritores e hifenação, copiado num novo arquivo (abaixo).

O processo de hifenação dos registros do novo arquivo é necessário para assegurar a presença de determinados termos significativos, tais como: indexação automática, estudo de Usuário, sistema de informação, São Paulo, países em desenvolvimento, etc., que se converterão, respectivamente, nos descritores: indexação-automática, estudo-de-usuário, sistema-de-informação, São-Paulo, países-em-desenvolvimento. De outra maneira, dado que na indexação automática considera-se como palavra qualquer cadeia delimitada por um separador (espaço, vírgula, ponto, ponto e vírgula, etc.), os descritores seriam: indexação, automática, sistema, informação, São, Paulo, países, desenvolvimento, etc.

É importante ressaltar que se deve realizar a hifenação com parcimônia, a fim de evitar a possível perda de determinadas palavras isgnificativas, assim como, por exemplo, a hifenação das palavras relações-bibliométricas levaria a perder a palavra biblométricas como descritor independente.

Para efetuar a indexação automática dos registros utilizaram-se novos programas, recentemente incorporados ao sistema BIB/DIALOGO, baseados numa nova filosofia de processamento. (+)

(+) ROBREDO, J. Indexação automática de textos e montagem de bases de dados para armazenagem e recuperação de informações. In: UNIVERSIDADE DE BRASÍLIA. Departamento de Biblioteconomia. Programa de ação para 1985 (Pesquisa em andamento).

As palavras do título e do resumo são comparadas em primeiro lugar, com as palavras vazias que figuram numa tabela previamente estabelecida, denominada WORDFIXED. Essa tabela contém preposições, conjunções, artigos, pronomes, advérbios, etc., que são as palavras instrumentais, fechadas, estáticas, que determinam a estrutura interna do sistema lingüístico e constituem aproximadamente 50% de um texto. Constan ainda dessa tabela palavras que apresentam poucas desinências e que pertencem ao sistema aberto da língua.

As palavras diferentes das que figuram na tabela WORDFIXED, após eliminação das palavras coincidentes são gravadas num arquivo intermediário, com o número de registro.

A seguir, as palavras que permaneceram são comparadas com as palavras vazias de significado, que figuram numa tabela denominada WORDROOTS.

Essa tabela contém palavras truncadas e raízes de palavras da morfologia lexical, que formam o sistema lingüístico aberto, dinâmico. São os verbos, adjetivos e os substantivos, que não remetem a um conceito na área de Ciência da Informação.

Eliminam-se, nessa etapa, aquelas palavras que coincidem com os radicais e as truncagens. Consideram-se, então, como descritores dos registros, as palavras do título e resumo que passaram por esses dois filtros de "não-significado" e não foram eliminadas.

O objetivo de se criar uma tabela de raízes e palavras truncadas vazias é aumentar a rapidez do processamento, diminuir o espaço necessário de armazenagem das tabelas e diminuir o ruído resultante da presença daquelas palavras vazias que pos-

suem desinências variadas, eventualmente não previstas numa tabela formada unicamente por palavras fixas. Por exemplo, a palavra truncada EXPLIC eliminará as palavras:

explica
 explicações
 explicação
 explicar
 explicativa
 explicativo
 explícita
 explícitas
 explícito
 explícitos
 explicou
 ... etc.

As tabelas WORDFIXED e WORDROOTS foram elaboradas especificamente para a área de Ciência da Informação, para língua portuguesa e para o processamento da indexação automática.

Foi utilizada, como ponto de partida, para a geração das duas tabelas, a tabela anteriormente aplicada para a elaboração dos índices KWIC dos periódicos Ciência da Informação e Revista de Biblioteconomia de Brasília. As palavras integrantes dessa tabela foram subdivididas em dois lotes, que vieram a constituir a lista de palavras vazias e a lista de raízes e palavras truncadas vazias.

As tabelas foram depuradas utilizando como instrumentos de avaliação os Índices KWIC de Revista de Biblioteconomia de Brasília e de Ciência da Informação, assim como dicionários e gramáticas da língua portuguesa.

Nas figuras 12 e 13 podem-se ver dois fragmentos, respectivamente, das tabelas WORDFIXED e WORDROOTS.

A figura 14 permite visualizar o processo de indexação automática dos registros bibliográficos a partir dos títulos e resumos.

5.6. Processamento do arquivo de dados indexado automaticamente

O arquivo de dados indexado automaticamente foi processado de maneira a se obter os mesmos tipos de listagens descritos no item 5.3. Com as listagens dos índices de assuntos, autores pessoais e entidades e com a listagem das referências bibliográficas foi montado um índice cumulativo dos 12 volumes iniciais do periódico Ciência da Informação, que se encontra no Depto de Biblioteconomia da UnB para consulta interna.

5.7. Formatação da base de dados a partir dos registros indexados automaticamente

A base de dados formatada a partir dos registros bibliográficos indexados automaticamente foi denominada INDEX-CI (Indexação de Ciência da Informação).

5.8. Escolha das perguntas e montagem das estratégias de busca

A fim de evitar qualquer julgamento subjetivo na escolha das perguntas que serviram para realizar o teste comparativo de recuperação das bases de dados (CINFORM, indexada

A	CONTRA
ABAIXO	CUSTA
ABRIL	DA
ACAO	CAI
ACERCA	DALI
ACIMA	DAO
ACCES	DAGUI
ACULA	DAR
ACORDO	DAS
ADEUS	DE
AFIM	DELA
AFINS	DELAS
AFORA	DELE
AGORA	DELES
AGOSTO	DEMAIS
AI	DENTRE
AINDA	DEPENDE
ALEM	DEFENDEM
AMANHÃ	DEPENDER
AMBAS	DEPOIS
AMBOS	DESDE
ANO	DESSA
ANOS	DESSAS
ANTE	DESSE
ANTES	DESSÉS
AO	DESTA
ACONDE	DESTAS
AOS	DESTF
APENAS	DESTES
APESAR	DIVISAO
APCS	DIVISOES
AGUI	DO
AGUILO	DOIS
AS	DOCS
ASSIM	DOZE
	DUAS

Figura 12 - Fragmento da tabela
Wordfixed

ABAL	CONUM
ABANDON	CONCED
ABERT	CONCEITU
ABORD	CUNCI
ABR	CONCO
ABSOL	CONFER
ABSOR	CONFL
ABSU	CONFOR
ABUS	CONFRA
ACAO	CONFRO
ACEIT	CONFU
ACENT	CONNECE
ACERT	CONNECID
ACH	CONJU
ACIUM	CUNGU
ACOLH	CONSEG
ACOMPANH	CONSER
ACRED	CONSIDER
ACRES	CONSTAT
ACUS	CONSTIT
ADENT	CUNTEN
ADENUA	CUNTEST
ADER	CUNTEXT
ADJA	CUNTID
ADMIR	CUNTIN
ADMIS	CUNTOR
ADMIT	CUNTRAO
ADOT	CUNTRAR
ADVLR	CUNTRAST
AFET	CUNTRIB
AFIRM	CUNTROLA
AFLIC	CUNTRULO
AFLIG	CUNTROVER
AFRONT	CUNVENC
AJUD	CUNVERS
ALGU	CUNVI

Figura 13 - Fragmento da tabela Wordroots

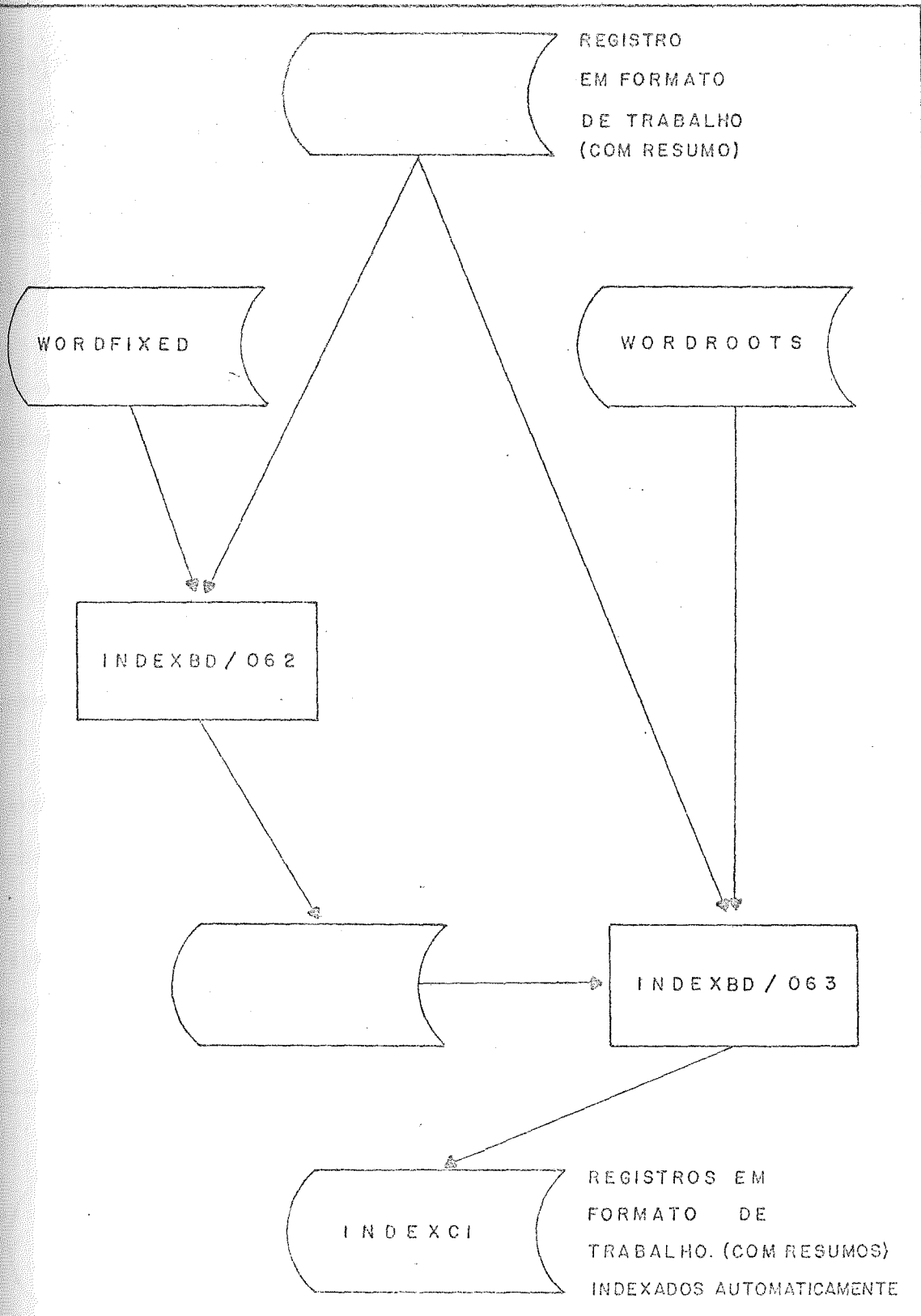


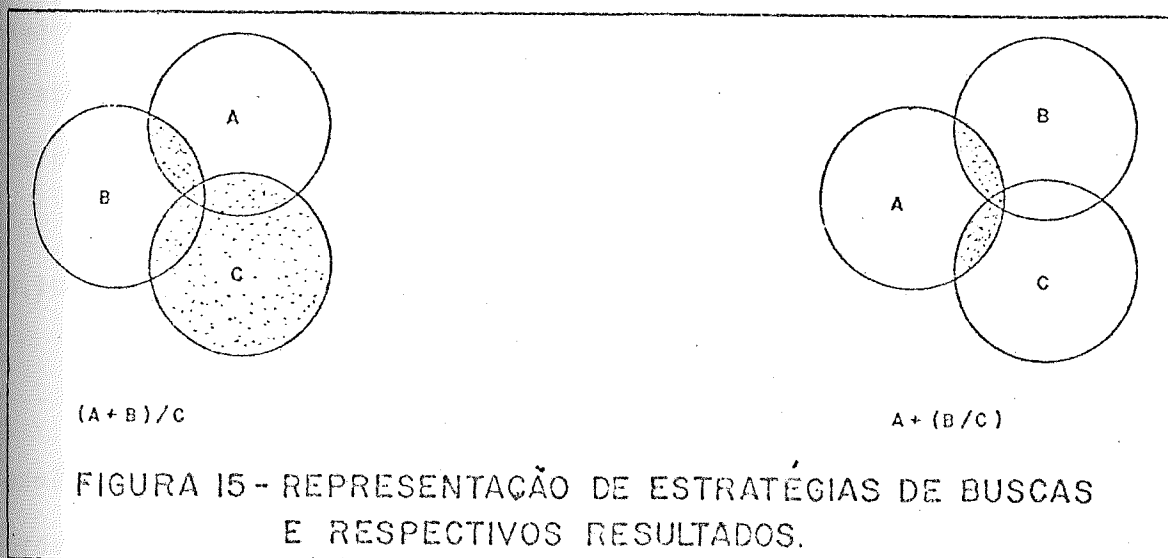
FIGURA 14

CONCEITUAÇÃO DO SUBSISTEMA DE INDEXAÇÃO AUTOMÁTICA, A PARTIR DOS TÍTULOS E RESUMOS.

manualmente, e INDEXCI, indexada automaticamente), foram consultados oito professores do Departamento de Biblioteconomia da UnB, que forneceram sugestões de buscas bibliográficas a serem realizadas. No total foram recolhidas 33 perguntas, uma vez eliminadas as coincidentes.

A partir dessas perguntas foram montadas 72 estratégias de busca, utilizando os operadores booleanos AND(E), representado pelo signo "+", e OR (ou), representado pelo signo "/" e combinações diversas de descritores (assuntos, autores ou assuntos e autores), até um máximo de três em cada estratégia. Essas condições obrigaram à realização de várias estratégias de busca para cobrir todos os aspectos de uma pergunta.

O sistema exige o uso de parênteses para delimitar claramente as relações entre os descritores utilizados, evitando ambigüidades. Assim, no caso da pergunta $A + B / C$, só o uso de parênteses permite distinguir entre $(A + B) / C$ e $A + (B / C)$, que conduziriam a resultados bastante diferentes, como mostra a figura 15.



O programa para busca e recuperação admite o uso de indicadores de truncagem à direita dos descritores, o que permite maior flexibilidade na escolha do nível de especificidade da estratégia, segundo as necessidades.

Durante a busca, o programa indica, na tela, o número de respostas correspondentes a cada descritor, independentemente, e à combinação de termos escolhidos em cada estratégia, o que permite eventuais modificações da mesma.

O resultado da busca é fornecido através de saídas na tela do terminal ou impressas, fornecendo o(s) número(s) do(s) registro(s) selecionado(s) ou a(s) referência(s) bibliográficas(s) completa(s).

Nas figuras 16 e 17 visualizam-se dois exemplos de saída impressa, das referências bibliográficas com a estratégia e resultados numa busca na base INDEXCI, e a outra na base CIFORM.

5.9. Interrogação das duas bases de dados

As perguntas às duas bases de dados foram realizadas de forma interativa, dirigindo-se, sempre, a mesma estratégia para as duas bases de dados. Na maioria dos casos partiu-se da estratégia geral para a específica.

5.10. Tabulação e análise dos resultados

Os descritores obtidos por indexação manual e automática e os existentes no índice KWIC foram arranjados em ordem alfabética, em uma única tabela com três colunas, con-

FUNDACAC UNIVERSIDADE DE BRASILIA

DEPARTAMENTO DE BIBLIOTECCNOMIA - CENTRO DE PROCESSAMENTO DE DADOS
SISTEMA BIBLIOLÓGICO

NOPE DO USUARIO : JAIME RCHREDO

ENDERECC DO USUARIO : DEPTO DE BIBLIOTECCNOMIA
UNIVERSIDADE DE BRASILIA
70910 DF

NUMERO DA PERGUNTA : 0001

PERGUNTA DO USUARIO : (LIVRS/INFORMACS)+PAISES-EM-DESENVOLVIMENTO

DESCRITORES :	TOTAL DE REF. NO ARGLIVO
LIVRS	==> 5
INFORMACS	==> 56
PAISES-EM-DESENVOLVIMENTO	==> 3

PERGUNTA DO USUARIO :	TOTAL DE REF. NO ARGLIVO
(LIVRS	/
INFORMACS)+
PAISES-EM-DESENVOLVIMENTO	==> 2

REFERENCIAS :

C100005
APREU, C. DE
SINDICATO NACIONAL DOS EDITORES DE LIVROS
LIVROS PARA OS PAISES-EM-DESENVOLVIMENTO
CIENCIA DA INFORMACAC, V.1, N.1, P.37-39, 1972
POLITICA-EDITORIAL DE LIVROS, ANALISE DOS PROBLEMAS DE FORNE
CIMENTO-DE-LIVROS E SUGESTOES PARA SUAS SOLUCOES

C100072
SCHLIE, I. H.
UNIVERSITY OF DENVER RESEARCH INSTITUTE
A UTILIZACAC DA INFORMACAC SOBRE PATENTES EM PAISES-EM-DESENVOLVIMENTO ; ESTUDO DE CASO EM ANDAMENTO
CIENCIA DA INFORMACAC, V.6, N.2, P.55-58, 1977
ABORDA A ATUACAC DO DENVER-RESEARCH-INSTITUTE NO USC-DA-INFORMACACAO SOBRE PATENTES EXISTENTES NOS ARGLIVOS DOS ESTADOS-UNIDOS

Figura 16 - Busca bibliografica realizada
na base de dados INDEXCI

FUNDAÇÃO UNIVERSIDADE DE BRASÍLIA

DEPARTAMENTO DE BIBLIOTECA - CENTRO DE PROCESSAMENTO DE DADOS

SISTEMA BIB/DIALOGO

NOME DO USUARIO : SIMONE BASTOS

ENDEREÇO DO USUARIO : CURSO DE MESTRADO
 DEPTO DE BIBLIOTECOLOGIA
 UNIVERSIDADE DE BRASÍLIA
 70910 BRASÍLIA DF

NÚMERO DA PERGUNTA : 0147

PERGUNTA DO USUARIO : (BIBLIOMETRIA/CITAC&)+DISSERTAC&

DESCRIÇÕES :	TOTAL DE REF. NO ARQUIVO
BIBLIOMETRIA	==> 18
CITAC&	==> 5
DISSERTAC&	==> 1

PERGUNTA DO USUARIO :	TOTAL DE REF. NO ARQUIVO
(BIBLIOMETRIA	/
CITAC&)+
DISSERTAC&	==> 1

REFERÊNCIAS :

C100121
 RODRIGUES, M. DA P.L.
 UNIVERSIDADE FEDERAL DO PARANÁ
 CITAC&ES NAS DISSERTAC&ES DE MESTRADO EM CIENCIA DA INFORMAC
 AO
 CIENCIA DA INFORMAC&O, V.11, N.1, P.35-59, 1982
 ANALISE DE 2310 CITAC&ES CONSTANTES EM 62 DISSERTAC&ES APRES
 ENTADAS AO CURSO DE MESTRADO EM CIENCIA DA INFORMAC&O DO IBI
 CI, NO PERIODO DE 1972 A 1979. CONCLUS&ES DA ANALISE

Figura 17 - Busca bibliográfica realizada
 na base de dados CIFORM

tendo a(s) freqüências(s) de ocorrência nos arquivos das bases de dados CIFORM e INDEXCI e no índice KWIC.

Com os elementos dessa tabela elaborou-se uma outra, contendo as raízes dos descritores manuais e automáticos em ordem alfabética. Essa tabela de raízes significativas foi dividida em três colunas, onde figuram a freqüência de ocorrência da raiz e das diversas desinências que dela se originam, incluindo os termos pré-coordenados e termos compostos, nas bases de dados CIFORM e INDEXCI e no índice KWIC.

Os resultados de cada estratégia de busca foram ordenados em seqüência alfabética por descritor de busca bibliográfica em uma tabela com três colunas, contendo o número de referências recuperadas, número de referências relevantes recuperadas e o número de referências recuperadas comuns (relevantes e irrelevantes) nas bases de dados CIFORM e INDEXCI. A tabela contém, também, o somatório do resultado de todas as estratégias realizadas para cada busca bibliográfica.

Essas três tabelas encontram-se, respectivamente, nos anexos 2,3 e 4.

O critério utilizado para o julgamento das referências relevantes foi o da existência de palavras significativas, pertinentes ao assunto da busca, no resumo e no título. Nos casos de ambigüidade, dúvida e inexistência de palavras significativas, recorreu-se ao artigo original referenciado.

Para a comparação entre os vocabulários obtidos por indexação automática e manual foi aplicada a lei de Bradford, através de programas computacionais desenvolvi-

dos no Departamento de Biblioteconomia da UnB.

A qualidade de recuperação das bases de dados CIFORM e INDEXCI foi medida através do índice de precisão. Utilizou-se o pacote SPSS (Statistical Package for the Social Science) para a realização do teste Mann-Whitney-U- Wildon

5.11. Limitações da pesquisa

Esta pesquisa foi realizada especificamente para a área de Ciência da Informação e para textos em língua portuguesa, utilizando-se, somente, os artigos do periódico Ciência da Informação, publicados entre 1972 e 1983, para a geração da base de dados.

A indexação manual foi realizada em linguagem livre, sem qualquer uso de instrumentos de controle de vocabulário. A estagiária encarregada de indexar os documentos não possuía experiência prévia a não ser a adquirida no curso de Biblioteconomia da Universidade de Brasília.

Os aspectos históricos, teóricos, matemáticos e lingüísticos da indexação não foram abordados. Deu-se ênfase aos estudos de frequência dos descritores e de precisão da recuperação.

6. APRESENTAÇÃO DOS RESULTADOS

6.1. Resultados dos estudos comparativos entre os vocabulários obtidos por indexação automática e manual

A letras I, C, K, nas tabelas a seguir descritas, representam, respectivamente, INDEXCI (base de dados indexada automaticamente, CIFORM (base de dados indexada manualmente) e índice KWIC.

Os percentuais apresentados nas tabelas 1 e 2 foram calculados sobre o valor 857, que representa o total de descritores da tabela alfabética comparativa (anexo 2). Cabe ressaltar que no vocabulário obtido por indexação automática figuram alguns termos em inglês e espanhol. Para facilitar a comparação com o vocabulário obtido por indexação manual, que não incluía nenhum termo estrangeiro, foram eliminados esses termos estrangeiros.

A tabela 1 demonstra os totais comparativos dos descritores comuns entre as duas bases de dados e entre estas e o índice KWIC.

Tabela 1

Descritores coincidentes

Base de dados	Descritores comuns	%
I e C	203	23,6
I e K	264	30,8
C e K	172	20,0

Na tabela 2 estão dispostos os totais dos descritores não-coincidentes.

Tabela 2
Descritores não-coincidentes

Base de dados	Descritores não-coincidentes	%
I	252	29,4
C	237	27,6
K	1	0,1

Nota-se, entre as tabelas 1 e 2, que os índices de descritores comuns e descritores não-coincidentes são relativamente semelhante ou seja, o índice de coincidência entre as bases de dados INDEXCI e CIFORM é relativamente baixo, principalmente se comparado ao índice de 60% mencionado por SALTON (97,98) como o normalmente encontrado.

Aplicando-se a fórmula descrita por SALTON (96,p. 336), mencionada na página 15 encontra-se como coeficiente da avaliação o valor de 0.22, comprovando-se o baixo grau de coincidência.

Na tabela 3 o índice de coincidência aumenta cerca de 20% entre I e C, em relação à tabela 1. Essa tabela representa a comparação entre descritores com frequência (f) de ocorrência, nas duas bases de dados, maior ou igual a quatro.

Tabela 3

Descritores com frequência ≥ 4

Bases de dados	Descritores $f \geq 4$	% (+)
I e C	65 ⁽⁺⁺⁾	43,0
C	19	12,5
I	39	25,0

(+) Cálculo realizado sobre o valor 151, que representa o total de descritores em I e C com frequência maior ou igual a quatro.

(++) Este valor equivale ao total de termos comuns em I e C, com frequência maior ou igual a quatro em pelo menos um deles.

As tabelas 4 e 5 apresentam, respectivamente, os totais de coincidência e não-coincidência dos radicais e truncamentos dos descritores. A lista de raízes encontra-se no anexo 3.

Os percentuais da tabela 4 foram calculados sobre o valor 371 (total de raízes e descritores truncados).

Tabela 4

Raízes e truncamentos coincidentes

Bases de dados	Raízes e truncamentos comuns	%
I e C	161	43
I e K	172	46
C e K	121	32

Ao se compararem os percentuais entre a tabela 1 e 4, verifica-se um aumento de coincidência de aproximadamente 10% a 20%.

Em relação à tabela 3, o índice de coincidência entre I e C é exatamente o mesmo-43%.

Na tabela 5 esses resultados continuam apresentando índices relativamente inferiores, exceto no valor encontrado em K.

Tabela 5

Raízes e truncamentos não-coincidentes

Bases de dados	Raízes e truncamentos não-coincidentes	%
I	97	26.1
C	40	10.7
K	1	0.26

Verificaram-se, também, na lista de raízes significativas das bases de dados INDEXCI e CIFORM (anexo 3), o total de descritores compostos e os com pré-coordenação. Encontraram-se para I e C, respectivamente, os totais de 147 e 312, o que significa que a indexação manual apresenta-se com um índice superior, de 52.9%, de descritores compostos e pré-coordenados em relação à indexação automática.

Nas tabelas 6, 7 e 8 estão dispostos os resultados da aplicação da lei de Bradford à distribuição dos descritores

nos vocabulários das bases de dados INDEXCI, CIFORM e na lista de raízes e palavras truncadas de INDEXCI.

Os dados das tabelas 6, 7 e 8 representam, na primeira coluna (A), a frequência de ocorrência dos descritores em ordem decrescente; na coluna B, o número de descritores que aparecem com uma determinada frequência; na terceira coluna ($\sum B$), o somatório acumulado dos descritores da coluna B; e na quarta coluna ($\sum AXB$) o somatório acumulado das frequências da coluna A.

Tabela 6

Lei de Bradford aplicada à INDEXCI

A	B	$\sum B$	$\sum (A \times B)$
50	001	001	0050
22	001	002	0072
14	001	003	0086
13	001	004	0099
12	002	006	0123
11	001	007	0134
10	004	011	0174
09	002	013	0192
08	005	018	0232
07	007	025	0281
06	016	041	0377
05	010	051	0427
04	016	067	0491
03	051	118	0644
02	106	224	0856
01	365	589	1221

Tabela 7

Lei de Bradford aplicada à CIFORM

A	B	ΣB	$\Sigma(A \times B)$
91	001	001	0091
51	001	002	0142
41	001	003	0183
34	001	004	0217
29	003	007	0304
26	002	009	0356
24	001	010	0380
23	002	012	0426
22	002	014	0470
21	001	015	0491
20	001	016	0511
19	002	018	0549
18	001	019	0567
16	002	021	0599
14	001	022	0613
13	002	024	0639
12	005	029	0698
11	001	030	0710
10	003	033	0740
09	007	040	0803
08	004	044	0835
07	005	049	0870
06	008	057	0918
05	013	070	0983
04	014	084	1039
03	022	106	1105
02	066	172	1237
01	346	518	1583

Tabela 8

Lei de Bradford aplicada à lista de
raízes e palavras truncadas de INDEXCI

A	B	ΣB	$\Sigma (A \times B)$
65	001	001	065
38	001	002	103
28	001	003	131
25	001	004	156
23	002	006	196
19	002	008	234
16	002	010	266
15	001	011	281
14	002	013	309
12	003	016	345
11	002	018	367
10	003	021	390
09	007	028	453
08	005	033	490
07	004	037	518
06	007	044	560
05	011	055	615
04	015	070	670
03	031	101	763
02	051	152	805
01	173	325	978

6.2. Resultados dos estudos comparativos entre a qualidade de recuperação em linha das bases de dados INDEXCI e CIFORM

A qualidade de recuperação, nesta pesquisa, foi medida através de precisão, que é inversamente proporcional à revocação e ao ruído, ou seja, quanto maior a precisão, menores serão a revocação e o ruído.

A precisão foi calculada através da fórmula mencionada anteriormente na página

A tabela 9 contém o total de referências recuperadas (R), o total de referências relevantes recuperadas (Rr) e o índice de precisão (P).

Para se diferenciar o resultado final de recuperação das buscas, acrescentou-se uma letra ao lado dos números das respectivas estratégias. Esse resultado é o somatório das referências recuperadas nas várias estratégias, eliminando-se aquelas referências coincidentes.

No total foram recuperadas em INDEXCI e CIFORM, respectivamente, 233 e 416 referências, 156 e 206 referências relevantes, e 204 referências coincidentes em ambas as bases de dados.

Tabela 9
Índice de Precisão

Nº	INDEXCI			CINFORM		
	R	Rr	P	R	Rr	P
1	11	8	0.727	4	2	0.5
1a	11	8	0.727	4	2	0.5
1b	2	2	1.000	1	1	1.000
2	15	13	0.866	36	17	0.472
2a	-	-	-	14	5	0.357
2b	2	2	1.000	-	-	-
2c	14	12	0.857	28	17	0.607
3	1	1	1.000	2	1	0.5
3a	1	1	1.000	2	1	0.5
3b	1	1	1.000	1	1	1.000
4	19	12	0.631	30	20	0.666
4a	19	12	0.631	30	20	0.666
4b	8	8	1.000	29	19	0.655
4c	9	8	0.888	27	18	0.666
4d	4	3	0.75	9	4	0.444
4e	2	1	0.5	6	4	0.666
5	5	3	0.6	3	1	0.333
5a	5	3	0.6	3	1	0.333
5b	-	-	-	-	-	-
6	5	3	0.6	5	2	0.4
6a	-	-	-	-	-	-
6b	1	-	-	1	-	-
6c	1	-	-	1	-	-
6d	1	1	1.000	1	-	-
6e	2	2	1.000	2	2	1.000
6f	-	-	-	-	-	-
6g	-	-	-	-	-	-
6h	2	2	2	2	2	1.000
7	12	12	1.000	20	17	0.85

Nº	INDEXCI			CINFORM		
	R	Rr	P	R	Rr	P
7a	11	11	1.000	20	17	0.85
7b	9	9	1.000	18	16	0.888
7c	1	1	1.000	8	8	1.000
8	-	-	-	-	-	-
8a	-	-	-	-	-	-
8b	-	-	-	-	-	-
8c	-	-	-	-	-	-
8d	-	-	-	-	-	-
8e	-	-	-	-	-	-
9	-	-	-	-	-	-
9a	-	-	-	-	-	-
9b	-	-	-	-	-	-
10	3	3	1.000	3	2	0.666
10a	3	3	1.000	1	-	-
10b	1	1	1.000	1	1	1.000
10c	-	-	-	1	1	1.000
11	6	4	0.666	9	5	0.555
11a	6	4	0.666	9	5	0.555
11b	1	1	1.000	1	-	-
12	11	9	0.818	16	9	0.562
12a	11	9	0.818	16	9	0.562
12b	3	3	1.000	12	8	0.666
12c	1	1	1.000	1	1	1.000
12d	2	2	1.000	1	1	1.000
12e	3	3	1.000	2	2	1.000
13	1	1	1.000	1	1	1.000
13a	1	1	1.000	1	1	1.000
14	-	-	-	-	-	-
15	1	1	1.000	1	1	1.000
15a	1	1	1.000	-	-	-
15b	1	1	1.000	1	1	1.000

№	INDEXCI			CINFORM		
	R	Rr	P	R	Rr	P
16	3	-	-	-	-	-
16a	3	-	-	-	-	-
17	10	8	0.8	34	12	0.352
17a	-	-	-	4	1	0.25
17b	1	1	1.000	-	-	-
17c	10	8	0.8	34	12	0.352
18	6	5	0.833	7	5	0.714
18a	1	1	1.000	7	5	0.714
18b	6	5	0.833	7	5	0.714
18c	1	1	1.000	-	-	-
19	1	1	1.000	13	2	0.153
19a	1	1	1.000	13	2	0.153
20	26	16	0.615	24	19	0.791
20a	26	16	0.615	24	19	0.791
21	10	10	1.000	13	12	0.923
21a	10	10	1.000	13	12	0.923
21b	4	4	1.000	13	12	0.923
21c	1	1	1.000	4	4	1.000
21d	-	-	-	2	2	1.000
22	2	2	1.000	3	3	1.000
22a	-	-	-	-	-	-
22b	1	1	1.000	3	3	1.000
22c	1	1	1.000	-	-	-
23	-	-	-	-	-	-
24	3	1	0.333	8	2	0.25
24a	3	1	0.333	8	2	0.25
24b	2	1	0.5	3	1	0.333
24c	1	1	1.000	1	1	1.000
25	10	10	1.000	22	19	0.863
25a	10	10	1.000	22	19	0.863
25b	6	5	0.833	22	19	0.863
25c	1	1	1.000	18	15	0.833

Nº	INDEXCI			CIFORM		
	R	Rr	P	R	Rr	P
26	-	-	-	-	-	-
27	3	2	0.666	3	2	0.666
27a	3	2	0.666	3	2	0.666
27b	3	2	0.666	2	2	1.000
28	2	-	-	3	2	0.666
28a	2	-	-	3	2	0.666
29	50	14	0.28	81	16	0.197
29a	50	14	0.28	81	16	0.197
29b	10	4	0.4	54	13	0.240
29c	1	-	-	-	2	0.5
29d	2	1	0.5	1	1	1.000
30	-	-	-	3	1	0.333
30a	-	-	-	1	-	-
30b	-	-	-	1	-	-
30c	-	-	-	3	1	0.333
31	1	-	-	4	1	0.25
31a	1	-	-	4	1	0.25
32	6	6	1.000	37	21	0.567
32a	6	6	1.000	33	18	0.545
32b	-	-	-	12	10	0.833
33	10	8	0.8	34	12	0.352
33a	10	8	0.8	34	12	0.352

7. INTERPRETAÇÃO DOS RESULTADOS

7.1. Análise comparativa entre os vocabulários obtidos por indexação automática e manual

Para melhor visualização e interpretação dos resultados elaboraram-se alguns gráficos e figuras, que são demonstrados a seguir.

Na figura 18 encontram-se de forma mais compacta os resultados apresentados nas tabelas 1, 2 e 3, onde se verifica o baixo índice de coincidência das bases de dados indexadas automaticamente e manualmente (23,6%).

Comparando-se os dados entre as tabelas 1 e 2, nota-se que os descritores obtidos através do índice KWIC estão incluídos, de fato, no arquivo INDEXCI, e uma grande maioria no arquivo CIFORM.

A ausência da palavra SERVIÇO(S) no arquivo INDEXCI se deve ao fato de ter sido incluída, erroneamente, esta palavra na lista de raízes e palavras truncadas vazias, WORDROOTS.

Essa total coincidência dos termos do KWIC com o arquivo INDEXCI e a quase total coincidência com o arquivo CIFORM significa que os títulos dos artigos, em língua portuguesa, do periódico Ciência da Informação, apresentam-se com palavras significativas para serem utilizadas como fonte de informação para representar o conteúdo dos documentos e, por consequência, para a indexação automática.

O resultado encontrado é semelhante aos mencionados no item 3.2.2 da revisão de literatura.

Em recente trabalho publicado, DIENER (129) verifi-

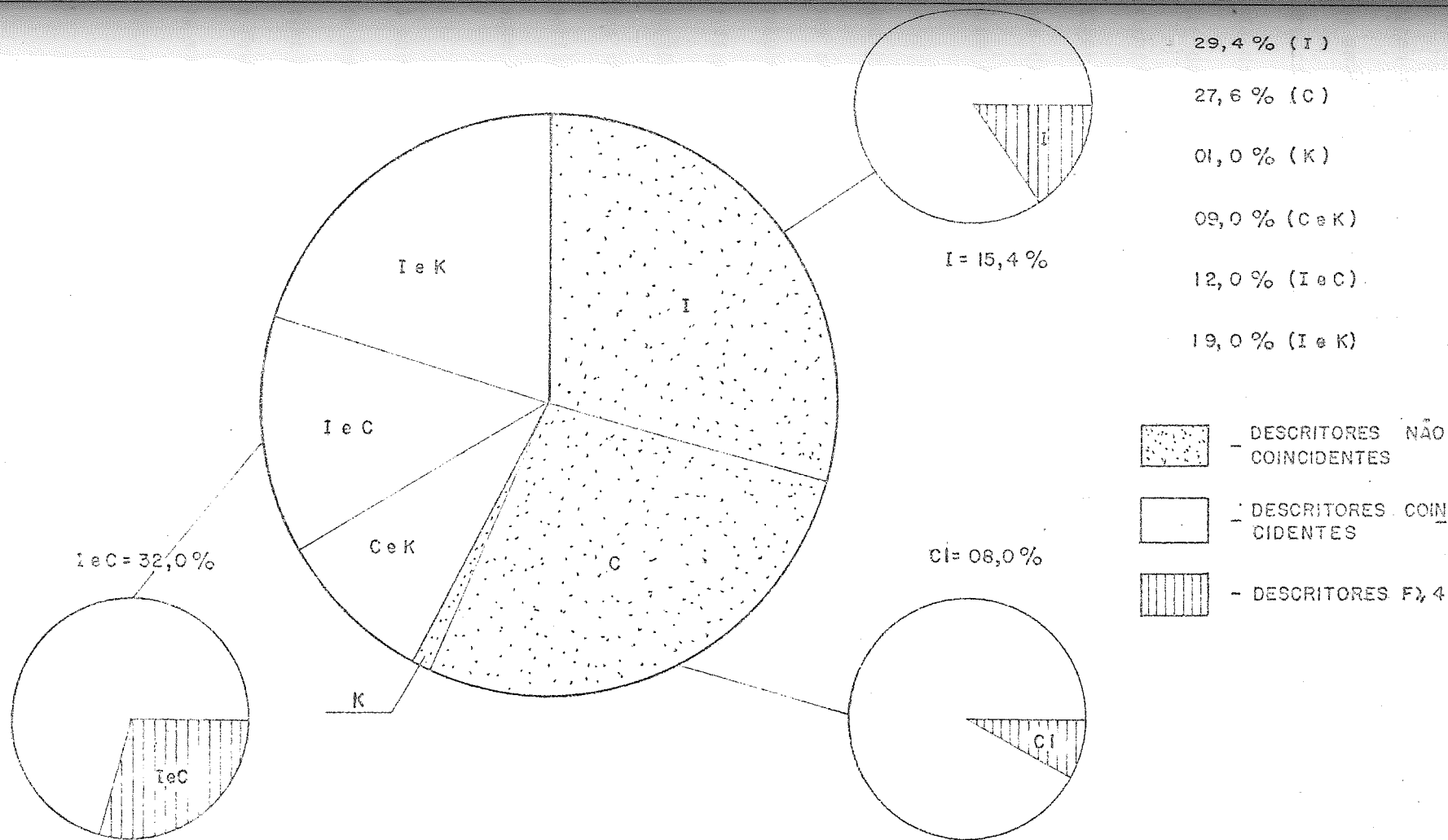


FIGURA 18= DESCRITORES COINCIDENTES E NÃO COINCIDENTES ENTRE AS BASES DE DADOS INDEXCI, CIFORM E O ÍNDICE KWIC

cou que o aumento de palavras significativas em títulos de artigos de periódicos é um fenômeno que cresce linearmente com os anos. Constatou que 62,5% dos substantivos existentes são descritores. Ao se transpor esse resultado para a área de Ciência da Informação, após constar-se nesta pesquisa, que os títulos dos artigos do periódico Ciência da Informação são, também, significativos, pode-se afirmar, então, que os títulos de artigos de periódicos podem ser utilizados, com eficiência, como uma das fontes para a realização de indexação automática na área de Ciência da Informação.

Verifica-se, também com os dados das tabelas 1 e 2, que o uso do resumo aumentou o número de palavras significativas em aproximadamente, 36%, o que torna aconselhável o uso do resumo para a indexação automática. Na revisão da literatura verificou-se (56,94,97) que o uso somente de títulos para indexação automática é menos eficiente para expressar o conteúdo dos documentos que o uso de títulos e resumos, mas em contrapartida alguns autores (9,33,113) constataram que o uso do resumo oferece um efeito negativo ao desempenho da recuperação.

Ao se comparar os resultados apresentados na tabela 3 com os das tabelas 1 e 2 verifica-se claramente o aumento de descritores coincidentes, em I e C, com freqüência de ocorrência maior ou igual a 4, demonstrando-se que, neste trabalho, a freqüência está diretamente relacionada com a coincidência, ou seja, à medida que a freqüência aumenta, aumentam os descritores coincidentes.

Constata-se, através da tabela contendo os vocabulários das duas bases de dados (anexo 2), que os descritores coincidentes são, também, significativos para a representação do conteúdo dos documentos em INDEXCI e CIFORM.

Na figura 19 encontram-se representados os dados das tabelas 4 e 5, onde, se constata claramente, o aumento da coincidência dos termos entre INDEXCI e CIFORM, quando estes se agrupam através de suas raízes e truncagens.

Observa-se que o total de raízes e termos truncados é 57% inferior ao total de descritores na tabela do anexo 2, significando que existem 57% de termos que podem ser considerados redundantes.

Dentre as duas bases de dados a que possui maior redundância é a indexada manualmente, pois apresenta-se com cerca de 53% de descritores compostos e pré-coordenados a mais do que a indexação automática. Esta redundância em CIFORM ocorre porque a pré-coordenação foi normalmente realizada com a repetição de um mesmo descritor, como se verifica nestes exemplos:

- AGEN - agência social, agente de informação, agentes.
- ANALIS - análise de citações, análise do documento, análise documentária, análise estrutural, análise lingüística, análise semântica.
- ARQUIV - arquivo(s), arquivo de ISI, arquivo particular, arquivo PASCAL.
- AUTOR - autores brasileiros, autores científicos, autores coincidentes, autores coletivos, autores individuais, autores ingleses, autores latinos, autores permanentes, autores portugueses, autores transientes, autoria

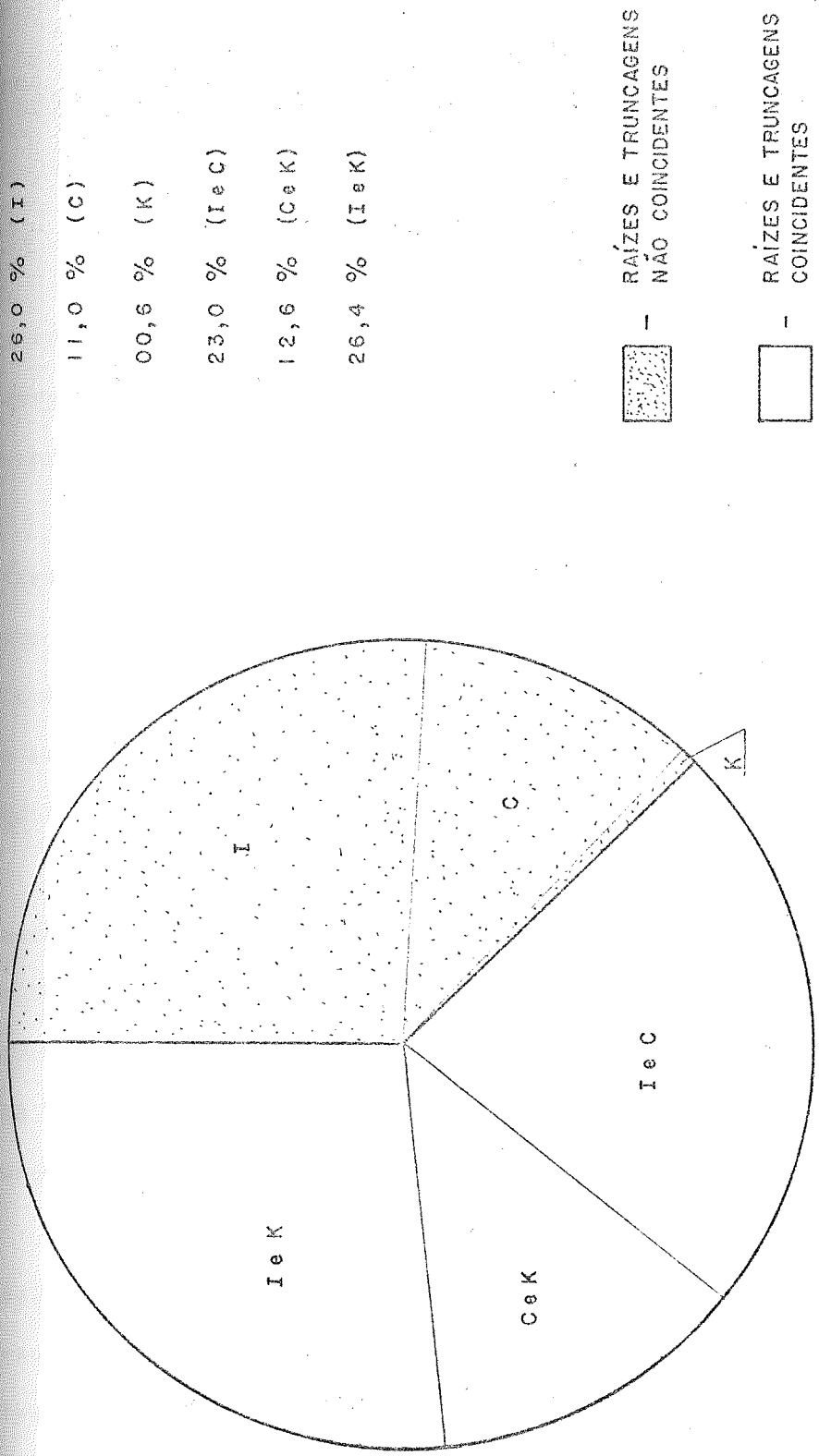


FIGURA 19- RAÍZES E TRUNCAGENS COINCIDENTES E NÃO COINCIDENTES ENTRE AS BASES DE DADOS INDEXCI, CINFOM E O ÍNDICE KWIC.

INDIC

- índice em cadeia, índice KWIC, índice KWOC, índice numérico, índice permutado, índice rotado, índice UNIDEK, índice(s) auxiliar(es), índices permutados.

Os índices de coincidência entre os descritores dos arquivos INDEXCI e CIFORM, nas tabelas 3 e 4, é o mesmo 43%. Pode-se observar que este índice é consideravelmente superior ao apresentado na tabela 1 (23,6%).

Isto confirma, de alguma maneira, a noção de que a frequência e a coincidência de ocorrência do termo nas duas bases estão relacionadas com o valor significativo do termo, já que a raiz de uma palavra ou a sua truncagem representam todas as variações morfológicas dessa palavra, em um só conceito significativo.

A figura 20 demonstra as curvas resultantes da representação dos dados existentes nas colunas A e B das tabelas 6, 7 e 8. No eixo X figura a frequência de ocorrência de termos, e no eixo y o número de termos que ocorrem em determinada frequência em CIFORM, INDEXCI e na lista de raízes e truncagens dos descritores do arquivo INDEXCI.

Nota-se que o ponto de inflexão, nas duas curvas, é próximo à frequência 4, e que abaixo dessa frequência aumenta-se consideravelmente o número de termos. Na frequência 1 a curva torna-se uma reta quase paralela ao eixo Y (assíntota). Em contrapartida, nas frequências acima de 10 ocorrem poucos termos, o que também acarreta uma aproximação da curva ao eixo X.

De uma certa forma os vocabulários de INDEXCI e

CINFORM

RAÍZES DE INDEXCI

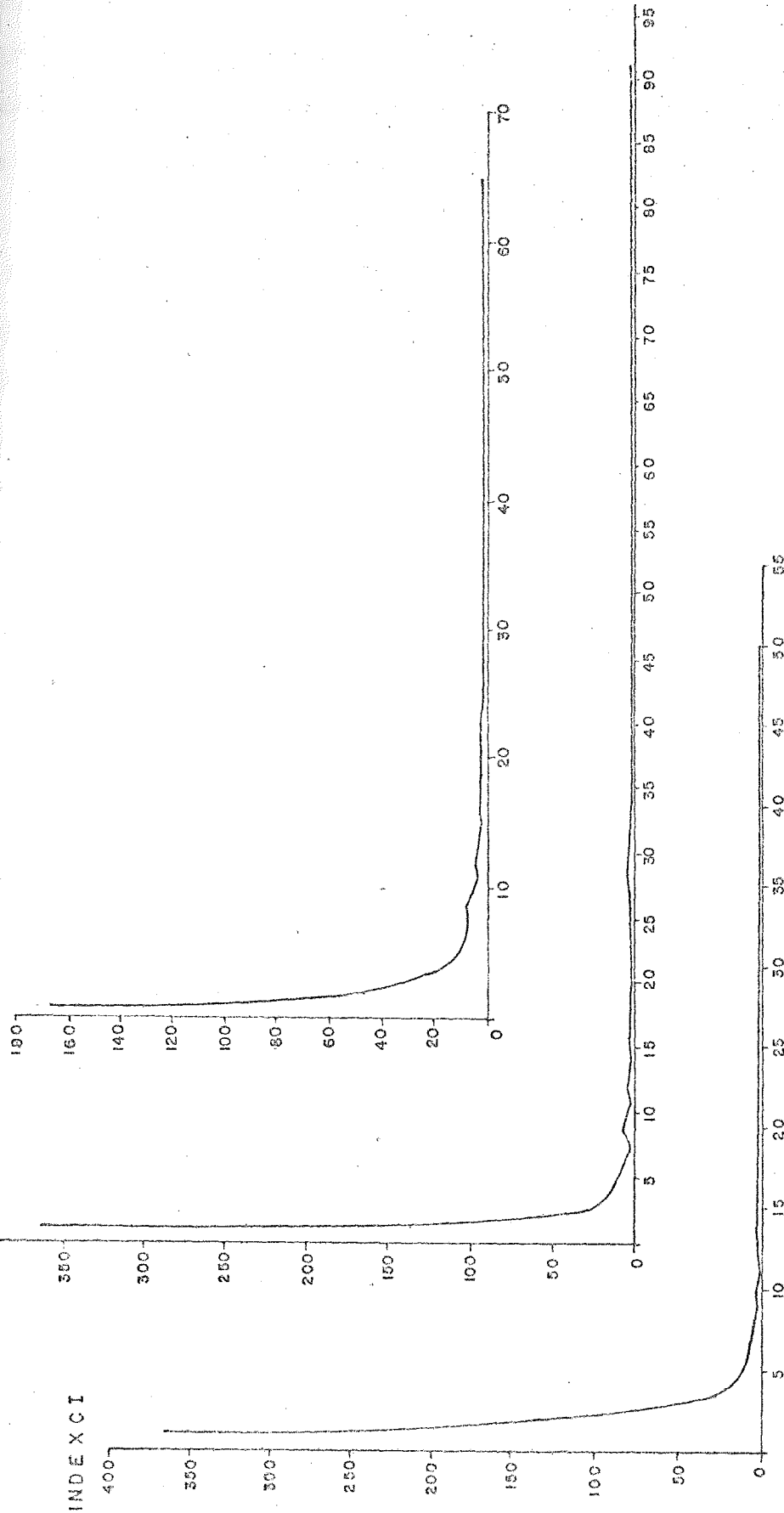


FIGURA 20 - DESCRITORES EM FUNÇÃO DA FREQUÊNCIA NAS BASES DE DADOS INDEXCI, CINFORM E NAS RAÍZES E TRUNCAGENS DE INDEXCI.

CINFORM comportam-se de maneira semelhante pois ambas as curvas apresentam o mesmo aspecto entre as frequências 4 e 10, indicando que entre estas frequências encontra-se o núcleo de termos coincidentes e significativos nas duas bases de dados.

Abaixo da frequência 4 encontram-se os termos que podem possuir significado próximo (ex.: autores brasileiros - autores latinos, índice KWIC - índice KWOC) e / ou serem muito específicos (ex.: ABNT, ASCA, ARIST, análise sintática do termo, autores latinos). Acima da frequência 10 localizam-se os macrodescritores que, de fato, podem ser redundantes com o assunto coberto pela base de dados (ex.: informação (ões), Ciência da Informação, Biblioteconomia).

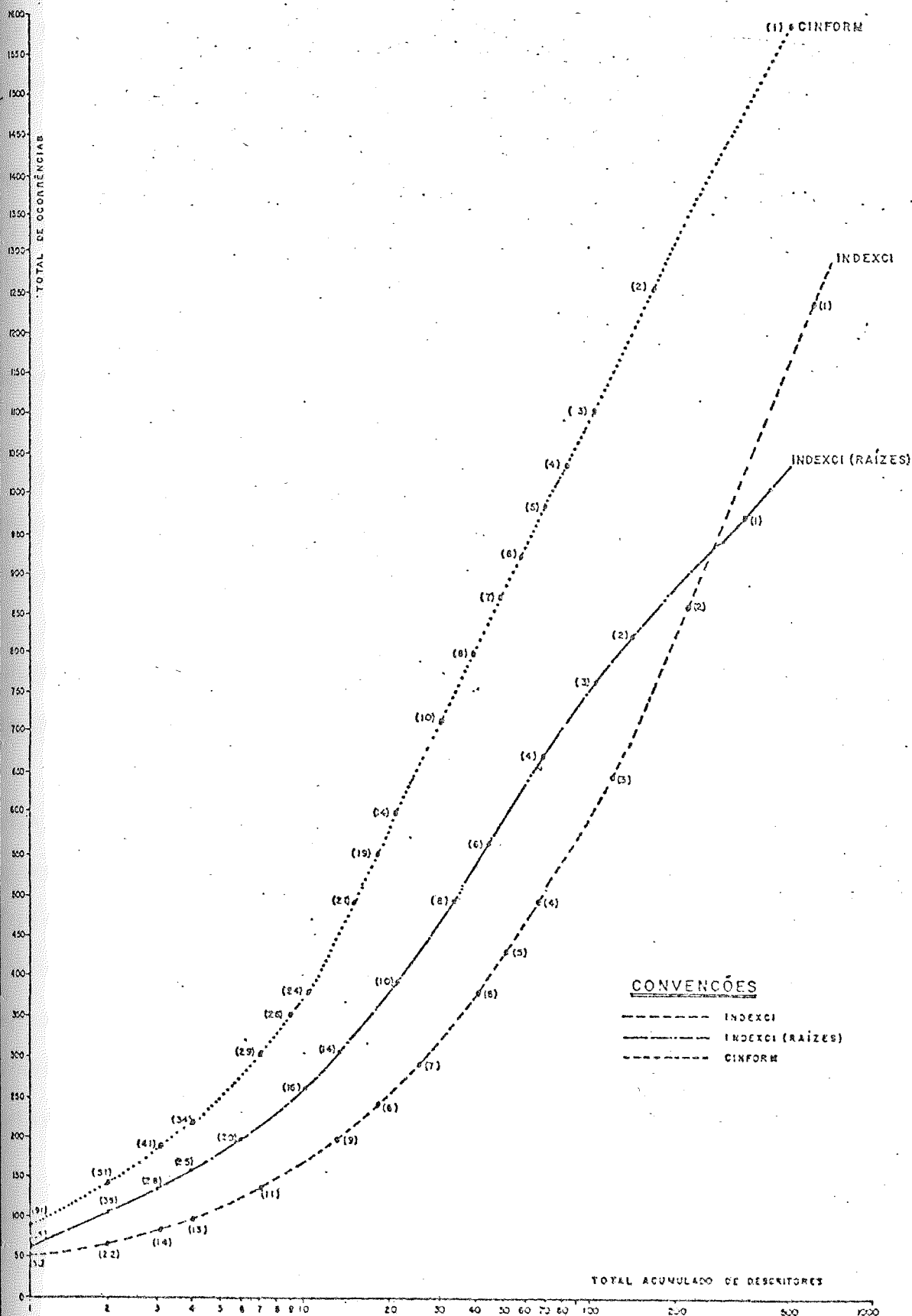
A figura 21 representa graficamente a lei de Bradford, aplicada aos vocabulários de INDEXCI, CINFORM e na lista de raízes e palavras truncadas de INDEXCI. No eixo das abscissas encontra-se o total acumulado de ocorrências (coluna ΣAXB) e no eixo das ordenadas figura o total acumulado de descritores (coluna ΣB). Esse tipo de gráfico já foi delineado anteriormente por ROBREDO (87) de acordo com adaptação à formulação de BROOKES (26).

Demonstra-se a diferença de comportamento dos vocabulários das duas bases e da lista de raízes e palavras truncadas de INDEXCI.

A dispersão em CINFORM caracteriza a redundância dos termos pré-coordenados e a existência de várias sinônimas. Em INDEXCI a dispersão deve-se ao significativo número de termos muito específicos, tais como siglas, datas, nomes de instituições, de países, que estão relacionados com a área de Ciência

FIGURA 21- REPRESENTAÇÃO EM ESCALA SEMILOGARÍTMICA DA VARIÇÃO DO NÚMERO DE OCORRÊNCIAS DOS DESCRITORES EM FUNÇÃO DO NÚMERO DE DESCRITORES IDENTIFICADOS, NAS BASES DE DADOS CIFORM E INDEXCI .

(OS NÚMEROS ENTRE PARENTESES INDICAM A FREQUÊNCIA CORRESPONDENTE DOS DESCRITORES)



da Informação. Deve-se, também, à existência de redundâncias através das variações morfológicas dos termos (ex.: autor e autoria) e à ocorrência de algumas palavras pouco significativas que não foram eliminadas através do filtro de qualidade, mas que facilmente podem ser acrescentadas às listas WORDFIXED ou WORDROOTS.

Na curva da tabela de raízes e palavras truncadas de INDEXCI verifica-se que praticamente não existe dispersão, pois o radical e o truncamento eliminam as redundâncias e as variações morfológicas das palavras na base de dados INDEXCI.

Verifica-se, novamente, a concentração e dispersão de termos nas frequências elevadas e baixas.

Utilizando-se terminologia adotada por ROBREDO (87), os termos que se localizam na parte inicial da curva, com frequência alta, são os descritores de escopo, que caracterizam subáreas da Ciência da Informação ou as categorias desta área, como por exemplo os descritores análise da informação, automação, informação científica, Documentação, Biblioteconomia, transferência de informação e tantos outros encontrados no anexo 2.

Os termos que se situam na parte central da curva, entre as frequências 10 e 4, são os descritores de facetas, que representam assuntos mais específicos, tais como análise bibliométrica, bibliotecas especializadas, catalogação, intercâmbio de informações, estudo de usuários e outros. São os descritores mais indicados para selecionar com rapidez e precisão documentos relevantes em uma busca bibliográfica, pois são eles que melhor caracterizam o conteúdo dos referidos documentos.

Em alguns casos os descritores de escopo podem se confundir com os descritores de facetas, dependendo da área de abrangência do assunto a ser pesquisado e do grau de revocação de

sojado.

Na última parte da curva encontram-se os descritores pontuais, que são numerosos e possuem frequência abaixo de 3. Esses descritores são os maiores responsáveis pela dispersão e caracterizam as sinonímias, quase - sinonímias, as variações morfológicas, assim como descritores com pouco valor significativo, que devem ser eliminados mediante um aprimoramento das listas utilizadas como filtros. Mas esses descritores representam, também, os descritores de alta especificidade e relevância.

Com estes dados pode-se afirmar que a indexação automática, aplicada aos títulos e resumos dos 144 registros bibliográficos existentes na base de dados, constituída pelos artigos publicados no periódico Ciência da Informação entre 1972 e 1983, identificou, de maneira equivalente à da indexação manual, os termos significativos que caracterizam essa base, o que confirma a primeira hipótese.

7.2. Análise comparativa da qualidade de recuperação em linha entre a base de dados indexada automaticamente e manualmente.

Para melhor comparação entre a precisão e o ruído obtidas através de buscas bibliográficas em linha, realizadas nas bases de dados INDEXCI e CIFORM, elaborou-se o teste estatístico Mann - Whitney U - Wilcoxon com objetivo de calcular o índice médio de precisão das duas bases. Para a primeira encontrou-se o valor 0,713 (ou 71,3%) e para a segunda 0,577 (ou 57,7%). Os dados referentes aos testes estatísticos en -

contram-se no anexo 5 e 6. Esses valores foram calculados tomando-se como base o resultado final de recuperação nas 33 buscas realizadas.

A precisão e o ruído são inversos, significando, que o índice médio de ruído apresentado em INDEXCI é 0,287 (ou 28,7%) e em CIFORM 0,423 (ou 42,3%)

Como se pode observar na figura 22 o índice médio de precisão apresentado em INDEXCI é significativamente superior ao encontrado em CIFORM. Isto se explica pelo próprio vocabulário existente em uma base e na outra.

Demonstrou-se, anteriormente, que INDEXCI apresentou menor redundância, o que acarretará, por consequência uma recuperação mais precisa.

A redundância em CIFORM, em contrapartida, fornecerá um número maior de referências recuperadas, mas que não significam, necessariamente, serem relevantes.

Proporcionalmente ao total de referências recuperadas, o número de referências relevantes em INDEXCI foi significativamente superior ao de CIFORM, que apresentou 50,4% de referências irrelevantes recuperadas.

Verificou-se, também, no item 7.1., que a compatibilidade de termos entre ambas é relativamente baixa, havendo maior coincidência em termos com frequência de ocorrência acima de 3. Isto acarreta diferentes resultados na recuperação, pois nem sempre um termo escolhido para a formulação da estratégia de busca em uma base é o mais adequado para a outra.

O conhecimento das diferenças entre os vocabulários

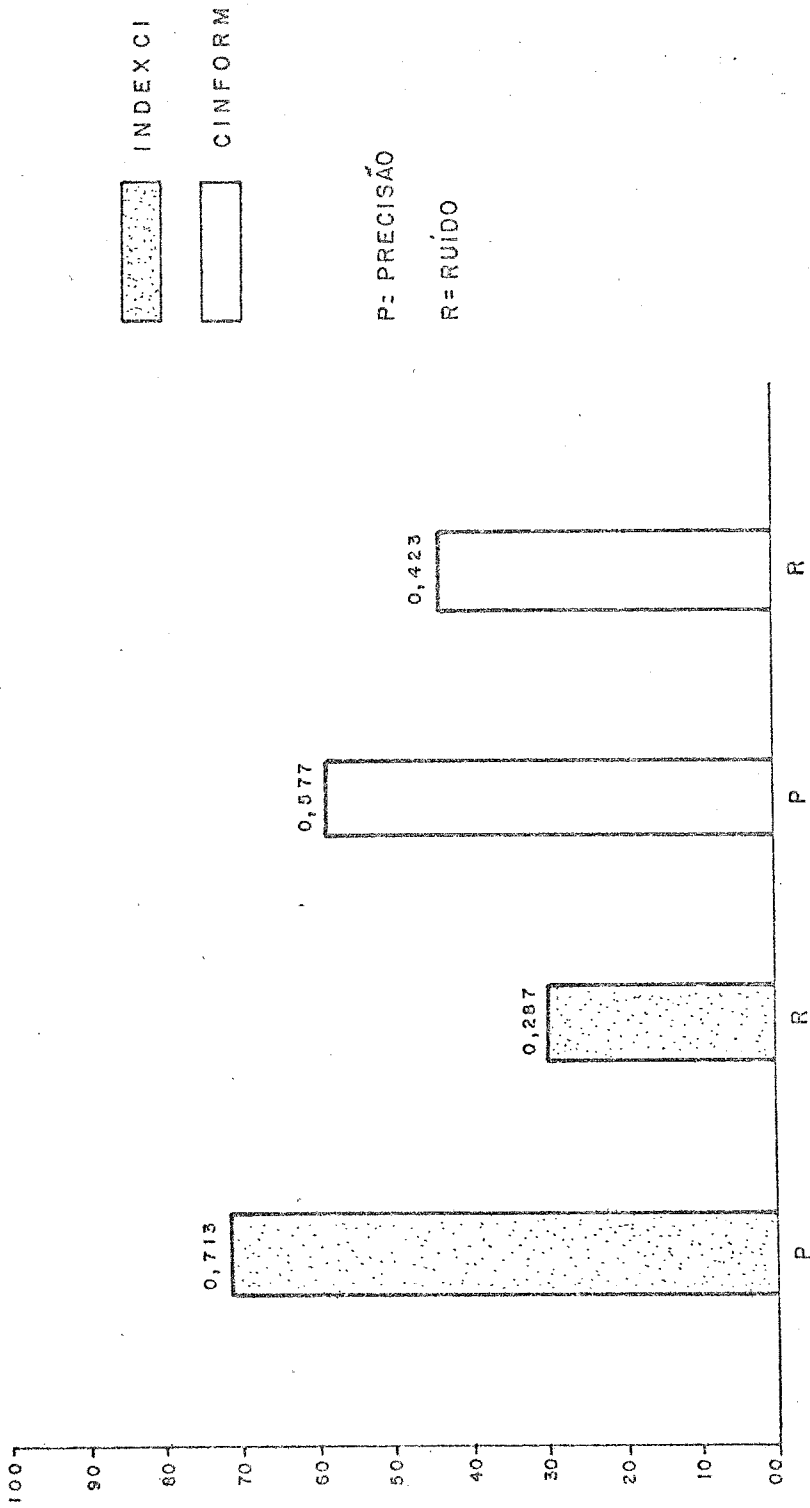


FIGURA 22- ÍNDICES DE PRECISÃO E RUIDO APRESENTADOS NAS BASES DE DADOS INDEXCI E CINFORM.

é importante para selecionar precisamente o descritor e formular-se a estratégia de busca. BATES (130) inclusive menciona este aspecto como um dos problemas encontrados em buscas em linha, isto é tão significativo que, segundo ela, alguns sistemas e serviços de informação criam instrumentos para minimizar esse problema.

A pré-coordenação existente no vocabulário da base de dados CINFORM é um fator que parece que deveria tornar a recuperação mais precisa, pois torna o descritor mais específico. Nesta pesquisa a pré-coordenação foi um fator que dificultou a elaboração da estratégia de busca, apresentando ruídos no resultado final. É possível que se a indexação manual tivesse sido realizada através de um tesouro para controlar a pré-coordenação, o resultado seria, provavelmente, um pouco diferente.

O vocabulário formado pela indexação automática foi mais adequado para a busca em linha, pois os termos ofereceram mais flexibilidade na formulação da estratégia, através da lógica booleana, realizando-se a coordenação adequada e desejada entre os descritores, no momento da pergunta e de acordo com o assunto.

Isto indica que, no caso deste trabalho, o uso da combinação de termos livres e simples é mais eficiente para a recuperação em linha do que o uso de vocabulários pré-coordenados.

Este resultado foi o mesmo a que chegaram o projeto CRANFIELD II (10) e as experiências de SALTON (97) e ROBREDO (87).

A precisão em INDEXCI apresentou-se, geralmente, superior à de CINFORM em perguntas gerais (com operador booleano "/") e em perguntas específicas (com operador booleano "+").

Em algumas buscas estabeleceram-se várias estratégias

na tentativa de se oferecer um resultado positivo. Na busca número 6, por exemplo, as combinações entre os descritores Bibliografia e Ciência e Tecnologia resultaram negativamente. Foi necessário, então, recorrer a um nível mais específico de descritores para se obter um resultado positivo.

Ocorreu falsa recuperação em somente uma busca de INDEXCI. Isto deve-se ao uso de palavras ambíguas e homônimas no resumo ou no título. Como no caso da busca sobre estágio de estudantes de Biblioteconomia, todas as referências recuperadas foram irrelevantes, pois esta palavra foi utilizada como sinônimo de "situação atual", que não expressa o conceito desejado.

Outra falsa recuperação poderia ter ocorrido caso fosse sugerida uma busca relacionando a Empresa Brasileira de Correios e Telégrafos (ECT) com por exemplo, custo de tarifas postais de envio de cópias, pois a sigla ECT existente na lista de descritores obtidos por indexação automática não corresponde ao conceito da sigla, porque foi um erro de digitação da palavra ETC.

Observa-se então que, para tornar a indexação automática mais eficiente, deve-se adotar algumas diretrizes simples no sentido de se evitar palavras ambíguas e erros de digitação no título e resumo ..

O resultado final negativo, da busca bibliográfica em ambas as bases, indica a inexistência do termo nos respectivos vocabulários. Quando negativo em INDEXCI, pode indicar que, em algumas buscas, as palavras, utilizadas como descritores, estão erroneamente, incluídas nas listas de palavras não significativas ou então, hifenadas de forma incorreta.

Os resultados demonstram que o bom desempenho, tanto

em nível de precisão como de revocação, está intrinsecamente relacionado com a seleção correta do descritor para identificar o conteúdo dos documentos.

Nesta pesquisa a indexação automática, em termos de precisão, mostrou um desempenho superior ao da indexação manual, o que confirma a segunda hipótese.

8. CONCLUSÃO

A indexação automática aplicada aos títulos e resumos dos artigos do periódico Ciência da Informação, entre 1972 e 1983, identificou de maneira equivalente à da indexação manual, os descritores que caracterizam a base de dados formada com os referidos artigos.

Os dois tipos de indexação apresentaram, basicamente as mesmas características em seus vocabulários, quando aplicada a análise das frequências dos descritores.

Constatou-se ao comparar os vocabulários das bases CIFORM (indexada manualmente) e INDEXCI (indexada automaticamente), que na primeira encontram-se descritores com frequência de aparecimento superiores aos de INDEXCI, o que parece que contribuiu para aumentar a redundância no vocabulário e o ruído na recuperação. Encontrou-se, também, maior quantidade de sinônimas e quase - sinônimas na base CIFORM.

O índice de coincidência entre os descritores das bases INDEXCI e CIFORM foi baixo, ocorrendo maior coincidência com descritores de frequências maiores do que 3, indicando que o aumento da frequência está relacionado com a coincidência.

Demonstrou-se, nesta pesquisa, que os descritores de facetas (média frequência) são os mais significativos para a identificação do conteúdo dos documentos. Estes, combinados com os descritores pontuais (baixa frequência), possibilitam uma melhor recuperação. Os descritores de escopo (alta frequência) permitem uma indexação macrocategorizada. Quando se consideram as raízes dos termos e as palavras truncadas, diminui a dispersão entre os vocabulários, aumentando a coincidência e concentração

dos termos significativos. Isto indica que o uso das raízes e palavras truncadas, também, é aconselhável nas buscas quando se deseja melhores índices de precisão e revocação.

Verifica-se que a análise de frequências das palavras, em língua portuguesa, contribui para os estudos semânticos de vocabulários formados por indexação automática ou manual. Através do número de ocorrências do termo, identifica-se seu nível de especificidade e o seu valor para a recuperação. O uso deste filtro de qualidade, a frequência, pode contribuir para os estudos terminológicos de linguagens de indexação e elaboração de tesouros.

A qualidade de recuperação apresentada na base de dados INDEXCI, medida em termos de precisão, apresentou índice superior ao de CIFORM.

Os termos livres e simples do arquivo INDEXCI apresentaram melhor flexibilidade para a formulação da estratégia de busca. Isto significa que a indexação automática, utilizando linguagem livre e pós-coordenação dos descritores, no momento da recuperação, oferece melhores resultados de recuperação.

O emprego da indexação automática parece aconselhável, face ao grande número de documentos existentes, o que torna cada dia mais difícil realizar a indexação manual com o mínimo de qualidade requerido para assegurar o acesso posterior à informação. Vários estudos estrangeiros já revelaram a sua validade e, neste trabalho, confirma-se a eficiência da indexação automática, quando aplicada aos títulos e resumos em língua portuguesa.

A criação de vocabulários controlados, hierarquiza-

dos ou não, é uma tarefa que requer um tempo considerável de estudos prévios e atualização constante, o que torna esses vocabulários rapidamente defasados em relação ao crescimento da terminologia, na área considerada.

É recomendável que os estudos sobre linguagens de indexação caminhem no sentido de se adotar e aprimorar técnicas de indexação automática, em linguagem livre, realizando-se o controle de vocabulário, ou sua coordenação, no momento da formulação da pergunta e utilizando-se lógicas adequadas a cada sistema de recuperação.

As técnicas de indexação automática se prendem em maior ou menor grau às características dos programas em que se baseiam. Para se obter os melhores resultados possíveis, faz-se mister seguir algumas regras, específicas a cada sistema como por exemplo na preparação dos registros, na conceituação das tabelas de palavras ou raízes não significativas e etc.

No caso específico da técnica de indexação automática no sistema BIB/DIALOGO é aconselhável o desenvolvimento de algumas diretrizes para aprimorar os resultados. Estas diretrizes estariam voltadas para:

- a) inclusão de novas palavras proibidas, detectadas nesta pesquisa e em outras que venham a surgir na área de Ciência da Informação, nas tabelas de controle terminológico;
- b) hifenação correta de alguns grupos de palavras, quando for necessário.

Com estas medidas a eficiência da indexação automática, para a área de Ciência da Informação, no sistema BIB/DIALOGO poderia aumentar, tornando o sistema bastante atrativo

para diversas aplicações.

LANCASTER (59, p. 229) menciona o futuro da indexação e afirma: "Eu realmente acredito que o rumo predominante da recuperação da informação, num futuro não tão distante, será o do aumento contínuo de bases de dados textuais, com a eliminação da técnica de indexação manual e o desenvolvimento de vocabulários controlados a posteriori".

8.1. Sugestões para estudos futuros

Estudos comparativos entre indexação automática (em títulos e resumos) e indexação manual, utilizando tesauros e listas de vocabulários controlado na área de Ciência da Informação e em outras áreas, no que diz respeito a:

- precisão e revocação relacionados com a demanda e satisfação do usuário;
- custo-benefício, relacionado com a precisão e revocação em buscas em linha;
- flexibilidade na formulação das estratégias de busca em diferentes bases de dados.

9. REFERÊNCIAS BIBLIOGRÁFICAS

01. ANDREEWSKY, Alexandre & RUAS, Vitoriano. Indexação automática baseada em métodos linguísticos e estatísticos e sua aplicabilidade à língua portuguesa. Rio de Janeiro, PUC - DI, 1982. 31p.
02. ARAUJO, V.M.R.H. de. Estudo dos canais informais de comunicação técnica: seu papel na transferência de tecnologia e na inovação tecnológica. Ciência da Informação. 8(2):79-100, 1979.
03. BARKER, F.H.; VEAL, D.C. & WYATT, B.R. Comparative efficiency of searching titles, abstracts and index terms in a free-text data base. Journal of Documentation, 28(1):22-36, Mar. 1972.
04. BARNES, C.I.; COSTANTINI, L. & PESCHKE, S. Automatic indexing using the SLC II System. Information Processing and Management, 14(2):107-119, 1978.
05. BARREIRO, S.C. Experiência em indexação do Centro de Informações Nucleares. In: REUNIÃO BRASILEIRA DE CIÊNCIA DA INFORMAÇÃO, 1., Rio de Janeiro, 1975. Anais. Rio de Janeiro, IBICT, 1978, p. 237-45.
06. BATTY, C.D. The automatic generation of index languages. Journal of Documentation, 25(2):142-51, June 1969.

07. BINAGRI. Guia brasileiro de pesquisa agrícola em andamento. Brasília, 1978. 2v. (Projeto PNUD/FAO/BRA/72/020: DOC./TEC. 78/061 e DOC./TEC./ 78062).
08. BIRD, P.R. & KNIGHT, M.A. Word count statistics of scientific papers. Information Scientist, 9(2):67-9, 1975.
09. BLOOMFIELD, M. Evaluation of indexing 2. The Simulated Machine Indexing Experiments. Special Libraries, 61(9):501-7, Nov. 1970.
10. BLOOMFIELD, M. Evaluation of indexing 4. A Review of the Cranfield Experiments. Special Libraries, 62(1):24-9, Jan. 1971.
11. BOOKSTEIN, A. & SWANSON, D.R. A decision theoretic foundation for indexing. Journal of American Society for Information Science, 26(1):45-50, 1975.
12. BOOKSTEIN, A. & SWANSON, D.R. A probabilistic models for automatic indexing. Journal of American Society for Information Science, 25(5):312-8, Sept./Oct.1974.
13. BOOTH, A.D. A "Law" of occurrences for words of law frequency. Information and Control, 10(4):386-93, 1967.
14. BORKO, H. Toward a theory of indexing. Information

Processing and Management, 13(6):355-66, 1977.

15. BOURNE, C.P. Evaluation of indexing systems. Annual Review of Information Science and Technology, 1:171-90, 1966.
16. BOYCE, Bert. & LOCKARD, Marta. Automatic and manual indexing performance in a small file of medical literature. Bulletin of Medical Library Association, 63(4):378-85, Oct. 1975.
17. BRAGA, L.M. Palavras de títulos e resumos como acesso ao conteúdo do documento: uma análise numérica. Rio de Janeiro, URFJ/IBICT, 1982. 181p. (Dissertação).
18. BROOKES, B.C. Bradford's law and the bibliography of science. Nature, 224:953-56, 1969.
19. BUXTON, A.B. & MEADOWS, A. J. The variation in the information content of titles of research papers with time and discipline. Journal of Documentation, 33(1):46-52, Mar. 1977.
20. CARROLL, John M. & ROELOFFS, Robert. Computer selection of keywords using word-frequency analysis. American Documentation, 20(3):227-33, July 1969.
21. CAVALCANTI, Cordélia Robalinho. Indexação & Tesouro; Metodologia e Técnicas. Ed. prelim. Brasília, ABDF, 89p.

22. CAVALCANTI, Cordélia Robalinho. Metodologia de indexação. Brasília, 1976. 8f.
23. CESARINO, M. A. da N. & PINTO, M.C.M.F. Análise de assunto. Revista de Biblioteconomia de Brasília, 8(1): 32-43, jan./jun. 1980.
24. CLEVELAND, D.B., CLEVELAND, A.D. & WISE, O.B. Less than fulltext indexing using a non-boolean searching model. Journal of the American Society for Information Science, 35(1):19-28, 1984.
25. CLEVERDON, C.W. Report on the testing analysis of an investigation into comparative efficiency of indexing systems. Cranfield, College of Aeronautics, ASLIB, Cranfield Research Project, 1962. n.p.
26. CUADRA, C.A. & KATTER, R.V. Experimental studies of relevance judgements. Final Report, v. 1 Project. Summary. Santa Monica, System Development Corp., 1967. (TM-3520 0001/00).
27. DILLON, Martin & FEDERHART, Peggy. The use of discriminant analysis to select content-bearing words. Journal of the American Society for Information Science, 33(4): 245-53, July, 1982.
28. EBINUMA, Y. Consideration for diffusion of keyword assignment in free term to articles in scientific

journals. International Information Communication and Education, 1(1):59-66, Mar. 1982.

29. FARRADANE, J. The testing of relational indexing procedures by diagnostic computer programs. Journal of Information Science, 2(6):285-97, Dec. 1980.
30. FEINBERG, H. Title derivative indexing techniques: a comparative study. Metuchen, Scarecrow Press. 1973. 297p.
31. FIELD, B.J. Towards automatic indexing. London, INSPEC Institution Electrical Engineers, 1975. 74p. (BLLD report 5235).
32. FIGUEIREDO, N. M. de. O contexto de relevância e suas implicações. Ciência da Informação, 6(2):75-8, 1977.
33. FIGUEIREDO, N.M. de. O processo de transferência da informação. Ciência da Informação, 8(2):119-38, 1979.
34. FOSKETT, A. C. The subject approach to information. 3. ed. London, C. Bingley, 1977. 476p.
35. FREUND, George Eduardo. Análise estrutural para aumentar a eficiência de pesquisa online. Ciência da Informação, 11(1):19-26, 1982.
36. FUGMANN, R. On the practice of indexing and its theoretic-

tical foundations. International Classification, 7
(1):13-20, Apr. 1980.

37. GARFIELD, E. A conceptual review of citation indexing.
In: _____. Citation indexing — its theory and
application in Science Technology, and Humanities. New
York, John Wiley and Sons, 1978. p.1-5
38. GARFIELD, E. A weekly subject index for Current Contents
/Life Sciences. In: ANNUAL MEETING OF THE MEDICAL LI
BRARY ASSOCIATION, 71., San Diego, June 11-15, 1972.
39. GHOSH, Jata S. Content representation in document titles:
a case study with prostaglandin literature. Aslib
Proceedings, 26(2):83-6, Feb. 1974.
40. GHOSH, Jata S. The information content of titles in
contraception literature. Journal of Chemical Infor
mation Computer Science, 17 (1):36-40, Feb. 1977.
41. GOFFMAN, W. A general theory of communication. In:
SARACEVIC, T. Introduction to information science.
New York, Bowker, 1970. p.726-47.
42. HALLER, Johann. Análise automática de textos em siste
mas de informação. Revista de Biblioteconomia de
Brasília, 11(1):105-113, jan./jun. 1983.

43. HALLER, Johann. Processamento de textos em linguagem natural. In: CONGRESSO NACIONAL DE INFORMÁTICA, 15., Rio de Janeiro, out. 1982. (Trabalhos apresentados). Rio de Janeiro, 1982. 9p.
44. HAMILL, Karen A. & ZAMORA, Antonio. The use of titles for automatic document classification. Journal of the American Society of Information Science, 31(6):396-402, Nov. 1980.
45. HARPER, D. J. & VAN RIJSBERGEN, C. J. An evaluation of feedback in document retrieval using co-occurrence data. Journal of Documentation, 34(3):189-216, 1978.
46. HARTER, Stephen P. A probabilistic approach to automatic Keyword indexing. Journal of the American Society of Information Science, 26(4):197-206, july/Aug.1975; 26(5):280-289, Sept./Oct. 1975.
47. HENZLER, R. G. Free or controlled vocabularies: some statistical user-oriented evaluations of biomedical information systems. International Classification, 5(1):21-6, Mar. 1978.
48. HUTCHINS, W. J. Languages of indexing and classification: a linguistic study of structures and functions. Stevenage, Peter Perenigrus Ltd., 1975. 148p.

49. JONES, Kevin P. How do we index a report of some Aslib Informatics Group Activity. Journal of Documentation, 39(1):1-23, Mar. 1983.
50. JONES, P. E.; GIULIANO, V.E. & CURTICE, R. M. Papers on automatic language processing - linear models for associative retrieval. Report ESD-TR-67-202. Ad Little, Inc., Cambridge, 1967. v.2.
51. JONKER, F. Indexing theory, indexing methods and research services. New York, Scarecrow Press, 1964. 124p.
52. KLINGBIEL, Paul H. Machine-aided indexing of technical literature. Information Storage and Retrieval, 9(2): 79-89, Feb. 1973.
53. KLINGBIEL, Paul H. & RINKER, Catherine C. Evaluation of Machine-Aided Indexing. Information Processing and Management 12(6):351-66, 1976.
54. KNIGHT, G. N. Treinamento em indexação: um curso da Society of Indexers. Rio de Janeiro, FGV, 1974. 216p.
55. KRAFT, D. H. A comparison of keyword in context (KWIC) indexing of titles with a Subject Heading Classification System. American Documentation, 15(1):48-52, Jan. 1964.

56. KWOK, K. L. Cited titles: a new source of keyword extraction for automatic document classification and retrieval. In: ASIS ANNUAL MEETING, 37, Atlanta, 13-17 Oct., 1974. Proceedings. Washington, ASIS, 1974. v.11, p.56-57.
57. KWOK, K.L. The use of title and cited titles as document representation for automatic classification. Information Processing and Management, 11(8/12):201-6, 1975.
58. LANCASTER, F. W. Evaluation of the operating efficiency of Medlars: final report. Bethesda, National Library of Medicine, 1968.
59. LANCASTER, F. W. Trends in subject indexing from 1957 to 2000. In: FID CONGRESS, 39., Edinburgh, 25-28 September 1978. New trends in documentation and information. London, Aslib, 1980. p.223-233.
60. LEONARD, L. E. Inter-indexer consistency studies, 1854 - 1975: a review of the literature and summary of study results. Illinois, University of Illinois, Graduate School of Library Science, 1977. 51p. (Occasional papers).
61. LESK, M. E. Word-word associations in document retrieval systems. In: CORNELL UNIVERSITY. Department of Computer Science. Report n° ISR-13 to the National

Science Foundation. Ithaca, N.Y. Jan. 1968. Section 9.

62. LUHN, H. P. The automatic creation of literature abstracts. IBM Journal of Research and Development, 2:159-165, 1958.
63. LUHN, H. P. A statistical approach to mechanized encoding and searching of literary information. IBM Journal of Research and Development, 1(4):309-317, Oct. 1957.
64. MACHADO, Norma & HAMAR, Alfredo A. Sistema de arquivamento e indexação por computador, do acervo de programas de um Centro de Processamento de Dados. In: CONGRESSO REGIONAL DE DOCUMENTAÇÃO, 2., REUNIÃO FID/CLA, 9., Rio de Janeiro, 1969. Anais. Rio de Janeiro, IBBD, 1970. p.237-41.
65. MAIA, E. L. e S. Comportamento bibliométrico da língua portuguesa como veículo de representação da informação. Ciência da Informação, 2(2):99-138, 1973.
66. MAIZELL, R. Value of titles for indexing purposes. Revue de la Documentation, 27:126-7, 1960.
67. MARCHESI, I. H. George Anderla e a informação em 1985. Ciência da Informação, 6(1):45-48, 1977.

68. MARON, M. E. & KUHNS, J. L. On relevance, probabilistic indexing and information retrieval. Journal of the Association for Computing Machinery, 7(3):216-43, 1960.
69. MONTGOMERY, C. & SWANSON, D. R. Title indexing. American Documentation, 13:359-64, 1962.
70. NEUFELD, M. L. et alii. Automatic title word indexing for a weekly current awareness service. In: ANNUAL MEETING OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE, 36, Los Angeles, October 21-25, 1973. v.10. Innovative Development in Information Systems: their benefits and costs. Ed. by Helen J. Waldron & F. Raymond Long. Washington, DC, ASIS and Westport, Conn., Greenwood Press, 1973. p.167-8.
71. OLIVEIRA, Elvia A. Automação da Bibliografia Brasileira de Ciências Sociais. In: CONGRESSO REGIONAL DE DOCUMENTAÇÃO, 3., REUNIÃO FID/CLA, 11., Lima, 1972. Anais. Rio de Janeiro, IBBD, 1972. p.59-61.
72. PAO, Miranda Lee. Automatic text analysis based on Goffman's transition phenomena of word occurrences. Journal of the American Society Information Science, 29(3):121-4, May. 1978.
73. PARKER, Lorraine M. Purgailis. Towards a theory of document learning. Journal of the American Society

for Information Science, 34(1):16-21, Jan. 1983.

74. PINHEIRO, Lena Vania Ribeiro. Medidas de consistência da indexação: interconsistência. Ciência da Informação, 7(2):109-14, 1978.
75. QUEIROZ, Mucio G. S. Um estudo comparativo de processos estatísticos para obtenção automática de resumos. Rio de Janeiro, PUC, 1973. (Dissertação).
76. REES, A. M. et alii. A field experimental approach to the study of relevance assessment in relation to document searching. v.2. Center for Documentation and Communication Research, Case Western Reserve University, Cleveland, Ohio, 1967. 475p.
77. REGAZZI, J. J. Evaluating indexing systems: a review after Cranfield, The Indexer, 12(1):14-21, Apr. 1980.
78. RESNICK, A. Relative effectiveness of document titles and abstracts for determining relevance of documents. Science, 134 (3484): 100-6, 6. Oct. 1961.
79. RIBEIRO, L. A. Aplicação dos métodos estatísticos e da teoria da informação e da comunicação na análise linguística: estudo da linguagem jornalística. Ciência da Informação, 3(2):151-4, 1974.

10. ROBERTSON, S. E. & SPARCK JONES, J, Relevance weighting of research terms. Journal of the American Society for Information Science, 27:129-146, 1976.
11. ROBREDO, J. BIB/Batch e BIB/Dialogo. In: UNESCO. International Inventory of Software Packages in information field. C. Keren, I. Sered, eds. Paris, 1983. p.119-24. (PGI/UNISIST, PGI/83/WS/28).
12. ROBREDO, J. BIB/Dialogo; descrição do sistema. Brasília, UnB, 1983. 28f.
13. ROBREDO, J. Documentação de hoje e de amanhã. Brasília, ABDF, 1976. viii, 172p.
14. ROBREDO, J. Dois novos sistemas com computadores para o processamento completo da informação documentária. Revista de Biblioteconomia de Brasília. 10(7):69-72, jan. 1982.
15. ROBREDO, J. A indexação automática como mecanismo básico no processo de transferência da informação. In: CONGRESSO LATINO-AMERICANO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 1., Salvador, 21-26 set., 1980. Anais, 19p.
16. ROBREDO, J. A indexação automática de textos: o presente já entrou no futuro. In: Machado, U. D., ed. Estudos Avançados em Biblioteconomia e Ciência da Informação. Brasília, ABDF, 1982. v.1, p.236-74.

87. ROBREDO, J. Otimização dos processos de indexação dos documentos e recuperação da informação mediante o uso de instrumentos de controle terminológico. Ciência da Informação, 11(1):3-18, 1982.
88. ROBREDO, J. & FERREIRA, J.A. de P. Conceituação de um programa para indexação automática de textos. Revista de Biblioteconomia de Brasília, 8(2):254-63, jul./dez., 1980.
89. ROBREDO, J. et alii. Construção de um núcleo de thesaurus em agricultura baseado no uso real dos descritores. In: REUNIÃO BRASILEIRA DE CIÊNCIA DA INFORMAÇÃO, 1., Rio de Janeiro, 1975. Anais. Rio de Janeiro, IBICT, 1978. v.1, p.289-303.
90. ROBREDO, J. et alii. Elaboración de um thesaurus agrícola baseado en critérios de eficiência del lenguaje en el proceso de comunicación. Brasília, SNIDA, 1975. 23p.
91. ROSENBERG, Victor. A study of statistical measures for predicting terms, used to index documents. Journal of American Society for Information Science, 22(1): 41-50, Jan./Feb. 1971.
92. ROWBOTTOM, Mary E. & WILLET, Peter. The effect of subject matter on the automatic indexing of full text. Journal of the American Society for Infor-

mation Science, 33(3):139-41, May 1982.

93. RUHL, M. J. Chemical documents and their titles: human concept indexing vs KWIC - Machine indexing. American Documentation, 15(2):136-41, Apr. 1964.
94. SALTON, G. Automated language processing. Annual Review of Information Science and Technology, 3:169-99, 1968.
95. SALTON, G. Automatic information systems. In: _____. Automatic information organization and retrieval. New York, Mac Graw-Hill Book Company. p.1-20.
96. SALTON, G. Automatic text analysis: automatic document indexing and classification methods are examined and their effectiveness assessed. Science, 168(3929):335-43, 17 Apr. 1970.
97. SALTON, G. A comparison between manual and automatic indexing systems. Computing Reviews, 10(6):274, June. 1969.
98. SALTON, G. A new comparison between conventional indexing and automatic text processing. Journal of the American Society of Information Science, 23(2):75-84, Mar./Apr. 1972.

99. SALTON, G.; WU, H. & YU, C.T. The measurement of term importance in automatic indexing. Journal of the American Society for Information Science, 32(3): 175-86, May 1981.
100. SALTON, G. & YANG, C.S. On the specification of term values in automatic indexing. Journal of Documentation, 29(4):351-72, Dec. 1973.
101. SALTON, G.; YANG, C. S. & YU, C. T. A theory of term importance in automatic text analysis. Journal of the American Society for Information Science, 26(1):33-34, Jan./Feb. 1975.
102. SARACEVIC, Tefko. Comparative effects of titles, abstracts and full texts on relevance judgments. In: ASIS ANNUAL MEETING, 32., San Francisco, Oct., 1-4, 1969. Cooperating information societies. Westport., Greenwood Publishing Corporation, 1969. v.6, p. 285-99.
103. SARACEVIC, Tefko. Relevance: a review of and a framework for thinking the notion in information science. Journal of American Society for Information Science, 26(6):321-43, Nov./Dec. 1975.
104. SCHUEGRAF, Ernest. & HEAPS, I. Indexing for associative processing. Canadian Journal of Information Science, 5:93-101, May 1980.

105. SIMMONS, R.F. Automated language processing. Annual Review of Information Science and Technology, 1:137-69, 1966.
106. SOERGEL, Dagobert. Automatic and semi-automatic methods as an aid in the construction of indexing languages and thesauri. International Classification, 1(1):34-9, May 1974.
107. SOERGEL, Dagobert. Indexing languages and thesauri: construction and maintenance. Los Angeles, Melville, 1974. 632p.
108. SOUZA, Eliane Santos. Estudo dos títulos de artigos de periódicos da área de Biblioteconomia e Ciência da Informação na década pós KWIC: 1960 a 1970. Ciência da Informação, 7(2): 115-7, 1978.
109. SPARCK JONES, Karen. Indexing term weighting. Information Storage and Retrieval, 9(11):619-33, Nov. 1973.
110. SPARCK JONES, Karen. The role of automatic indexing in operational online retrieval systems. In: FID CONGRESS, 39., Edinburg, 25-28 September 1978. New trends in documentation. London, Aslib, 1980. p. 33-8.
111. SPARCK JONES, Karen. A statistical interpretation of term specificity and its application in retrieval.

Journal of Documentation, 28(1):11-21, Mar. 1972.

112. STEINACKER, Ivo. Indexing and automatic significance analysis. Journal of American Society for Information Science, 25(4):237-41, July/Aug. 1974.
113. SVENONIUS, Elaine. An experiment in index term frequency. Journal of the American Society for Information Science, 23(2):109-21, Mar./Apr. 1972.
114. SWANSON, D. R. The evidence underlying the Cranfield results. Library Quarterly, 35(1):1-20, Jan. 1965.
115. TANIMOTO, T. T. An elementary mathematical theory of classification and predication. IBM, 1958. n.p.
116. TEIXEIRA, I. L.R. Sistema KWIC no controle da documentação especializada. In: CONGRESSO REGIONAL SOBRE DOCUMENTAÇÃO, 2.; REUNIÃO DA FID/CLA, 9. Rio de Janeiro, IBBD, 1970. p.270-7.
117. TOCATLIAN, J. J. Are titles of chemical papers becoming more informative? Journal of the American Society for Information Science, 21:345-50, 1970.
118. TORRES FILHO, Paulo Roberto Pinheiro. Um sistema semi-automático para o apoio à indexação de documentos técnicos. Rio de Janeiro, PUC, 1983. 82p. (Dissertação).

119. THE UNISIST draft on indexing principles: test and comments. International Classification, 4(1):29-34, May 1977.
120. VAN DER MEULEN, W. A. & JANSSEN, P. J. F. C. Automatic versus manual indexing. Information Processing and Management, 13(1):13-21, 1977.
121. VAN RIJSBERGEN, C. J. A theoretical basis for the use of co-occurrence data in information retrieval. Journal of Documentation 33(2):106-19, June 1977.
122. VICKERY, B.C. Structure and function in retrieval languages. Journal of Documentation, 27(2):69-82, June 1971.
123. WEISMAN, H. M. Information systems, services and centers. New York, Becker & Hayes, 1972.
124. WESSEL, A. E. Free-text indexing the problems and paradoxes. In: _____. Computer/aided information retrieval. Los Angeles, Melville Publishing Company, 1975. p.11-8.
125. WESSEL, A. E. Indexing and analysis of information-some preliminary computs. In: _____. Computer/aided information retrieval. Los Angeles, Melville Publishing Company, 1975. p.1-10.

126. YU, C. T. & SALTON, G. Precision wighting-an effective automatic indexing method. Journal of Association for Computing Machinery, 23:76-88, 1976.
127. ZAHER, C.R. & DUARTE, Y. C. Sistema KWIC versus descritores. In: CONGRESSO SOBRE DOCUMENTAÇÃO, 2.; FID/CLA Reunião 9., Rio de Janeiro, 1969. Anais. Rio de Janeiro, IBBD, 1969. p. 195-206.
128. ZAHER, C. L. et alii. Automação da informação em Física no Brasil. In: SEMINÁRIO SOBRE INFORMÁTICA, Rio de Janeiro, 1968. Anais. Rio de Janeiro, IBBD, 1969. p. 39-52.
129. DIENER, Richard A.V. Informational dynamics of journal article titles. Journal of the American Society for Information Science, 35(4): 222-7, 1984.
130. BATES, Marcia J. Search technique. Annual Review of Information Science and Technology, 16: 139-69, 1981

As referências 129 e 130 não se encontram em ordem alfabética porque esse documentos foram localizados na fase final de redação dessa dissertação.

10. ANEXOS

10.1 ANEXO 1

Descrição das funções dos programas
utilizados

ANEXO I

NOMES E PRINCIPAIS FUNÇÕES DOS PROGRAMAS INTEGRANTES DO SISTEMA BIB/DIALOGO, UTILIZADOS NA PESQUISA.

<u>NOME</u>	<u>FUNÇÃO</u>
INDEXBD/011	Captura e validação de dados em modo interativo. Geração de um arquivo de dados em disco ou fita, em formato de trabalho para processamento posterior.
INDEXBD/012	Reformata o arquivo gerado pelo programa anterior, de modo a permitir atualizações e correções dos registros em modo interativo.
INDEXBD/013	Reformata o arquivo anterior, após correção, em formato de trabalho.
INDEXBD/023	Gera um índice KWIC a partir dos títulos dos registros em formato de trabalho (utiliza uma tabela de palavras vazias "STOPWORDLIST").
INDEXBD/025	Gera um índice de entidades (autores corporativos, filiação dos autores) a partir de registros em formato de trabalho.

INDEXBD/026

Gera uma listagem de referências bibliográficas completas (sem descritores) ordenadas, sequencialmente, a partir dos números dos registros bibliográficos.

INDEXBD/031

Gera um arquivo invertido de assuntos (descriptor versus número do registro), em disco. Mediante o utilitário "SYSTEM/DUMPALL" é obtido um índice de assuntos impresso.

INDEXBD/032

Gera um arquivo invertido de autores (autor versus número do registro), em disco. Mediante o utilitário "SYSTEM/DUMPALL" é obtido um índice de autores impresso.

INDEXBD/033

Reúne num arquivo só os dois arquivos anteriores, gerando um arquivo invertido, em disco, de assuntos mais autores versus número dos registros.

INDEXBD/034

Cria, a partir do arquivo anterior, os arquivos "Invertfile" e "Linkfile", integrantes da base de dados para interrogação em linha.

INDEXBD/035

Cria a partir do arquivo contendo registros em formato de trabalho, os arquivos "Masterfile" e "Accessfile", integrantes da base de dados para interrogação em linha.

INDEXBD/036

Permite a interrogação da base de dados integrada pelos arquivos "Invertfile", "Linkfile", "Accessfile" e "Masterfile", selecionando as referências pertinentes. Exibe na tela e / ou imprime os números ou as referências dos registros recuperados.

INDEXBD/041

Imprime uma relação alfabética dos descritores, no arquivo, indicando suas respectivas ocorrências. Imprime uma relação alfabética dos descritores que ocorrem mais de 4 vezes, indicando os descritores que com eles se associam nos documentos, com as respectivas frequências de associação.

INDEXBD/042

Estabelece uma relação alfabética de descritores, indicando suas respectivas ocorrências. Estabelece uma relação de descritores com suas respectivas frequên-

cias, ordenados segundo as frequências crescentes.

INDEXBD/043

Imprime uma tabela com os dados necessários para desenhar a curva de frequências acumuladas em função do número de ocorrências dos descritores segundo a lei de Bradford - Zipf.

INDEXBD/061

Cria e / ou atualiza a lista de palavras vazias (Wordfixed) e a lista de raízes vazias (Wordroots) utilizadas pelos programas seguintes.

INDEXBD/062

Cria um arquivo intermediário em disco, contendo as palavras do título e do resumo que passaram pelo filtro da comparação com a tabela de palavras vazias (Wordfixed), após eliminação das palavras, repetidas.

INDEXBD/063

Utiliza o arquivo intermediário anterior, considerando como descritores as palavras que não contem as raízes que figuram na tabela de raízes vazias (Wordroots), colocando-os nos subcam

pos reservados para os descritores, no registro em formato de trabalho.

INDEXBD/099

Permite criar e / ou atualizar a tabela "STOPWORDLIST" utilizada pelo programa INDEXBD/023.

INDEXBD/093

Edita uma listagem dos registros apresentado as informações completas (referência e resumo) em linhas de 60 caracteres, seguidas dos correspondentes descritores.

10.2 ANEXO 2

Descritores obtidos por indexação
manual e automática

DESCRITORES	INDEXCI		CINFORM		KWIC	
	<4	74	<4	74	<4	74
ABNT	1		-	-	-	-
Acadêmico	1		-	-	-	-
Acervo	-	-	1		-	-
Acervo bibliográfico	-	-	2		-	-
Acessibilidade da informação	1		1		1	
Acesso	2		1		-	-
Acesso a informação	1			5	1	
Acoplamento bibliográfico	-	-	1		-	-
Acumulação	1		-	-	-	-
Administração	1		-	-	-	-
Administração de projetos	1		1		-	-
Administrar	1		-	-	-	-
Agências bibliográficas	-	-	1		-	-
Agência social	-	-	1		-	-
Agente de informação	-	-	1		-	-
Agente Social	-	-	1		-	-
Agentes	1		-	-	-	-
Agrícola	3		-	-	5	-
Agricultura	1	-		5	-	-
Algebricamente	1		-	-	-	-
Algebricos	1		-	-	-	-
Ambiental	1		-	-	1	
Ambiente	2				-	-
America Latina	1		1		-	-
Americano	1		-	-	1	
Amostra	1		-	-	-	-
Amostrada	1		-	-	-	-
Análise		22	-	-	-	-
Análise bibliométrica	-	-	-	8	-	-
Análise da informação	-	-		13	-	-
Análise de assuntos	1		-	-	-	-
Análise de citações	-	-	2		1	
Análise do conteúdo	1		-	-	-	-
Análise do documento	-	-	2		-	-
Análise documentária	-	-	1		-	-
Análise econômica	1		-	-	-	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Análise estrutural	1		1		-	-
Análise linguística	-	-	1		-	-
Análise quantitativa	1		-	-	-	-
Análise semântica	-	-	1		-	-
Análise sintática	1		-	-	-	-
Análise sintática do termo	-	-	1		-	-
Ante - projeto	1		-	-	-	-
Aquisição	-	-	1		-	-
Aquisição planejada	1		1		-	-
ARIST	-	-	1		-	-
Armazenagem da informação	-	-	1		-	-
Armazenamento da informação	-	-	2		-	-
Arquivo (s)	3			4	-	-
Arquivo do ISI	-	-	1		-	-
Arquivo particular	-	-	1		-	-
Arquivo Pascal	-	-	1		-	-
Arquivo principal	1		1		-	-
Arquivo privado	1		-	-	1	
Artigos	1		-	-	-	-
Artigos de periódicos	1		-	-	1	
Asca	1		1		1	
Associação	1		-	-		
Associação de descritores	-	-	1		-	-
Assuntos	1		-	-	-	
Atuação profissional	-	-	1		-	
Atualização do depósito legal	-	-	1			
Automação		5		26	4	
Automação de bibliografia	-	-	1		1	
Automação de serviços	-	-	1		-	
Automatizado (s)		4	-	-	2	
Autonomia	1		-	-	-	
Autor	3		-	-	1	
Autores	-	7	-	-	2	
Autores brasileiros	-	-	1		1	
Autores científicos	-	-	3		-	
Autores coincidentes	-	-	1			
Autores coletivos	-	-	1			

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Autores individuais	-	-	1		-	
Autores ingleses	-	-	1		-	-
Autores latinos	-	-	1		-	-
Autores permanentes	-	-	2		-	-
Autores portugueses	-	-	1		-	-
Autores transientes	-	-	2		-	-
Autoria	2		-	-	1	
Autoria múltipla	-	-	1		-	-
Autoria única	-	-	1		-	-
Autoridade	1		-	-	1	
Avaliação		6	-	-	3	-
Avaliação de coleções	-	-	1		1	
Avaliação do sistema	-	-	2		1	
Avaliar	1		-	-		
Banco	1		-	-		
Banco de dados		4		9	3	
Barreiras	1		-	-	1	
Barreiras culturais	1		1		1	
Base(s) de dados	-	-		6	-	-
Base de dados multilingue	-	-	1		-	-
Bases de dados referenciais	1		-	-	-	-
Benefício do sistema	-	-	1		-	-
Bibliografia(s)		5		15		6
Bibliografia brasileira	2		1		2	
Bibliografia classificada	-	-	1		-	-
Bibliografia corrente	-	-	1		-	-
Bibliografia de Solla Price	-	-	1		-	-
Bibliografia do xisto	1		1		1	
Bibliografia estatística	1		1		1	
Bibliografia nacional	-	-	1		-	-
Bibliografias setoriais	-	-	1		-	-
Bibliográfica (s)	2		-	-	2	
Bibliometria	1			18	1	
Bibliométrica (o) (s)		8	-	-		7
Biblioteca(s)		18		34		10
Biblioteca (s) brasileira (s)	-	-	2		1	

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥4	< 4	≥4	< 4	≥4
Biblioteca do Congresso	1		1		-	-
Biblioteca(s) especializada(s)	-	-	3		-	-
Biblioteca nacional	3		-	5	3	
Biblioteca pública	-	-	2		-	-
Biblioteca Regional de Medicina	1		1		1	
Biblioteca(s) universitária(s)	2		3		1	
Bibliotecário (s)		5		11	1	
Bibliotecário como agente social	-	-	1		-	-
Bibliotecário de sistemas	-	-	1		-	-
Bibliotecários de referência	-	-	1		-	-
Biblioteconomia		9		22		8
Biográficos	1		-	-	-	-
Biologia	-	-	1		-	-
Biological Abstracts	-	-	1		-	-
Boletim de alerta corrente	1		-	-	-	-
Botânica	1		1		1	
Bradford	3		-	-	1	
Brasil		12		23	1	
Brasileir (a) (as) (o) (os)		14	-	-	-	-
Bulletin - Signaletique	1		-	-	-	-
Burocratização	-	-	1		-	-
Busca	1		-	-	-	-
Busca retrospectiva	-	-	1		-	-
Buscando	1		-	-	-	-
Cadastro de pesquisas	1		1		-	-
Cadeia	1		-	-	1	-
Campos de informação	-	-	1		-	-
Canais de informação	-	-	1		-	-
Canais informais	1		-	-	1	
CAPESQ	1		-	-	1	
Capitalismo	-	-	1		-	-
Características		7	-	-	-	-
Cartões	1		-	-	-	-
Caso	3		-	-	-	-
Catálogo	1			4	1	
Catálogo (s)	3		1		1	

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Catálogo Coletivo Nacional	-	-	1		-	-
Catálogo multindexado	1		1		1	
Catálogo sistemático	1		1		-	-
Catálogo(s) de biblioteca	-	-	1		-	-
Catálogos de editores	-	-	1		-	-
Categorias	1		-	-	-	-
Cooperação	1	-	-	-	-	-
Centros	1		-	-	-	-
Centros de análise	1		-	-	-	-
Centro(s) de documentação	2			16	-	-
Centros de informação	1		1		1	
Chagas	1		-	-	1	
Ciclo	1		-	-	-	-
Ciência(s)		10		12		10
Ciência-Brasil	1		-	-		
Ciência da Informação		13		24		10
Ciência e Tecnologia		6	-	-		5
Ciência para o progresso	1		-	-	-	-
Ciência pura	1		-	-	-	-
Ciências Biológicas	1		2		1	
Ciências Humanas	2		1		2	
Cienciometria	2		-	-	2	
Científica (o) (os)		12	-	-		20
Científica e técnica	1		-	-	2	
Científica e tecnológica		8	-	-		6
Cientista (s)	2		1		1	
Cientista da informação	-	-	2		-	-
CIPC	1		-	-	-	-
Circulação	3		-	-	1	-
Citação (oês)		4	-	-	3	
Citações-bibliográficas	1			5	-	-
Citados	2		-	-	1	-
Citation Indexes	1		1			
Classes	1		-	-	-	-
Classificação	2			9		6
Classificação automática	-		2		-	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Classificação bibliográfica	1		-	-	-	-
Classificação Decimal	-	-	1		-	-
Classificação Decimal Universal	3			5	3	
Classificação enumerativa	1		1		1	
Classificação(ões) facetada(s)	1		2		1	
CNRS	1		-	-	-	-
Cobertura	1		-	-	-	-
Coincidência	-	-	1		-	-
Coleção (oes) de referência	1		1		1	
Coleções	1		-	-	-	-
Colégio invisível	-	-		4	-	-
Coleta	1		-	-	-	-
Comparação	1		-	-	-	-
Computação	1	-	-	-	-	-
Computador	1			16	-	-
Comunicação (oões)		8	2			8
Comunicação científica	1			12	1	
Comunicação com o sistema	-	-	1		-	-
Comunicação da informação	-	-		8	-	-
Comunicação formal	1		3		1	
Comunicação informal	2			6	2	
Comunicação rural	1		1		1	
Comunicação técnica	1		1		1	
Comunidade científica	1		1			
Comunidade européia	-	-	1	-	-	-
Comutação	-	-	1		-	-
Conceito (s)		11	2		2	
Conceito de biblioteca	-	-	1		-	-
Conceitos heterogêneos	-	-	1		-	-
Condicionamentos	1		-	-	-	-
Conhecimento (s)		8	-	-	2	
Conhecimento científico	-	-	1		1	
Conhecimento tecnológico	-	-	2		1	
Consenso	1		-	-	-	-
Consistência	1		-	-	1	-
Consistência da Indexação	-	-	1		-	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Constantes	1		-	-	-	-
Consultado (tas)	2		-	-	-	-
Conteúdo	2		-	-	-	-
Conteúdo do(s) documento(s)	-	-	2		-	-
Controle	1		-	-	1	
Controle bibliográfico universal	-	-	2		-	-
Controle terminológico	1		1		1	
Cooperação		4	-	-	-	-
Cooperação entre bibliotecas	-	-	1		-	-
Cooperação internacional	-	-	1		-	-
Coordenação	1		-	-	1	
COPPE/UFRJ	1		-	-	1	
Corrente (s)	2		-	-	1	
Correspondência particular	1		-	-	-	-
Crescimento	1		-	-	2	
Crescimento da informação	-	-	1		-	-
Crescimento da literatura	2	-		4	1	
Criptografia	1		1		-	-
Crises	1		-	-		
Cultura(is)	3		1		2	
Curricular	1		-	-	-	-
Currículo	-	-	1		-	-
Curso (s)	2		-	-	1	
Curso de mestrado	2		2		-	-
Curso de pós-graduação	-	-	2		-	-
Custo (s)	2		-	-	-	-
Custo do sistema	-	-	2		-	-
Dados		7	-	-		6
Dados bibliográficos	2		-	-	-	-
Decimal de Dewey (res)	1		-	-	-	-
Decorrente	1		-	-		
Demanda	1		-	-	1	
Demanda de informação	2		1		-	-
Denver Research Institute	1		-	-	-	-
Departamentos	2		-	-	1	
Dependência tecnológica	1		1		1	
Dependente	1		-	-	1	

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Índice em cadeia	1		2		1	
Índice KWIC	-	-	2		-	-
Índice KWOC	-	-	1		-	-
Índice numérico	-	-	1		-	-
Índice permutado	-	-	1		-	-
Índice rotado	-	-	1		-	-
Índice UNIDEK	-	-	1		-	-
Índice(s) auxiliar(es)	1		1		-	-
Índices permutados	-	-	1		-	-
Individuais	2		-	-	-	-
Indústria da informação	1		1			
Indústria editorial	-	-	1		-	-
Infantil	1		-	-	1	
Informação (ões)		58		91		71
Informação agrícola	1		2		1	
Informação ambiental	1		1		1	
Informação bibliográfica	1		2		1	
Informação científica	3			41		7
Informação nuclear	-	-	2		-	-
Informação técnica			1		2	
Informação tecnológica	-	-	-	29		4
Informacionais	1		-	-	-	-
Informática	3			6	3	
Informática agrícola	1		1		1	
Infra-estrutura	2		-	-	-	-
Inglêses	1		-	-	-	-
Iniciativa privada	-	-	1		-	-
Iniciativa pública	-	-	1		-	-
Inovação tecnológica	1		-	-	1	
Instituição(ões) de pesquisa	1		1		1	
Institute for Scientific Information	1		-	-	-	-
Instituto (s)		4	-	-	5	
Instituto de pesquisa	-	-	1		-	-
Instituto Nacional do Livro	-	-	1		-	-
Intensão do conceito	-	-	1		-	-
Intercâmbio	1		-	-	-	-
Intercâmbio de informação (ões)	1			6	-	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Depósito legal	1		3		-	-
Derek de Solla Price	1		1		1	
Descreve (vendo)		4	-	-	-	-
Descritiva	1		-	-	-	-
Descritor (es)	-	-		7	-	-
Desempenho	1		-	-	-	-
Devido	1		-	-	-	-
Disciplina (s)	2	-	-	-	-	-
Dispersão	1	-	-	-	-	-
Dispersão da literatura	2		3		-	-
Disponibilidade de documentos	1		1		1	
Disponíveis (vel)	2	-	-	-	-	-
Disseminação	1		-	-	-	-
Disseminação da informação	-	-	-	12	-	-
Disseminação seletiva da inf.	1		-	-	1	
Dissertações	1		-	-	1	-
Dissertações de mestrado	-	-	1		1	
Dissonâncias	1		-	-	-	-
Distribuição	1		-	-	1	
Divulgação	1		-	-	-	-
Documentação		5		26	3	
Documentação agrícola	1		-	-	-	-
Documentação tecnológica	-	-	1		-	-
Documentalista	-	-	2		-	-
Documentária (s)	2		-	-	1	-
Documento (s)		8		25	3	
Documento original	-	-	2		-	-
Documentos científicos	-	-	1		-	-
Documentos de patente	-	-	1		-	-
Doença	2		-	-	2	
Doença de Chagas	1		1		1	
DSI	1		2		-	-
ECO-UFRJ/IBICT	1		-	-	-	-
Ecológica	1		-	-	-	-
Economia da informação	-	-	1		-	-
Econômico (a)	2		-	-	1	
ECT - erro	1		-	-	-	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	<4	≥4	<4	≥4	<4	≥4
Edições	1	-	-	-	-	-
Editôra	-	-	1	-	-	-
Editores	2	-	3	-	-	-
Editores de enciclopédias	-	-	1	-	-	-
Éditores estrangeiros	-	-	1	-	-	-
Editoriais	1	-	-	-	-	-
Educação	3	-	1	-	3	-
Educação profissional	-	-	2	-	-	-
Efeito de Mateus	-	-	2	-	-	-
Efeito Mathew	-	-	2	-	-	-
Efeitos	2	-	-	-	-	-
Efetividade	1	-	-	-	-	-
Efetividade da biblioteca	-	-	1	-	-	-
Efetividade do sistema	-	-	1	-	-	-
Eficamente	1	-	-	-	-	-
Eficiência	2	-	-	-	1	-
Eficiência do sistema	-	-	1	-	-	-
Eficientemente	1	-	-	-	-	-
Egressos	1	-	-	-	1	-
Elaboração	3	-	-	-	-	-
Elementos	1	-	-	-	-	-
ELVCDU 1,2,3, (n)	3	-	-	-	-	-
EMBRAPA	1	-	1	-	1	-
Emissor	1	-	1	-	-	-
Empresa de fundações	1	-	1	-	1	-
Empréstimo	1	-	1	-	-	-
Empréstimo entre bibliotecas	-	-	1	-	-	-
Enciclopédia (s)	1	-	1	-	1	-
Enciclopédias(s) brasileira. (s)	1	-	1	-	-	-
Energia	1	-	-	-	-	-
Energia nuclear	-	-	2	-	-	-
Engenharia	1	-	1	-	1	-
Engenharia Química	1	-	1	-	1	-
Engenheiros	1	-	-	-	-	-
Ensino	1	-	-	-	-	-
Entrada. (s)	2	-	-	-	1	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Entrada de informação	-	-	1		1	
Enunciados	1		-	-	-	-
Epidemiológico	1		-	-	1	
Epistemologia	1		-	-	1	
Epistemologia Social	1		1		1	
Epistemológicos	1		-	-	-	-
Equipamento (s)	2		-	-	-	-
Escritos	1		-	-	-	-
Espaço quase métrico	-	-	1		1	
Especialista da informação	-	-	1		-	-
Estados Unidos	1		-	-	-	-
Estágio	3		-	-	-	-
Estatísticas (os)	1		-	-	1	
Estrangeiros	1		-	-		
Estrutura (1)	1		-	-	1	
Estrutura organizacional	-	-	1		-	-
Estruturação	3		-	-	1	
Estudo comparativo	2		-	-	-	-
Estudo de caso	2		-	-	-	-
Estudo (s) de usuário (s)	1			4	2	
Estudo (s) recapitulativo (s)	1		1		-	-
Etapas	1		-	-	-	-
EURONET	1		1		1	
Europa (eria)	2		1		2	
Ex-alunos	1		-	-	-	-
Exame	1		-	-	-	-
Explosão bibliográfica	-	-		4	-	-
Exportar	1		-	-	-	-
Extensão do conceito	-	-	1		-	-
Extensionista agrícola	-	-	1		1	
Extensionistas	1		-	-	1	
Exterior	1		-	-	-	-
Externos	1		-	-	-	-
Extraída	1		-	-	-	-
Fase	1		-	-	-	-
Fenômeno (s)	3		-	-	-	-
Filosofia	1		-	-	1	

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Filosofia da classificação	-		1		1	
Filosofo	1		-	-	1	
Filtragem	1		-	-	-	-
Filtragem da literatura	-	-	1		-	-
Filtros de qualidade	1		-	-	1	
Física	1		1		1	
Fluxo(s)		4	-	-	1	-
Fluxo de(a) informação	1			11	1	-
Fluxo de entrada de informação	-	-	1		-	-
Folheto(s)	2		2		1	
Fonte(s)	2		-	-	2	-
Fontes de informação	1			5	1	
Formação	2		-	-	1	-
Formação profissional	1			7	1	
Formato	1		-	-	-	-
Formato CALLO	-	-	1		-	-
Fórmula de transição	1		-	-	-	-
Fórmula de transição de offman	-	-	1		-	-
Fornecedores de(a) informação	1		1			
Fornecimentos de livros	1		1		-	-
França	1		1		1	
Frases	1		-	-	-	-
Frente de pesquisa	3			9	2	
Frequência	2		-	-	-	-
Frequência à biblioteca	-	-	3		-	-
Frequência biblioteca	-	-	1		-	-
Frequência de descritores	-	-	1		-	-
Função da biblioteca	-	-	1		-	-
Futurologia (lista de raiz)	-	-	1		1	
Gatekeepers	1		3		1	
Geologia(ca)	1		1		1	
Georges Anderla	1		-	-	1	
Geração (do)		4			1	
Gerente de informação	2		-	-	-	-
Gestão	-	-	1		1	

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Getúlio Vargas	1		1		1	
Goffman	3		-	-	-	-
Governamentais	1		-	-	-	-
Governo	1		1	-	-	-
Grupo de elite	-	-		5	-	-
Hábitos	2		-	-	-	-
Hardware	-	-	2		-	-
Histórico		4	-	-	1	-
Histórico Biblioteca Nacional	-	-	1		-	-
Histórico econômico sociais	1		-	-	-	-
Humanas	1		-	-	2	
Humanidade	1		-	-	-	-
IBBD	1		-	-	1	
IBICT		4	2		3	
Idioma	1		-	-	-	-
Impulso	1		-	-	-	-
Imunologia	1		1		1	
Incidência de citações	-	-	2		-	-
Independência	1		-	-	-	-
Indexação		8		21		4
Indexação automática	1			6	1	
Indexação automática simulada	-	-	1		-	-
Indexação coordenada	-	-	1		-	-
Indexação ponderada	-	-	1		-	-
Indexação pré-coordenada	-	-	1		-	-
Indexador	-	-	1		-	-
Indexados	1		-	-	-	-
Indicativo autoridade	-	-	1		-	-
Indicativo estado	-	-	1		-	-
Indicativo língua	-	-	1		-	-
Indicativo país	-	-	1		-	-
Indicativo território	-	-	1		-	-
Índice (s)		4		10	3	
Índice(s) alfabético(s)	1		2		1	
Índice de citação (oês)	-	-		6	-	-
Índice de Price	-	-	1		-	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Interconsistência	1		1		1	
Interdisciplinaridade	1		1		-	
Intraconsistência	-	-	1		-	-
ISI	1		1		1	
Internacional (is)		6	-	-	-	-
Interrelaçi (onada) (onamento) (ões) 3			-	-	-	-
Intersecção	1		-	-	-	-
ISA	1		-	-	-	-
ISI	1		1		1	
ISI -III	1		-	-	-	-
ISO/R-639	1		-	-	-	-
Jornais	1		1		1	
Jornal científico	1		-	-	-	-
Jornal do Brasil	1		-	-	-	-
KWIC	1		-	-	1	
Lei (s)		4	-	-	-	-
Lei da dispersão da literatura	-	-	1		-	-
Lei de Bradford	2			8	1	
Lei de Lotka	1			4	-	-
Lei de ZIPF	2			7	-	-
Lei do crescimento exponencial	-	-	1		-	-
Lei do elitismo	-	-	1		-	-
Leitores	1		-	-	-	-
Leitura	1		-	-	-	-
Leitura do documento	-	-	2		-	-
Levantamento bibliográfico	1		1		-	-
Library - Literature	1		-	-	-	-
Língua	1		-	-	1	
Língua Portuguesa	3		1		2	
Linguagem	1		-	-	-	-
Linguagem artificial	-	-	1		-	-
Linguagem de busca	1		1		1	
Linguagem (ns) de indexação	1		2		1	
Linguagem (ns) de recuperação	1		2		-	-
Linguagem documentária	1		1		1	
Linguagem jornalística	1		1		1	

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Linguagem (n) natural (is)		5		5	-	-
Linguística (s) (cos)		5		9	3	
Linguística computacional	-	-	1		-	-
LISA	1		1			
Literatura	1		-	-		8
Literatura científica	1		1		1	
Literatura de Química	-	-	1		1	
Literatura especializada	-	-	1		-	-
Literatura geológica	1		1		1	
Literatura Siderúrgica	-	-	1		-	-
Livro (s)		5		6	3	
Localização	1		-	-	-	-
Lotka	1		-	-	-	-
Luis Floren	1		1		1	
Macrossistema	1		-	-	-	-
Macrothesaurus	1		-	-	1	
Macrothesaurus da OCDE	-	-	1		1	
Matéria	2		-	-	-	-
Material	2		-	-	-	-
Medicina			1		1	
Médias	1		-	-	-	-
Medidas	2		-	-	-	-
Medidas de informação	-	-	1		-	-
Medlars-on-line	1		-	-	-	-
MEDLINE	1		-	-	-	-
Meio ambiente	-	-	1		-	-
Meios de informação	1		-	-	-	-
Mercado de informação	-	-	1		-	-
Mestrado	1		-	-	1	
Meta	1		-	-	-	-
Método (s)		8	-	-	-	-
Metodologia (s)		5	-	-	2	
Método (s) estatístico (s)	1		3		-	-
Métodos linguísticos	-	-	1		-	-
Microbiologia	1		1		1	
Microcomputadores	-	-	1		-	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
MIP	1		-	-	1	
Mitos	1		-	-	-	
Modelo determinístico	-	-	1		-	-
Monolíngues	1		-	-	-	-
Multilíngues	1		-	-	-	-
Multilínguismo	1		-	-	1	
Múltipla	1		-	-	-	-
Nacionais (1)		9	-	-	-	8
Não-usuário(s)	3		2		1	
NATIS	-	-	1		-	-
NB - 69	1		-	-	1	
Necessidade(s) de informação	-	-		14	-	-
Nível de satisfação de usuários	-	-	1		-	-
Nomenclatura	1		-	-	-	-
Nordeste	1		1		1	
Norma	1		1		1	
Norma técnica	-	-	1		-	-
Normalização	2		2		1	
Normalização de documentos	-	-	1		-	-
Normalização de originais	-	-	1		1	
Normalização internacional	-	-	1		-	-
Normalização no Brasil	-	-	1		-	-
Núcleo	1		-	-	-	-
Numeração-progressiva	1		-	-	1	
Número	1		-	-	-	-
Obra (s)	3		-	-	-	-
OCDE	1		-	-	1	
ONLINE	1		-	-	1	
Operacionalização	1		-	-	-	-
Operacionalização do sistema	1		-	-	-	-
Ordenadas	1		-	-	-	-
Organização		6	-	-	3	
Organização de conceitos	-	-	1		1	
Organização do sistema	-	-	1		-	-
Organização social	-	-	1		-	-
Organização-Panamericana de Saúde	1		-	-	1	

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥4	< 4	≥4	<4	≥4
Orientação (da)	3		-	-	1	
Origem geográfica	1		-	-	-	-
Originais	1		-	-	1	-
Padrão (oes)	3		-	-	-	-
País (es)		4	-	-		4
Países desenvolvidos	-	-	1	-	-	
Países em desenvolvimento	3			7	3	
Países industrializados	1		-	-	-	-
Palavra(s) chave	1		3		-	-
Papéis (l)		7	-	-	1	
Papel social da biblioteca	-	-	1		-	-
Parasitologia	1		1		1	
PASCAL	1		-	-	1	
Patente (s)	2		3		2	
Periódico (s)		4		9	2	
Periódico(s) científico(3)	1			14	1	
Periódico diário	-	-	1		-	-
Permanencia	1		1		-	-
Pertinência	-	-	1		-	-
Pesquisa(s)		8	2			7
Pesquisa agrícola	1		1		-	-
Pesquisa bibliográfica	1		1		-	-
Pesquisa bibliométrica	-	-	1		-	-
Pesquisa científica	1		1		1	
Pesquisa de leitura	-	-	1		-	-
Pesquisa em semântica	-	-	1		-	-
Pesquisa industrial	1		1		1	
Pesquisador(es)		4	-	-	2	
Pesquisas em andamento	-	-	2		-	-
Pesquisas online	-	-	1		1	
Pessoal	1		-	-	-	-
Planejamento	2		-	-	1	
PNB - 27	1		-	-	1	
PNB - 286	1		1		1	
Política(s)		4	-	-		6
Política(s) científica(s)	2			6	1	

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Política científica e tecnologia	1	-	-	-	-	-
Política de administração	1	-	-	-	-	-
Política de aquisição	1	-	1	-	-	-
Política de informação	-	4	-	19	2	-
Política editorial	1	-	2	-	-	-
Política nacional de informação	3	-	-	5	1	-
Político	2	-	-	-	-	-
População	2	-	-	-	-	-
Portugueses	1	-	-	-	-	-
Pragmática	-	-	1	-	-	-
Precisão	-	-	1	-	1	-
Preservação	1	-	-	-	-	-
Privacidade da informação	-	-	1	-	-	-
Privada	1	-	-	-	-	-
Processamento de dados	-	-	1	-	-	-
Processamento distribuído	-	-	1	-	-	-
Processo de comunicação	-	-	1	-	-	-
Produção	-	9	-	-	-	-
Produção científica	-	6	-	9	-	-
Produção do sistema	-	-	1	-	-	-
Produção tecnológica	-	-	1	-	-	-
Produtividade absoluta	1	-	1	-	-	-
Produtividade de autor	-	-	2	-	-	-
Produtividade de periódicos	-	-	1	-	-	-
Produtividade do sistema	-	-	1	-	-	-
Produtividade relativa	1	-	1	-	-	-
Produto de informação	-	-	1	-	-	-
Produtos	3	-	-	-	-	-
Professores	1	-	-	-	1	-
Profissão	1	-	-	-	-	-
Profissionais(1)	-	4	-	-	2	-
Profissional da informação	1	-	-	-	-	-
Programa(s)	-	5	-	-	-	4
Programa educacional	1	-	-	-	1	-
Programas internacionais	1	-	-	-	1	-
Projeto(s)	-	4	-	-	1	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Projeto CADESQ	-	-	1	-	-	-
Projeto INPUT	-	-	1	-	-	-
Projeto MARC	-	-	1	-	-	-
Pública	2	-	-	-	2	-
Publicação (oes)	-	6	-	-	1	-
PUC/RJ	2	-	-	-	1	-
Quantidade	1	-	-	-	2	-
Qualificação	1	-	-	-	-	-
Quantificar	1	-	-	-	-	-
Química	2	-	3	-	3	-
Racionalização	1	-	-	-	-	-
RANGANATHAN	2	-	1	-	1	-
RECALL	1	-	-	-	-	-
Receptor	2	-	1	-	-	-
Recomenda-se(ções)	3	-	-	-	-	-
Recuperação	-	8	-	-	-	-
Recuperação automática	1	-	-	-	-	-
Recuperação de ^(a) (as) inf.(ões)	-	10	-	29	-	4
Recuperadas	1	-	-	-	-	-
Recursos	3	-	-	-	-	-
Recursos audiovisuais	1	-	-	-	-	-
Recursos de informação	-	-	3	-	-	-
Recursos humanos	1	-	-	-	-	-
Rede(s)	-	5	-	-	-	4
Rede de bibliotecas	-	-	1	-	-	-
Rede de comunicação	-	-	1	-	-	-
rede(s) de informação	1	-	-	20	1	-
Rede europeia de informação	-	-	2	-	1	-
Rede nacional	1	-	-	-	-	-
Rede nacional de informação	1	-	-	-	1	-
Redes de bases de dados	-	-	1	-	-	-
Referência(s) bibliográfica(s)	1	-	1	-	-	-
referências	1	-	-	-	2	-
Refleite	1	-	-	-	-	-
Reforma Biblioteca Nacional	-	-	1	-	1	-
Regional	2	-	-	-	3	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Registro	1	-	-	-	-	-
Registro bibliográfico	-	-	1	-	-	-
Registro científico	1	-	-	-	-	-
Registro de informação	-	-	2	-	-	-
Registros gráficos	-	-	1	-	-	-
Regras	1	-	-	-	1	-
Regras de catalogação	-	-	1	-	1	-
Relacionadas(do)	2	-	-	-	-	-
Relações bibliométricas	1	-	-	-	-	-
Relevância	-	6	3	-	2	-
Representação	1	-	-	-	1	-
Reprografia	-	-	1	-	-	-
Resumo	1	-	1	-	-	-
Revisão	3	-	-	-	-	-
Revisão(ões) de Literatura	3	-	2	-	1	-
Revistas	1	-	-	-	-	-
Revocação	1	-	1	-	1	-
Rotadas	1	-	-	-	-	-
Ruídos	1	-	-	-	-	-
Ruídos da comunicação	-	-	1	-	-	-
Satisfação do usuário	-	-	1	-	-	-
Science Citation Sndex	-	-	1	-	-	-
SDI/CIN/CNEN	1	-	-	-	1	-
Seções do documento	-	-	1	-	1	-
Segurança da informação	-	-	1	-	1	-
Seleção	3	-	2	-	-	-
Seleção de usuários	-	-	1	-	-	-
Semântica(s) (os)	-	5	-	8	3	-
Semântica Geral	1	-	1	-	1	-
Semânticamente	1	-	-	-	-	-
Semanticistas	1	-	-	-	-	-
Serviço de alerta	-	-	1	-	-	-
Serviço de alerta corrente	-	-	1	-	-	-
Serviço(s) de informação	-	-	-	17	2	-
Serviço(s) de referência	-	-	1	-	1	-
Serviço de SDI	-	-	1	-	2	-
Serviços "online"	-	-	1	-	-	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
serviços de biblioteca	-	-	3	-	1	-
serviços de utilidade pública	-	-	1	-	-	-
serviço (s)	-	-	-	-	-	10
siderurgia	1	-	-	-	1	-
siderurgia brasileira	1	-	1	-	1	-
SIEF	1	-	-	-	-	-
significação	1	-	-	-	-	-
significado	-	-	2	-	-	-
similaridades	1	-	-	-	-	-
sintaxe	1	-	4	-	-	-
Sistema (s)	-	23	-	-	-	-
Sistema de documentação	-	-	1	-	-	-
Sistema de indexação SPIRIT	-	-	1	-	-	-
Sistema(s) de informação (ões)	-	11	-	54	-	-
Sistema(s) de informação agrícola	2	-	-	-	-	-
Sistema Nacional de inf.	2	-	-	-	-	-
Sistema MEDLINE	-	-	1	-	-	-
Sistema PASCAL	-	-	1	-	-	-
Sistema UNIDEK	-	-	1	-	-	-
Sistemas online	-	-	1	-	-	-
SNICT	1	-	-	-	-	-
Social	-	4	-	-	3	-
Sociedade	-	5	-	-	2	-
Sociedade informatizada	-	-	1	-	-	-
Software	-	-	1	-	-	-
Software de aplicação	-	-	1	-	-	-
Source - Indexes	1	-	1	-	-	-
SPIRIT	1	-	-	-	-	-
Subcomissão	1	-	-	-	-	-
Subdesenvolvimento	1	-	-	-	1	-
Subjetivos	1	-	-	-	-	-
Tabelas	1	-	-	-	1	-
Técnica(s) (o) (os)	-	11	-	-	1	-
Técnico-econômico	1	-	-	-	-	-
Tecnologia (s)	-	6	-	20	-	8
Tecnologia da informação	2	-	3	-	-	2

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Tecnológica (o) (os)		11	1			8
Tema da pesquisa	1		1		1	
Teoria (s)		7	-	-		4
Teoria da comunicação	-	-	2		-	-
Teoria(s) da informação	1		3		1	
Teoria de Goffman	-	-		4	-	-
Teoria de Ranganathan	-	-	1		-	-
Teoria de Shannon	-	-	1		-	-
Teoria de Solla Price	-	-		5	-	-
Teoria do processo epidêmico	-	-	1		-	-
Teoria semântica da inf.	-	-	1		1	
Teoria unificada da inf.	-	-	1		-	-
Terminologia	1		3		-	-
Terminologia de documentação	1		-	-	-	-
Terminologia teórica e aplicada	-	-	1		-	-
Termos	3		-	-	-	-
Termos de classificação	-	-	1		-	-
Termos significativos	1		-	-	-	-
Territórios	1		-	-	1	-
Tesouro	-	-	-	6	-	-
Tesauros especializados	-	-	1		-	-
Testes	1		-	-	-	-
Texto	1		-	-	-	-
Tipo	1		-	-	-	-
Tipologia de erros	-	-	1		-	-
Tipologias	1		-	-	1	
Título (s)	3		-	-	1	
Tradução	1		-	-	-	-
Tradutologia	-	-	1		-	-
Transciência	1		-	-	1	
Transfêrencia	3		-	-		4
Transferência da informação		6		13	3	
Transferência de tecnologia	-	-		6	-	-
Transiência	-	-	2		-	-
Transientes	1		-	-	-	-
Transmissão	1		-	-	-	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	< 4	≥ 4	< 4	≥ 4	< 4	≥ 4
Tratamento	2		-	-	-	-
Tratamento bibliográfico	1		-	-	-	-
Tratamento da informação	-	-	2		-	-
Treinamento bibliotecário	-	-	1		-	-
Treinamento do bibliotecário	1		1		-	-
Treinamento profissional	-	-	1	-	-	-
Troca de informações	1		-	-	1	
Truncamento arbitrário	1		-	-	-	-
Truncamento de termos	-	-	1		-	-
UFMG	1		-	-	1	
UNESCO	1		1		1	
Unidade(s)	2		-	-	-	-
Unidade de informação	-	-	1		-	-
UNIDEK	1		1		1	
Unificação	1		-	-	-	-
Universidade	1		-	-	1	
Universidade Federal de Pernambuco	1		-	-	-	-
Universitaria	1		-	-	-	-
Universos	1		-	-	-	-
Uso da biblioteca	-	-	1	-	-	-
Usuário(s)		10		34		7
Usuário (em) potencial	-	-	3		-	-
Utilidade da informação	-	-	2		-	-
Utilidade pública	-	-	1		1	
Utilidade pública da inf.	-	-	1	-	-	-
Utilização da coleção	-	-	1	-	-	-
Utilização da informação	-	-		12	-	-
Valor da informação	-	-		4	-	-
Veículo	1		-	-	1	
Velocidade	1		-	-	-	-
Vida média	-	-	1		-	-
Vocabulário(s)	1		3		-	-
Vocabulário especializado	-	-	2		-	-
Vocabulário monolíngue	-	-	1		-	-
Vocabulário multilíngue	-	-	1		-	-
Volume	1		-	-	-	-

DESCRITORES	INDEXCI		CINFORM		KWIC	
	≤ 4	JP 4	≤ 4	7/ 4	≤ 4	2/4
Xisto	1		1		1	
Xisttos-ole igamos	1		-	-	-	-
Zipf	1		-	-	-	-
1909	1		-	-	1	
1930 " 1939	1		-	-	-	-
1948	1		-	-	-	-
1959	1		-	-	-	-
1960	1		-	-	-	-
1963 - 1979	1		-	-	1	-
1970	1		-	-	1	-
1971	2		-	-	-	-
1971 " 1972	1		-	-	1	-
1972	1		-	-	-	-
1973	2		-	-	1	
1978	1		-	-	-	-
1966	1		-	-	1	
1965	1		-	-	1	

10.3 ANEXO 3

Raízes e palavras truncadas
dos descritores

RAIZ

DESINÊNCIAS

Termos compostos
Pré - coordenados

	I	C	K	I	C	K	I	C	K
ABNT	1	-	-	-	-	-	-	-	-
ACADEM	1	-	-	1	-	-	-	-	-
ACERV	-	3	-	-	1	-	-	2	-
ACCESS	4	7	2	2	1	2	1	1	-
ACOPLAMENT	-	1	-	-	-	-	-	1	-
ACUMUL	1	-	-	1	-	-	-	5	-
ADMINISTR	3	1	1	2	-	1	1	1	-
AGEN	1	-	4	1	-	4	-	1	-
AGRIC	4	-	5	4	-	5	-	-	-
ALGEBR	2	-	1	2	-	1	-	-	-
AMBIENT	3	-	1	3	-	1	-	-	-
AMERIC	2	1	1	1	-	1	1	-	-
AMOSTR	2	-	-	2	-	-	-	-	-

RAIZ	DESINÊNCIAS			Termos compostos Pré-coordenados		
I C K	I C K	I C K	I C K	I C K	I C K	I C K
ANALIS	28 30 1 e(s)	22 - -	- - -	- - -	8 - -	- - -
				bibliométrica(s)	13 - -	- - -
				da informação	1 - -	- - -
				de assuntos	- - -	- - -
				de citações	3 1 -	- - -
				de conteúdo	1 - -	- - -
				do(s) documento(s)	- 2 -	- - -
				documentária	- 1 -	- - -
				econômica	1 - -	- - -
				estrutural	1 1 -	- - -
				linguística	- 1 -	- - -
				quantitativa	- 1 -	- - -
				semântica	- 1 -	- - -
				sintática	1 1 -	- - -
AQUIS	1 2 - ão	- 1 -	- 1 -	Planificada	1 1 -	- - -
ARIST	- 1 -	- 1 -	- 1 -		- - -	- - -
ARMAZEN	- 3 -	- - -	- - -	da(s) informação(ões)	- 3 -	- - -
ARQUIV	5 9 1 o(s)	2 5 1	1 do ISI		- 1 -	- - -
				Particular	- 1 -	- - -
				PASCAL	- 1 -	- - -
				Principal	1 1 -	- - -
				Privado	2 - -	- - -

RAIZ	DESINÊNCIAS			Termos compostos Pré - coordenados		
	I	C	K	I	C	K
ARTIG	2	-	1 (s)	1	-	1 de periódico(s)
ASCA	1	1	1	1	1	
ASSOCIAC	1	1	- Ao(ões)	1	-	de descritores
ASSUNT	1	-	- o(s)	1	-	
ATUAÇ	-	1	- ão (ões)	-	-	profissional
ATUALIZ	-	1	- ação, ada(s), ado(s), ar	-	-	do depósito legal
AUTOM	-	-	- ação	-	26	6 de bibliografia
			atização	9	-	de serviços
			atizado(s)	-	-	
			atizada(s)	-	-	
AUTONOM	1	-	- a(s), ia, o(s)	1	-	
AUTOR	-	-	-	3	-	1 brasileiro(s)
	12	12	5 es	-	-	científico(s)
						coletivos
						individual(eis)
				9	-	3 ingles(es)
						latin(os)
						portugues(es)
AUTOR			ia	-	-	múltipla
						única

RAIZ

DESINÊNCIAS

Termos compostos
Pre-coordenados

	RAIZ			DESINÊNCIAS			Termos compostos Pre-coordenados		
	I	C	K	I	C	K	I	C	K
AUTORIDAD	-	1	1	-	1	1	-	-	-
AVALIA	7	3	5	7	-	3	-	1	1
								2	1
BANC	5	9	3	1	-	-	4	9	3
BARREIR	2	1	2	-	-	-	-	-	-
BAS	1	7	-	-	-	-	1	6	-
								1	-
BENEFICI	-	1	-	-	-	-	-	1	-
BIBLIOGRAF	-	-	-	7	15	8	2	1	2
	11	23	12	-	-	-	-	1	-
								1	-
								1	-
								1	-
							1	1	1
							1	1	1
							-	1	-
							-	1	-
							-	1	-

ica(s)

ico(s)

RAIZ

DESINÊNCIAS

Termos compostos
Pré- Coordenados

	I	C	K	I	C	K	I	C	K
BIBLIOMETR	9	-	7	ia, ica(s), ico(s)	9	-	7	-	-
BIBLIOTEC	25	51	16	a(s)	19	35	10		
				brasileiras				-	2
				do Congresso				1	1
				especializada(s)				-	1
				nacional(is)				3	5
				publica(s)				-	1
				Regional de Medicina				1	1
				universitárias				2	3
BIBLIOTECAR				ia(s)				-	-
	5	14	-	io(s)	5	13	-	-	1
BIBLIOTECON				omia				-	-
				omica(s)				-	-
	9	22	8	omico(s)	9	22	8	-	-
BIOGRAF	1	-	-	ia(s), ica(s)	1	-	-	-	-
				ico(s)				-	-
BIOLOG	-	1	-	ia, ica(s), ico(s)	-	-	-	-	1
BOLET	1	-	-	a(s), o(s)	-	-	-	1	-
BOTANIC	1	1	1	a(s), o(s)	1	1	1	-	-

RAIZ	DESINÊNCIAS			Termos compostos Pré Coordenados	I	C	K
	I	C	K				
BRADFORD	3	-	1		-	-	-
BRASIL	12	23	1		-	-	-
BRASILEIR	14	-	-	a(s), o(s)	-	-	-
BULLETIN SIGNALETIQUE	1	-	-		-	-	-
BUROCRA	-	1	-	cia, tica(s), tico(s), tização	-	-	-
BUSC	2	1	-	a(s)	2	-	-
CADASTR	1	1	-	ada(s), ado(s), o(s)	-	-	-
CADEI	1	1	-	a(s)	1	1	-
CAMP	-	1	-	o(s)	-	-	-
CAN	1	1	1	al, ais	-	-	-
CAPEQ	1	-	1		1	-	1
CARACTERISTIC	7	-	-	a(s), o(s)	7	-	-
CART	1	-	-	ao, ces	1	-	-
CAS	3	-	-	o(s)	3	-	-

RAIZ DESINÊNCIAS

Termos compostos
Pré - Coordenados

RAIZ	I	C	K	I	C	K	I	C	K
CATALOG	5	10	2	3	5	1			
							Coletivos Nacional		
							de biblioteca(s)		
							de editor(es)		
							multindexado		
							sistemático		
COBERTURA	1	-	-	1	-	-			
COINCID	-	1	-	-	1	-			
COLEC	2	1	1	1	-	-	de referência		
COLÉGIO INVISIVEL	-	4	-	-	4	-			
COLETA	1	-	-	1	-	-			
COMPARA	1	-	-	-	-	-			
							ção(ões), tiva(s) -		
							tivo(s)		
COMPUT	2	16	-	2	16	-			
							ação, dor(es)		
COMUNICA	14	34	14	8	2	8	científica		
							com o sistema		
							da informação		
							formal(ais)		
							informal (ais)		
							rural		
							técnica(s)		

RAIZ

Termos compostos
Pré - Coordenados

DESINÊNCIAS

I C K I C K I C K

COMUNID	3	2	-	ade(s)	1	-	-	científica	1	1	-
COMUTAÇÃO	-	1	-		-	1	-	européia	1	1	-
CONCEIT	11	4	2	o(s), nal(eis)	11	4	2		-	-	-
CONDICIONA	1	-	-	da(s), do(s), mento(s)	1	-	-		-	-	-
CONHECIMENT	8	3	4	o(s)	8	-	2	científico(s) tecnológico(s)	-	1	1
CONSENS	2	-	-	o(s), val	2	-	-		-	-	-
CONSISTEN	1	1	1	cia, te(s)	1	-	1	da indexação	-	1	-
CONSTANT	1	-	-	e(s)	-	-	-		-	-	-
CONSULT	4	-	-	a(s), da(s), do(s) or(es)	4	-	-		-	-	-
CONTEUD	-	2	-	o(s)	-	-	-	do(s) documento(s)	-	2	-
CONTROL	2	3	1	ada(s), ado(s), e(s)	1	-	-	bibliográfico universal terminológico	-	2	-
COOPER	5	3	1	ação, ante(s), ar tiva(s), tivo(s)	5	1	1	entre bibliotecas internacional	-	1	-
									-	1	-

RAIZ
DESINÊNCIAS
Termos compostos
Pré-Coordenados

RAIZ	I	C	K	DESINÊNCIAS	I	C	K	Termos compostos Pré-Coordenados	I	C	K
COORDEN	1	-	1	ação, ada(s), ado(s)	1	-	1		-	-	-
COPPE/UFRG	1	-	-		1	-	-		-	-	-
CORRENT	2	-	1	e(s)	2	-	1		-	-	-
CORRESPOND	1	-	-	ência, ente	-	-	-	particular	1	-	-
CRESC	3	5	2	imento	1	-	1	da informação da literatura	-	1	-
CRIPTOGR	2	1	-	afia,eme(s)	2	1	-		-	-	-
CRIS	1	-	-	e(s)	1	-	-		-	-	-
CULTUR	3	1	2	e, al(ais)	3	1	2		-	-	-
CURRIC	1	1	-	ular(es), ulo(s)	1	1	-		-	-	-
CURS	4	4	1	o(s)	2	-	1	de mestrado de pós-graduação	2	-	-
CUST	2	-	2	o(s)	2	-	-	do(s) sistema(s)	-	4	-
DAD	9	-	6	o(s)	7	-	6	bibliográficos	2	-	-

RAIZ

DESINÊNCIAS

Termos compostos
Pré-Coordenados

I C K I C K I C K

DECIMAL DE DEWEY	1	-	-	1	-	-	-	-	-	-	-
DEMAND	3	1	1	a(s)	3	-	-	de informação	-	1	1
DENVER RESEARCH INSTITUTE	1	-	-		1	-	-		-	-	-
DEPARTAMENT	2	-	1	al(a)is),o(s)	2	-	1		-	-	-
DEPENDEN	1	1	1	cia,te(s)	-	-	-	tecnológica	1	1	1
DEPOSIT	1	3	-	o(s)	-	-	-	legal	1	3	-
DEREK DE SOLLA PRICE											
DESCR	1	1	1		1	1	1		-	-	-
	5	7	-	ever,ição(ções)	5	7	-		-	-	-
				itor(es),							
				itiva(s)itivo(s)							
DESEMPENHO	1	-	-		1	-	-		-	-	-
DISCIPLIN	2	-	-	a(s)	2	-	-		-	-	-
DISPERS	3	3	-	ão	1	-	-	da literatura	2	3	-
DISPONI	3	1	1	bilidade,	2	-	-	dos documentos	1	1	1
				vel(veis)							

RAIZ	DESINÊNCIAS			Termos compostos Pré - Coordenados		
I C K	I C K	I C K	I C K	I C K	I C K	I C K
DESSEMIN	2 12 1	eda(s),ado(s), ar ação	2 - 1	1 da informação seletiva da informação	- 12 -	- - -
DISSERTAÇ	1 1 1	ão(oês)	1 - 1	1 de mestrado	- 1 -	- - -
DISTRIBUI	1 - -	ção, da(s), do(s)	1 - 1		- - -	- - -
DIVULG	1 - -	ção,ada(s) ado(s), r	1 - -		- - -	- - -
DOCUMENT	16 59 7	ção,alista(s) aria(s),ario(s) o(s)	15 54 7	7 agrícola tecnológica científico(s) de patente original(ais)	1 - 1 - 1 - 1 - 2 -	- - - - -
DOENÇ	3 1 2	e(s)	2 - 1	1 de chagas	1 1 1	1 1 1
OSI	1 2 -		1 2 -		- - -	- - -
ECO-UFRJ/IBICT	1 - -		1 - -		- - -	- - -
ECOLOG	1 - -	ia,ica(s), ico(s)	1 - -		- - -	- - -

	I	C	K	I	C	K	I	C	K
ECONOM	2	1	1	1	2	-	1	1	-
				ia,ice(s), ico(s)			da informação		
EDI	4	7	-	ção(ções), tor(a,as)(es) torial(a(is)	4	5	1	1	-
							de enciclopédias estrangeiros		
EDUC	3	3	3	ação	3	1	3	2	-
							profissional		
EFEITO	-	4	-		-	-	-	2	-
							de mateus mathew		
EFETIV	1	2	-	idade(s),a(s) o(s)	1	-	-	1	-
							da bibliotece do sistema		
EFIC	4	2	-	acia,az,azment, iencia(s) iente_ mente	4	1	-	1	-
							do sistema		
EGRESS	1	-	1	o(s)	1	1	1	-	-
ELVCDU	3	-	-	1,2, 3	3	-	-	-	-
EMBRAPA	1	1	1		1	1	1	-	-
EMISS	1	-	1	ão(ões),or(es)	1	-	1	-	-
EMPRESA	1	1	1		-	-	-	1	1
							de fundações		

RAIZ
Termos compostos
Pré - Coordenados

DESINÊNCIAS

	I	C	K	I	C	K	I	C	K		
EMPREST	1	2	-	imo(s)	-	-	-	entre bibliotecas	-	1	-
ENCICLOPED	1	1	1	ie(s)	-	-	-	brasileira(s)	-	-	-
ENERG	1	2	-	ia	1	-	-	nuclear	-	2	-
ENGENH	1	1	1	aria, eico(s)	-	-	-	química	1	1	1
ENSINO	1	-	-		1	-	-		-	-	-
ENTRAD	2	1	2	a(s)	1	1	-	da informação	1	-	1
ENUNCIA	1	-	-	do(s)	1	-	-		-	-	-
EPISTEMOLOG	4	1	3	ia, ica(s), ico(s)	3	-	2	societ	1	1	1
EQUIP	2	-	-	umento(s)	2	-	-		-	-	-
ESCRIT	1	-	-	o(s)	1	-	-		-	-	-
ESPAÇO	-	1	1		-	-	-	quase-métrico	-	1	1
ESPECIAL	-	1	1	istefs)	-	-	-	da informação	-	1	1
ESTADOS UNIDOS	1	-	-		-	-	-		-	-	-
ESTAGI	4	-	1	aria(s), ario(s), o(s)	4	-	1		-	-	-

RAIZ	DESINÊNCIAS			Termos compostos Pré - Coordenados				
I	C	K	I	C	K	I	C	K
ESTATIST	1	-	1	ica(s),ico(s)	1	-	1	-
ESTRANG	1	-	1	eira(s), eiro(s)	1	-	1	-
ESTRUTR	3	1	2	a(s),ção, ada(s),ado(s), al(ais)	3	1	2	-
ESTUD	6	5	2	ante(s),o(s)	1	-	-	comparativo(s) de caso de usuário(s) recapitulativo(s)
ETAP	1	-	-	a(s)	1	-	-	-
EURONET	1	1	1		1	1	1	-
EUROP	2	1	2	a, eia(s) eio(s)	2	1	2	-
ALUN	1	-	-	a(s),o(s)	1	-	-	-
EXAM	1	-	-	e	1	-	-	-
EXPLOS	-	4	-	ões)	-	-	-	bibliográfica

	I	C	K		I	C	K		I	C	K
EXPORT	1	-	-	ação(ões)	1	-	-		-	-	-
EXTENS	1	2	2	ão,ionista(s)	1	-	1	do conceito agrícola	-	1	-
EXTER	2	-	-	ior(es),na(s) no(s)	-	-	-		-	-	-
FILOSOF	2	1	1	ia, o(s)	2	-	-	da classificação	-	1	1
FILTR	2	1	-	agem(ns),o(s)	1	-	-	da literatura de qualidade	-	1	-
FOLHET	2	2	1	o(s)	2	2	1		-	-	-
FONT	3	5	1	e(s)	2	-	-	de informação	1	5	1
FORMAÇ	3	7	2	ão(ões)	2	-	1	profissional	1	7	1
FOR MAT	1	1	-	ação, o(s)	1	-	-	CALCO	-	1	-
FORMULA	1	-	-		-	-	-	de transição	1	-	-
FORNEC	2	2	-	edos(es), imento	-	-	-	da(e)informação de livros	1	1	-
FRANÇA	1	1	1		-	-	-		-	-	-
FRAS	1	-	-	e(s)	1	-	-		-	-	-

RAIZ

DESINÊNCIAS

Termos compostos
Pré - Coordenados

	I	C	K		I	C	K		I	C	K
FRENT	3	9	2	e(s)	-	-	-	de pesquisa	3	9	2
FREQUEN	2	5	-	cia	2	-	-	a biblioteca de descritores	-	4	-
FUNÇÃO	-	1	-		-	-	-	da biblioteca	-	1	-
FUTUR	-	1	1	ologia	-	1	1		-	-	-
GATEKEEPER	1	3	1		1	3	1		-	-	-
GEOLOGIA	1	1	1		1	1	1		-	-	-
GEORGES ANDERLA	1	-	1		1	-	1		-	-	-
GER	4	-	1	ação, ente(s)	4	-	1		-	-	-
GETULIO VARGAS	1	1	1		1	1	1		-	-	-
GOFFMANY	3	-	-		3	-	-		-	-	-
GOVERN	2	1	-	amental(aís),o	2	1	-		-	-	-
GRUPO	-	5	-		-	-	-	de elite	-	5	-
HABIT	2	-	-	o(s)	2	-	-		-	-	-
HARDWARE	-	2	-		-	2	-		-	-	-
HISTOR	5	1	1	ia, ica(s), ico(s)	5	1	1		-	-	-

RAIZ	DESINÊNCIAS			Termos compostos Pré-Coordenados						
	I	C	K	I	C	K				
HUMAN	2	-	2	a(s), idade(s), o(s)	2	-	2	-	-	-
IBBD	1	-	1		1	-	1	-	-	-
IBICT	4	2	3		4	2	3	-	-	-
IDIOM	1	-	-	e(s)	1	-	-	-	-	-
IMUNOLOGI	1	1	1	a,co(s)	1	1	1	-	-	-
INCIDÊNCIA	1	2	-		-	-	-	de citações	1	2
INDEPENDÊNCIA	1	-	-		-	-	-		-	-
INDEXA	10	33	5	ção, de(s), dor(es),do(s)	9	23	4	automática	1	6
								automática simulada	-	1
								coordenada	-	1
								pré-coordenada	-	1
INDIC	-	-	-	ativo	-	-	-	autoridade	-	1
								estado	-	1
								língua	-	1
								país	-	1
								território	-	1

RAIZ DESINÊNCIAS

Termos compostos
Pré-Coordenados

RAIZ	I	C	K	I	C	K	I	C	K
INDIC	-	-	-	-	-	-	1	2	1
			e(s)						
							1	1	-
							-	6	-
							-	1	-
							1	2	1
							-	2	-
				5	1	-	-	2	-
			es				-	1	-
							-	2	-
							-	1	-
							-	2	-
							-	1	-
							-	1	-
INDUSTR	1	2	-	1	-	-	-	1	-
			ia, iel(ais)						
							-	1	-
							-	1	-
INFANT	1	-	1	1	-	1	-	-	-
			il(is)						
INFORMAC	-	-	-	59	92	71	1	2	1
			ão(ões)						
							1	1	1
							1	2	1
							3	41	7
							-	2	-
							-	1	2

	I	C	K	I	C	K	I	C	K
INFORMAC	-	-	-	ã	o	s			
	65	169	87	ional	(1s)				
INFORMAT	4	7	4	ica		1	1	1	agrícola
INFRA-ESTRUTURA	2	-	-			-	2	-	
INGL	1	-	-	es	(es)	1	-	-	
INICIATIVA	-	2	-			-	-	-	PRIVADA, Publica
INOVAÇÃO	1	-	1			-	-	-	Tecnológica
INTITUIÇ	1	1	1	ão	(ões)	-	-	-	de pesquisa
INSTITUI	5	3	6	e,	o(s)	4	1	6	for Scientific informa- tion de pesquisa nacional do livro
INTERCÂMBIO	2	6	-			1	-	-	de informações
INTERCONSISTÊNCIA	1	1	1			1	1	1	
INTERDISCIPLINARIDADE	1	1	-			1	1	-	
INTRACONSISTÊNCIA	-	1	-			-	1	-	

RAIZ	DESINÊNCIAS			Termos compostos Pré-coordenados		
I C K	I C K	I C K	I C K	I C K	I C K	I C K
ISI	1 1 1	1 1 1	1 1 1	-	-	-
INTERNACION	6 - -	al (ais)	6 - -	-	-	-
INTERRELA	3 - -	ção (ões), cionada(s), cionado(s), cionamento(s)	3 - -	-	-	-
INTERSECÇÃO	1 - -		1 - -	-	-	-
ISA	1 - -		1 - -	-	-	-
ISI	2 1 1		1 1 1	1	-	-
ISO/R-639	1 - -		1 - -	-	-	-
JORN	3 1 1	al(ais)	1 1 1	1 científico(s) do Brasil	1	-
KWIC	1 - -		1 - -	-	-	-
LEI	9 22 1		7 - -	da dispersão da literatu- ra de Bradford de lotka de Zipf do Crescimento exponencial do elitismo	1	-
					8	1
					4	-
					7	-
					1	-
					1	-

RAIZ	DESINÊNCIAS			Termos compostos Pré-coordenados						
	I	C	K	I	C	K				
LEIT	2	2	-	2	-	-	do documento	2	-	
LEVANTAMENT	1	1	-	-	-	-	Bibliográfico(s)	1	1	
LING	-	-	-	7	10	4	portuguesa artificial (ais) de busca de indexação de recuperação documentária(s) jornalística(s) natural(ais)	3	1	2
LISA	20	20	10	-	-	-	uagem (ns) computacional	-	1	-
LITERAT	3	5	11	1	1	8	científica de geologia de química especializada siderurgica	1	1	1
LIVR	5	5	3	5	5	3	o(s)	-	-	-

RAIZ DESINÊNCIAS

Termos compostos
Pré-coordenados

	I	C	K	I	C	K	I	C	K
LOCALIZAÇÃO	1	-	-	1	-	-	-	-	-
LOTKA	1	-	-	1	-	-	-	-	-
LUIS FLOREN	1	1	1	1	1	1	-	-	-
MACROSSISTEMA	1	-	-	1	-	-	-	-	-
MACROTHESAURUS	1	1	2	1	-	1	da OCDE	1	1
MATER	4	-	-	4	-	-	ia, iac	-	-
MEDICINA	-	1	1	-	1	1		-	-
MÉDIAS	1	-	-	1	-	-		-	-
MEDIO	2	1	-	-	-	-	a(s)	de informação	1
MEDLARS	1	-	-	1	-	-		-	-
MEDLINE	1	-	-	1	-	-		-	-
MEI	1	1	-	-	-	-	o(s)	ambiente	1
MERCADO	-	1	-	-	-	-		de informação	1
MESTRADO	1	-	1	-	-	-		da informação	1
METOD	14	4	2	13	-	-	o(s), ologia	estatísticos	1
								linguísticos	3
									1

RAIZ	DESINÊNCIAS			Termos compostos Pré - coordenados		
I C K	I C K	I C K	I C K	I C K	I C K	I C K
MICROBIOLOGIA	1 1 1	1 1 1	1 1 1	-	-	-
MICROCOMPUTAD	- 1 -	or (es)	- 1 -	-	-	-
MICROF	1 1 4	ilme, imagem	1 1 4	-	-	-
MIT	1 - 1	o(s)	1 - 1	-	-	-
MODELO	- 1 -		- - -	-	-	1 -
MONOLINGU	- 1 -	e(s), ismo	- 1 -	-	-	-
MULTILINGU	1 - 1	e(s), ismo(s)	1 - 1	-	-	-
NACION	10 - 8	el (eis)	10 - 8	-	-	-
NÃO-USUAR	3 2 1	io(s)	3 2 1	-	-	-
NATIS	- 1 -		- 1 -	-	-	-
NB - 69	1 - -		1 - -	-	-	-
NECESSIDAD	- 14 -	e(s)	- - -	-	-	14 -
NIVEL	- 1 -		- - -	-	-	1 -
NOMENCLATURA	1 - -		1 - -	-	-	-
NORDESTE	1 1 1		1 1 1	-	-	-

RAIZ
DESINÊNCIAS
Termos compostos
Pré-coordenados

	I	C	K	I	C	K	I	C	K
NORMA	-	-	-	3	1	-	-	1	-
	(s), alização Técnica(s)								
NÚCLEO	1	-	-	1	-	-	-	-	-
NUMER	2	-	1	1	-	-	1	-	1
OBR	3	-	-	3	-	-	-	-	-
OCDE	1	-	1	1	-	1	-	-	-
ONLINE	1	-	1	1	-	1	-	-	-
OPERACION	1	-	-	-	-	-	1	-	-
ORGANIZAÇÃO	9	3	4	3	-	3	-	1	-
	do sistema de conceitos do sistema social Panamericana da Saúde								
ORIENT	3	-	3	3	-	3	-	-	-
	ação, adas(s), ado(s)								

RAIZ DESINÊNCIAS

Termos Compostos
Pré-coordenados

RAIZ	I	C	K	DESINÊNCIAS	I	C	K	Termos Compostos Pré-coordenados	I	C	K
PESQUIS								bibliométrica(s)	-	1	-
								Científica(s)	1	1	1
								da literatura	-	1	-
								em andamento	-	2	-
								em semântica	-	1	-
								industrial	1	1	1
								online	1	-	1
				odor(es)							
PESSOAL	1	-	-		1	-	-		-	-	-
PLANEJAMENTO	2	-	1		2	-	1		-	-	-
PNB-27	1	-	1		1	-	1		-	-	-
PNB-266	1	1	1		1	1	1		-	-	-
POLITIC	19	34	4		6	-	-	científica(s)	2	6	1
								científica e tecnológica	1	-	-
								de administração	1	-	-
								de aquisição	1	1	-
								de informação	11	19	2
								editorial	1	2	-
				o(s)				nacional de informação	3	5	1

RAIZ	I	C	K	DESINÊNCIAS	I	C	K	Termos compostos Pré - coordenados	I	C	K
POPULAÇÃO	2	-	-		2	-	-		-	-	-
PORTUGU	1	-	-	e(s);es(a); eses(as)	1	-	-		-	-	-
PRAGMÁTICA	-	1	-		-	1	-		-	-	-
PRECISÃO	-	1	1		-	1	1		-	-	-
PRESERVAÇÃO	1	-	-		1	-	-		-	-	-
PRIVA	1	1	-	cidade,de(s), do(s)	-	-	-	da informação	-	1	-
PROCESS	-	3	-	amento, o(s)	-	-	-	de dados distribuídos de comunicação	-	1	-
PRODUC	15	12	-	ão	-	9	9	científica do sistema tecnologia absoluta de autor de periódicos do sistema relativa	6	9	-
PRODUTIV	2	6	-	idade	-	-	-		1	1	-
									-	2	-
									-	1	-
									-	1	-
									-	1	-
									1	1	-

RAIZ	DESINÊNCIAS			Termos compostos Pré - coordenados		
	I	C	K	I	C	K
PRODUT	3	1	- o(s)	3	1	-
PROFESS	1	1	- or (er)	1	1	-
PROFISS	6	-	2 ão(ões), ional (ais)	6	-	2 da informação
PROGRAM	7	-	2 a(s)	-	5	- educacional internacionais
PROJET	4	-	4 o(s)	4	-	1 CAPEQ. INPUT MARC
PUBLIC	8	-	3 a(s), ação(ões), o(s)	-	-	-
PUC/RJ	2	-	1	2	-	1
QUALID	1	-	2 ade(s)	1	-	2
QUALIF	1	-	- ação (oês)	1	-	-
QUANTIC	1	-	- ação, ar	1	-	-
QUÍMICA	2	3	3	2	3	3
RACIONAC	1	-	- ização	1	-	-

RAIZ DESINÊNCIAS

Termos compostos
Pré-coordenados

I C K

RAIZ	I	C	K	DESINÊNCIAS	I	C	K	Termos compostos Pré-coordenados	I	C	K
RANGANATHAN	2	1	1		2	1	1		-	-	-
RECEPTOR	2	1	-		2	1	-		-	-	-
RECOMEND	2	-	-	ação(oes), ar	2	-	-		-	-	-
RECUPERA	20	29	4	ção, da(s), do(s)	9	-	-	automática	1	-	-
								da(s) informação(ões) de informação(ões)	10	29	4
RECURS	6	-	3	o(s)	-	-	-	audiovisuais	1	-	-
								de informação humanos	-	3	-
RED	8	25	7	e(s)	5	-	5	de bases de dados	-	1	-
								de biblioteca(s)	-	1	-
								de comunicação(ões)	-	1	-
								de informação	-	20	1
								EUROPEA de informação	-	2	1
								NACIONAL	1	-	-
								Nacional de informação	1	-	1
REFERENC	1	-	2	ia(s)	-	-	-	bibliográfica(s)	1	-	-
REFORMA	1	-	1		-	-	-	Biblioteca Nacional	1	-	1

RAIZ DESINÊNCIAS

Termos compostos
Pré-coordenados

I C K I C K

REGION	2	-	2	al(a)is)	2	-	2	-	-	-	-	-
REGISTR	2	3	-	o(s)	1	-	-	-	-	-	1	-
REGR	1	2	2	a(s)	1	1	1	1	1	1	1	1
RELAC	2	-	-	ão (ões)	1	-	-	-	-	1	-	-
RELEVAN	6	3	2	cia, te(s)	6	3	2	-	-	-	-	-
REPRESENTAC	1	-	1	ão (ões)	1	-	1	-	-	-	-	-
REPROGRAF	-	-	1	ia	-	-	1	-	-	-	-	-
RESUM	1	1	-	o(s)	1	1	-	-	-	-	-	-
REVISÃO	6	-	2	(oes)	3	-	-	-	-	3	-	2
REVISTA	1	-	-	(s)	1	-	-	-	-	-	-	-
REVOCAÇÃO	1	1	1		1	1	1	-	-	-	-	-
ROTA	1	-	-	ção, da(s), do(s)	1	-	-	-	-	-	-	-
RUIO	-	1	-	(s)	-	-	-	-	-	-	1	-
SATISF	-	1	-	eção	-	-	-	-	-	-	1	-

RAIZ DESINÊNCIAS

Termos compostos
Pré-coordenados

	I	C	K	I	C	K	I	C	K
SCIENCE CITATION INDEX	-	1	-	-	1	-	-	-	-
SDI/CIN/CNEN	1	-	1	1	-	1	-	-	-
SEC	-	1	1	ão (ões)	-	-	do(s) documento(s)	-	1
SEGURANÇ	-	1	1	a (s)	-	-	da informação	-	1
SELEÇÃO	3	3	-		3	2	de usuários	-	1
SEMANTIC	8	9	4	a(s), lista(s), o(s)	8	9	4	-	-
SERVIÇ	-	25	15	o(s)	-	-	10	de alerta	-
								de alerta corrente	-
								de informação	-
								de SDI	-
								online	-
								de biblioteca	-
								de referência	-
								de utilidade pública	-
SIDERURGIA	2	1	2		1	-	1	brasileira	1
SIEF	1	-	-		1	-	-		-
SIGNIFIC	1	2	-	ção, ado	1	2	-		-

RAIZ	DESINÊNCIAS			I	C	K	Termos compostos Pré coordenados	I	C	K
	I	C	K							
SIMILARIDAD	1	-	-	1	-	-		-	-	-
SINTAXE	1	4	-	1	4	-		-	-	-
SISTEM	38	60	-	23	-	-	de documentação de indexação de informação de informação agrícola MEDLINE PASCAL UNIDEK ONLINE	-	1 1 54 2 1 1 1 1	-
SNICT	1	-	-	1	-	-		-	-	-
SOCI	9	1	5	9	-	5	al(a)s), idade informatizada	-	1	-
SOFTWARE	-	2	-	-	1	-	de aplicação	-	1	-
SPIRIT	1	-	-	1	-	-		-	-	-
SUBCOMISSÃO	1	-	-	1	-	-		-	-	-
SUBDESENVOLVIMENTO	1	-	1	1	-	1		-	-	-
SUBJETIV	1	-	-	1	-	-	a(s), idade,o(s)	-	-	-

RAIZ	DESINÊNCIAS			Termos compostos Pré-coordenados		
	I	C	K	I	C	K
TABEL	1	-	1	1	-	1
TECNIC	12	-	1	11	-	1
TECNOLOGI	19	24	18	17	21	16
TEMA	1	1	1	-	-	-
TEORI	8	19	6	7	1	4
TERMINOLOGI	2	4	-	1	3	-
TERMO	4	1	-	3	-	-
TERRITORIO	1	1	-	1	1	-

RAIZES	DESINÊNCIAS			Termos compostos Pré-coordenados							
	I	C	K	I	C	K					
TESAURO	-	7	-	-	6	-	-	1	-	-	
TESTE	1	-	-	1	-	-	-	-	-	-	
TEXTO	1	-	-	1	-	-	-	-	-	-	
TIP	2	-	1	o(s), ologia(s)	2	-	-	de erros	-	1	-
TITUL	3	-	1	o(s)	3	-	1	-	-	-	
TRADU	1	-	-	cao(oes), todologia	1	-	-	-	-	-	
TRANSCIENCIA	1	-	1	-	1	-	1	-	-	-	
TRANSFER	6	19	3	encia	-	-	-	de informação de tecnologia	6	13	3
TRANSIÊNCIA	-	2	-	-	-	2	-	-	-	-	-
TRANSMISS	1	-	-	ão(ões)	1	-	-	-	-	-	-
TRATAMENTO	3	2	-	(s)	2	-	-	bibliográfico(s) da(s) informações	1	-	-
TREINA	1	3	-	mento	1	-	-	bibliotecário do(s) bibliotecário(s) profissional	-	1	-
									-	1	-

RAIZES	DESINÊNCIAS			Termos compostos Pré-coordenados					
	I	C	K	I	C	K			
TROCA	1	-	1	-	-	-	1	-	1
TRUNCA	1	1	-	-	-	-	1	-	-
								1	-
UFMG	1	-	1	1	-	1	-	-	-
UNESCO	1	1	1	1	1	1	-	-	-
UNIDADE	2	-	1	(s)	-	-	-	1	-
UNIDEK	1	1	1	1	1	1	-	-	-
UNIFICAÇÃO	1	-	-	1	-	-	-	-	-
UNIVERSI	4	-	3	dade(s), taria(s)	3	-	3	Federal de Pernambuco	1
				tário(s)					
UNIVERS	1	-	-	(os)	1	-	-	-	-
US	10	38	7	o(s), uário(s)	10	34	7	de biblioteca(s)	1
								em potencial	3
UTILID	-	3	1	ade(s)	-	-	1	da informação	1
								pública	1
								pública de informação	1

DESINÊNCIAS

Termos compostos
Pré-coordenados

RAIZES

	I	C	K		I	C	K		I	C	K
UTILIZ	-	13	-	ação (oes)	-	-	-	de(s) coleção (oês)	-	1	-
VALOR	-	4	-		-	-	-	de(s) informação(oês)	-	12	-
VEÍCULO	1	-	1		1	-	1	da informação	-	4	-
VELOCIDADE	1	-	-		-	-	-		-	-	-
VIDA	-	1	-		-	-	-	MÉDIA	-	1	-
VOCABULARI	1	7	-	o(s)	1	3	-	especializado(s)	-	2	-
								monolingue(s)	-	1	-
								multilingue(s)	-	1	-
XIST	2	1	1	o(s)	1	1	1	oleigenos	1	-	-
ZIPF	1	-	-		1	-	-		-	-	-

10.4 ANEXO 4

Buscas Bibliográficas

ESTRATÉGIAS DE BUSCA	INDEXCI		CINFORM			Referências Coincidentes		
	R	Rf	R	Rf	Ri	Rr	Ri	TOTAL
1. Administração de sistemas de informação	11	8	4	2	2	2	2	4
1a. (Administ& / Plane&) / Organiz&	11	8	4	2	2	2	2	4
1b. (Administração / Planejamento) & Organização do Sistema	2	2	1	1	-	-	-	-
2. (análise de conteúdo e análise temática da informação	15	13	36	17	13	13	2	15
2a. (Análise da inf. / análise do documento)/análise documental	-	-	14	5	-	-	-	-
2b. (Análise de assunto&/Análise de Conteúdo)/ Análise do Conteúdo	2	2	-	-	-	-	-	-
2c. (Indexaç & / Linguage&)	14	12	28	17	12	12	2	14
3. Aquisição planificada	1	1	2	1	-	1	-	1
3a. (Aquis&)	1	1	2	1	-	1	-	1
3b. (Aquisição planificada)	1	1	1	1	-	1	-	1
4. Automação	19	12	30	20	12	12	1	13
4a. (Informat& / Autom&)	19	12	30	20	12	12	1	13
4b. (Informática / Automação)	8	8	29	19	8	8	-	8
4c. (Automac& / Automat&)	9	8	27	18	8	8	1	9
4d. (Autom & + Bibliot &)	4	3	9	4	3	3	1	4
4e. (Autom& + Bibliot&) + Sistem&	2	1	6	4	1	1	-	1

ESTRATÉGIAS DE BUSCA	INDEXCI		CINFORM		Referências Coincidentes		
	R	Rf	R	Rf	Rf	Ri	TOTAL
5. Aveliação de sistemas de in- formação	5	3	3	1	1	1	2
5a. (Sistema& / Bibliotec&) + Avalia&	5	3	3	1	1	1	2
5b. (Rede& / Centro&) + Avalia&	-	-	-	-	-	-	-
6. Bibliografias especializadas em Ciência e Tecnologia.	5	3	5	2	2	2	4
6a. (Bibliografia& + Ciência e Tecnologia)	-	-	-	-	-	-	-
6b. (Bibliograf& + Biologia)/Imu- nologia	1	-	1	-	-	1	1
6c. (Bibliograf& + Engenharia Química) / Engenharia	1	-	1	-	-	1	1
6d. (Bibliograf& + Física/Side- Purg&)	1	1	1	-	-	-	-
6e. (Bibliografie + Xiglo). / Botê- nica.	2	2	2	2	2	-	2
6f. (Biologia / Imunologia) + Bi- biograf&	-	-	-	-	-	-	-
6g. (Engenharia / Engenharia Quí- mica) + Bibliografie&	-	-	-	-	-	-	-
6h. (Xisto/Botânica) + Bibliogra- fia&	2	2	2	2	2	-	2
7. Bibliometria - lei Zipf, Lei de Bradford	12	12	20	17	11	-	11
7a. (Lei de Bradford / Lei de Zipf) / Bibliom&	11	11	20	17	11	-	11
7b. (Bradford / Zipf) / Bibliom&.	9	9	18	16	9	-	9
7c. (Lei de Zipf / Lei de Bradford Bibliom&	1	1	8	8	1	-	1

	ESTRATÉGIAS DE BUSCA	INDEXCI		CINFORM		Referências Coincidentes		
		R	Rf	R	Rf	Rf	Ri	TOTAL
8.	Bibliometria aplicada à Medicina e Saúde.	-	-	-	-	-	-	-
8a.	(Medic& / Ciência Saúde)+Bibliom&.	-	-	-	-	-	-	-
8b.	(Anali& + Bibl&) + Medic&	-	-	-	-	-	-	-
8c.	(Anali& + Bibl&) + Saúde	-	-	-	-	-	-	-
8d.	(Anali& + Bibl&) + Ciências da Saúde.	-	-	-	-	-	-	-
8e.	(Biomed& + Bibl)	-	-	-	-	-	-	-
9.	Bibliotecário biomédico	-	-	-	-	-	-	-
9a.	(Bibliotecários biomédicos / Bibliotecário biomédico)	-	-	-	-	-	-	-
9b.	(Bibl& + Biomed&)	-	-	-	-	-	-	-
10.	Biblioteconomia/Ciência da Informação - aspectos sociais	3	3	3	2	1	-	1
10a.	(Ciência da Informação/Bibliot& + Soci&	3	3	1	-	-	-	-
10b.	(Ciência da Informação / Biblioteconomia)+Epistem.Soci&	1	1	1	1	1	-	1
10c.	(Ciência da Informação / Biblioteconomia) + Agência Social	-	-	1	1	-	-	-
11.	Biblioteconomia / Ciência da Informação-conceito Filosofia	6	4	9	5	4	2	6
11a.	(Bibliot& + Ciência da Informação).	6	4	9	5	4	2	6
11b.	(Bibliot& / Ciência da Informação) + Conceit&	1	1	1	-	-	-	-

	ESTRATÉGIAS DE BUSCA	INDEXCI		CINFORM		Referências Coincidentes		
		R	Rr	R	Rr	Rr	Ri	TOTAL
12.	Catologação teoria, filosofia Classificação	11	9	16	9	9	2	11
12a.	(Catalog & / Classific &)	11	9	16	9	9	2	11
12b.	(Catalogação / Classificação)	3	3	12	8	3	-	3
12c.	(Catalog & / Classific &) + Filosofia	1	1	1	1	1	-	1
12d.	(Catalog & / Classific &) + Teoria	2	2	1	1	1	-	1
12e.	(teoria & / Conceito &) + Classificação	3	3	2	2	1	-	1
13.	Coleções & desenvolvimento	1	1	1	1	1	-	1
13a.	(Colec &)	1	1	1	1	1	-	1
14.	Edson: Nery de Fonseca - estudos biográficos	-	-	-	-	-	-	-
15.	Epistemologia aplicada a Bi- blioteconomia	1	1	1	1	1	-	1
15a.	(Epistemologia + Bibliotecono- mia)	1	1	-	-	-	-	-
14b.	(Epistemologia Soc & + Biblio- teca)	1	1	1	1	1	-	1
16.	Estágio	3	-	-	-	-	-	-
16a.	(Estágio)	3	-	-	-	-	-	-
17.	Estudos de usuário	10	8	34	12	8	2	10
17a.	(Estudo de usuário / Estudo de usuários)	-	-	4	1	-	-	-

	ESTRATÉGIAS DE BUSCA	INDEXEXCI		CINFORM		Referências Coincidentes		
		R	Rr	R	Rr	Rr	Ri	TOTAL
17b.	(Estudos de usuários / Estudo de usuários)	1	1	-	-	-	-	-
17c.	(Estudos de usuário / Usuário / Estudos de usuário)	10	8	34	12	8	2	10
18.	Formação Profissional	6	5	7	5	4	-	4
18a.	(Formação profissional/Educação profissional)	1	1	7	5	1	-	4
18b.	(Formação profissional/Formação) / Profissional	6	5	7	5	4	-	4
18c.	(Ensino)	1	1	-	-	-	-	-
19.	Indexação automática	1	1	13	2	1	-	1
19a.	(Indexação + Automática) / Indexação automática	1	1	13	2	1	-	1
20.	Informação científica e tecnologia	26	16	24	19	16	2	18
20a.	(Inf. tecnol / inf. científica)	26	16	24	19	16	2	18
21.	Linguística e Ciência da Informação terminologia	10	10	13	12	10	-	10
21a	(Linguística / Semântica) / Sintaxe	10	10	13	12	10	-	10
21b.	(Linguística/Semântica)/Sintaxe	4	4	13	12	4	-	4
21c.	(Linguística + Semântica) + Sintaxe	1	1	4	4	1	-	1
21d.	(Linguística + Ciência da Inf.) + index	-	-	2	2	-	-	-
22.	Matemática aplicada e Biblioteconomia	2	2	3	3	1	-	1

ESTRATÉGIAS DE BUSCA	INDEXCI		CINFORM		Referências Coincidentes		
	R	Rf	R	Rf	Rf	Ri	TOTAL
22a. (Matemática + Bibliotec&)	-	-	-	-	-	-	-
22b. (Método estatísticos /método estatístico)	1	1	3	3	1	-	1
22c. (Algebr&)	1	1	-	-	-	-	-
23. Metodologia, métodos de pesquisa	-	-	-	-	-	-	-
23a. (Metodo& + pesquisa&)	-	-	-	-	-	-	-
24. Pesquisa em Biblioteconomia e Ciência da Informação	3	1	8	2	1	-	1
24a. (Ciência da Informação / Bibliotec& + Pesquisa.	3	1	8	2	1	-	1
24b. (Pesquisa-científica /Pesquisa) + Ciência da Informação	2	1	3	1	1	-	1
24c. ((Pesquisa científica/Pesquisa) + Biblioteconomia.	1	1	1	1	1	-	1
25. Política da informação	10	10	22	19	10	-	10
25a. (polit& + informac&)	10	10	22	19	10	-	10
25b. (Política-cientif& /Política tecnológica)/Política de inf.	6	5	22	19	5	-	5
25c. (Informação, científ&/inf. tecn&) + polit&	1	1	18	15	1	-	1
26. Promoção (marketing) de bibliotecas	-	-	-	-	-	-	-
27. Seleção	3	2	3	2	2	1	3
27a. (Selec&)	3	2	3	2	2	1	3

	ESTRATÉGIAS DE BUSCA	INDEXCI		CINFORM		Referências Coincidentes		
		R	Rf	R	Rf	Rf	Ri	TOTAL
27b.	(Seleção)	3	2	2	2	2	-	2
28.	Serviços de referência	2	-	3	2	-	-	-
28a.	(Serviço de referência / serviços de referência)/Referências	2	-	3	2	-	-	-
29.	Sistema de inf., biblioteca, centro de inf., centro de doc house, centro ref.-conceitos.	50	14	81	16	14	25	49
29a.	(Sistema&/Bibliote&)/Centro&	50	14	81	16	14	25	49
29b.	(Sistemas de inf./Sistema de inf.) /Clearinghouse	10	4	54	13	4	6	10
29c.	(Conceito de biblioteca/Centros de inf.)/C.referencial	1	-	2	1	-	-	-
29d.	(Sistema& / rede&)+ Conceit&	2	1	1	1	-	-	-
30.	Tesouro e descritor composto	-	-	3	1	-	-	-
30a.	(Thes& / Tes&) + Descri&	-	-	1	-	-	-	-
30b.	{descritores/descriptor} + tesouro	-	-	1	-	-	-	-
30c.	{Tesouro + Descritores}/ Descritor.	-	-	3	1	-	-	-
31.	Treinamento de usuários	1	-	4	1	-	1	1
31a.	(Usuar&/Bibliot&) + Trein&	1	-	4	1	-	1	1
32.	Uso da inf. e bibliotecas	6	6	37	21	6	-	6

10.5 ANEXO 5

Índices médios de precisão
na base de dados INDEXCI

Anexo 5 - Índices médios de precisão na
base de dados INDEXCI

COMPARAÇÃO ENTRE BASES MANUAIS E AUTOMÁTICA
TESE SIMONE - BIBLIOTECONOMIA
FILE NOMATE (CREATION DATE : 11/09/84)

VO5 ÍNDICE DE PRECISÃO

CATEGORY LABEL	CODE	ABSOLUTE FREQ.	RELATIVE FREQ. (PCT)	ADJUSTED FREQ. (PCT)	COM. FREQ. (PCT)
	0.000	3	11.1	11.1	11.1
	0.280	1	3.7	3.7	14.8
	0.333	1	3.7	3.7	18.5
	0.600	2	7.4	7.4	25.9
	0.615	1	3.7	3.7	29.6
	0.632	1	3.7	3.7	33.3
	0.667	2	7.4	7.4	40.7
	0.727	1	3.7	3.7	44.4
	0.800	2	7.4	7.4	51.9
	0.818	1	3.7	3.7	55.6
	0.833	1	3.7	3.7	54.3
	0.867	1	3.7	3.7	63.0
	1.000	10	37.0	37.0	100.0
	TOTAL	27	100.0	100.0	
MEAN	0.713				
MODE	1.000	0.063	MEDIAN	0.804	
KURTOSIS	0.400	0.323	VARIANCE	0.107	
MINIMUM	0.000	-1.151	RANGE	1.000	
VALID CASES	27	1.000			
MISSING CASES	0				

10.6 ANEXO 6

Índices médios de precisão
na base de dados CIFORM

Anexo 6 - Índices médios de precisão na
base de dados CIFORM

COMPARAÇÃO ENTRE BASES MANUAIS E AUTOMÁTICA
TESE SIMONE - BIBLIOTECONOMIA
FILE NOMATE (CREATION DATE : 11.09.64)

V05 ÍNDICE DE PRECISÃO

CATEGORY LABEL	CODE	ABSOLUTE FREQ.	RELATIVE FREQ. (PCT)	ADJUSTED FREQ. (PCT)	COM. FREQ. (PCT)
	0.154	1	3.7	3.7	3.7
	0.193	1	3.7	3.7	7.4
	0.250	2	7.4	7.4	14.8
	0.333	2	7.4	7.4	22.2
	0.353	2	7.4	7.4	29.6
	0.400	1	3.7	3.7	33.3
	0.472	1	3.7	3.7	37.0
	0.500	2	7.4	7.4	44.4
	0.556	1	3.7	3.7	48.1
	0.563	1	3.7	3.7	51.9
	0.668	1	3.7	3.7	54.0
	0.667	1	14.8	14.8	70.4
	0.714	1	3.7	3.7	74.1
	0.792	1	3.7	3.7	77.8
	0.850	1	3.7	3.7	81.5
	0.864	1	3.7	3.7	85.2
	0.923	1	3.7	3.7	88.9
	1.000	3	11.1	11.1	100.0
TOTAL		27	100.0	100.0	
MEAN	0.577				
MODE	0.667				
KURTOSIS	-1.011				
MINIMUM	0.164				
STD ERR	0.050				
STD DEV	0.257				
SKLWNESS	0.161				
MAXIMUM	1.000				
MISSINGN CASES	0				
VALID CASES	27				
			MEDIAN	0.563	
			VARIANCE	0.066	
			RANGE	0.846	