

COMPARAÇÃO ENTRE UM ÍNDICE KWOC (KEYWORD-OUT-OF-CONTEXT) E UM ÍNDICE  
EM CADEIA DERIVADO DA CDU (CLASSIFICAÇÃO DECIMAL UNIVERSAL)

Gláucia H. B. Pereira de Sousa

CURSO DE MESTRADO EM CIÊNCIA DA  
INFORMAÇÃO (IBBD/UFRJ)  
1972/1973

Dissertação apresentada ao IBBD/UFRJ  
para obtenção do grau de mestre em Biblioteconomia  
e Documentação.

Orientador: Prof. Gilda M. Braga

Rio de Janeiro, 1975

Para

Flávio, por seu incentivo, sua  
compreensão, seu exemplo

Gilberto, meu maior obstáculo e  
também meu maior estímulo

Mamãe e Papai, por me transmitirem  
o prazer de estudar

## RESUMO

Neste estudo comparou-se dois sistemas de indexação que haviam sido empregados na coleção de livros da biblioteca da FINEP - Financiadora de Estudos e Projetos: um índice KWOC (Keyword-out-of-Context) e um índice em cadeia derivado da CDU (Classificação Decimal Universal). Foi também comparado ao índice em cadeia um segundo índice KWOC formado pela união de cada duas palavras-chave consecutivas dos títulos.

Dois métodos diversos foram utilizados nesta comparação: 1) medida de similaridade dos índices, onde o cabeçalho de nível 1 do índice em cadeia foi tido como padrão e os títulos a ele comparados, tendo cada palavra ou frase-chave do título recebido um valor da escala de 100%, 90%, 80%, 70% e 0% segundo o maior ou menor grau de sua similaridade; 2) medida de especificidade dos índices onde se apurou o número de termos especificadores de cada um para cada obra analisada.

Apurou-se na medida de similaridade dos índices a média geral (títulos comparados aos cinco primeiros cabeçalhos da cadeia indexadora) de 19,03% para as palavras-chave dos títulos e 22,61% para as frases-chave. A média geral obtida com a comparação dos títulos aos cabeçalhos do nível 1 da cadeia, tidos como padrão, foi de 38,85% (palavras-chave) e 47,41% (frases-chave). Todas essas médias atestaram o desentorno de vocabulário entre os dois índices, e uma análise dos termos-chave que receberam o mais baixo valor de similaridade, 0%, mostrou que a maioria deles eram significativos, o que contou ponto para os títulos como indexadores.

No segundo método verificou-se ser o índice em cadeia no caso estudado mais específico do que os termos-chave dos títulos.

Concluiu-se não serem possíveis as generalizações dos resultados, mas sim dos métodos e que qualquer sistema de indexação pode ser até certo ponto manuseado para servir melhor às necessidades dos usuários de uma dada coleção.

## SUMÁRIO

1. Introdução.....	11-26
1.1. - Representação do documento: o problema da indexação...	11-14
1.2. - Índices rotados: KWIC (Keyword-in-Context) e KWOC (Keyword-out-of-Context).....	14-16
1.3. - CDU: o Sistema de Classificação Decimal Universal.....	16-17
1.4. - Índice em cadeia.....	17-19
1.5. - Metodologia para avaliação de índices.....	19-26
2. Material.....	27-33
3. Metodologia.....	35-40
3.1. - Apuração do nível de similaridade entre as palavras e frases-chave dos títulos e os cabeçalhos da cadeia indexadora.....	35-39
3.2. - Apuração do nível de especificidade destes dois índi ces.....	39-40
4. Resultados e Comentários.....	41-78
4.1. - Nível de similaridade.....	41-75
4.2. - Nível de especificidade.....	76-78
5. Conclusões.....	79-82
6. Citações Bibliográficas.....	83-85
Apêndice 1 - Relação dos títulos dos livros por ordem numé rica e cabeçalhos da cadeia com as apurações de similaridade e especificidade	
Apêndice 2 - Amostras das listagens do índice KWOC:	
2.1.- Bibliografia dos 100 livros	
2.2.- Palavras-chave dos títulos	
2.3.- Frases-chave dos títulos	
Apêndice 3 - Relação alfabética das palavras-chave dos tí tulos	
Apêndice 4 - Relação das palavras e frases-chave dos títu los resultantes do índice reformado	
Apêndice 5 - Classes de assunto da amostra segundo a CDU	

## TABELAS

1. Distribuição de frequência dos assuntos.
2. Distribuição de frequência dos valores de similaridade das palavras-chave dos títulos versus cabeçalhos de nível 1 da cadeia indexadora.
3. Distribuição da frequência dos valores de similaridade das palavras-chave dos títulos versus cabeçalhos do nível 2 da cadeia indexadora.
4. Distribuição da frequência dos valores de similaridade das palavras-chave dos títulos versus cabeçalhos de nível 3 da cadeia indexadora.
5. Distribuição da frequência dos valores de similaridade das palavras-chave dos títulos versus cabeçalhos do nível 4 da cadeia indexadora.
6. Distribuição da frequência dos valores de similaridade das palavras-chave dos títulos versus cabeçalhos do nível 5 da cadeia indexadora.
7. Distribuição da frequência dos valores de similaridade das palavras e frases-chave dos títulos versus cabeçalhos de nível 1 da cadeia indexadora.
8. Distribuição da frequência dos valores de similaridade das palavras e frases-chave dos títulos versus cabeçalhos de nível 2 da cadeia indexadora.
9. Distribuição de frequência dos valores de similaridade das palavras e frases-chave dos títulos versus cabeçalhos de nível 3 da cadeia indexadora.
10. Distribuição de frequência dos valores de similaridade das palavras e frases-chave dos títulos versus cabeçalhos do nível 4 da cadeia indexadora.
11. Distribuição de frequência dos valores de similaridade das palavras e frases-chave dos títulos versus cabeçalhos do nível 5 da cadeia indexadora.
12. Valores de similaridade dos títulos versus diversos níveis da cadeia indexadora antes e depois de formação de frases-chave.
13. Valores de similaridade dos títulos versus cabeçalhos do nível 1 da cadeia indexadora.

Influência do valor de similaridade 0% das palavras-chave dos títulos sobre a sua média geral.

14. Valores de similaridade dos títulos após a formação de frases chave versus cabeçalhos do nível 1 da cadeia indexadora.

Influência dos valores 0% sobre a média apresentada.

15. Diferença entre os valores de similaridade atribuídos às palavras-chave dos títulos versus cabeçalhos de nível 1 da cadeia indexadora e às frases e palavras-chave dos títulos em comparação com estes mesmos cabeçalhos.

16. Comparação entre as palavras-chave dos títulos e o número de frases-chave formadas com indicação da diferença entre os dois valores de similaridade obtidos no cabeçalho de nível 1 da cadeia indexadora

17. Moda dos valores de similaridade das palavras-chave dos títulos com os diversos níveis da cadeia indexadora

18. Comparação entre o número de obras com maior nível de especificidade nos títulos ou nos cabeçalhos dos diversos níveis da cadeia indexadora, ou em ambos.

## 1. INTRODUÇÃO

"Indexação e recuperação da informação permanecerão por um longo tempo tanto uma arte como uma ciência".

Nada mais acertada, a nosso ver, do que esta afirmação de Moss sobre o complexo assunto da recuperação da informação e mais ainda sobre o seu problema básico, o tão conhecido gargalo dos sistemas de informação, o índice.

A indexação utiliza primordialmente a linguagem para expressar ou comunicar pensamentos, idéias e conhecimento. Nenhum destes elementos, porém, pode ser satisfatoriamente definido, pois há na linguagem, seu veículo de expressão, não somente um elemento factual, mas também uma emoção ou sentimento. Mesmo no trabalho científico, onde estes últimos aspectos não são, (ou, ao menos, não deveriam ser), motivos de preocupação, eles sub-repticiamente se apresentam em estilos de redação técnica que mostram, por exemplo, uma certa dissimulação, que alguma afinidade tem com emoção ou arte. Daí a afirmação que a ciência com a qual cientistas da informação lidam não ser nunca inteiramente científica.

Empenhamo-nos, pois, neste trabalho, através de uma avaliação comparativa de dois índices, a colaborar com os demais cientistas da informação na transformação desta arte em ciência.

### 1.1. Representação do documento: o problema da indexação.

O crescente volume da literatura publicada é um assunto que vem sendo levantado e debatido continuamente nos últimos anos assim como o problema que ele traz para o tratamento e representação da informação, onde as mais variadas soluções vem sendo perseguidas.

Richmond afirma ser comum dividir-se o campo da representação da informação em duas grandes partes: a primeira incluindo sistemas baseados em classificação estruturada a partir de várias teorias

e com formatos que usualmente estão relacionados a um código de manipulação; e a segunda parte incluindo sistemas relacionadas basicamente com palavras e cadeias de palavras (word strings). Estes dois tipos de sistema estão, entretanto, intimamente relacionados e o uso de um excluindo inteiramente o outro, é praticamente impossível. - Sistemas de classificação não indexados são desajeitados e frustrantes, da mesma forma que sistemas de indexação sem uma estrutura são ambíguos e indefinidos.

O nosso trabalho é um exemplo da inter-relação entre estes dois tipos de sistemas: - comparamos o sistema de indexação de "palavras chave-fora-do-contexto" (Keyword-out-of-context-KWOC) com o sistema de indexação em cadeia feito para o sistema de classificação Decimal Universal (CDU). O índice KWOC existe por si só, mas o índice em cadeia é um complemento de um sistema de classificação, quer seja a CDU (no caso), quer seja uma classificação facetada, ou outras - mais.

Os sistemas de classificação, por sua vez, dividem-se também em duas categorias principais: classificações semânticas e classificações matemáticas (formadas através da aplicação de métodos matemáticos ou estatísticos).

A classificação Decimal Universal situa-se entre as classificações semânticas. Outros exemplos deste tipo são a classificação Decimal de Dewey, a Classificação dos Dois Pontos de Ranganathan (Colon Classification), etc.

Os sistemas de indexação são em grande numero, sendo o KWOC e a Indexação em Cadeia, dois deles. Pesquisas sobre todos os aspectos da indexação proliferam em nossos dias e elas giram em torno dos inúmeros aspectos que cercam o processo de indexação, desde a sua elaboração até sua avaliação e interação com o usuário.

O índice, segundo Lancaster, é essencialmente um filtro, seu propósito sendo deixar passar os documentos procurados e segurar os não procurados. Para tanto, ele deve ser construído de forma a

refletir as características e requisitos dos usuários da coleção de documentos.

Muitos aspectos são então analisados na elaboração de uma linguagem de índice. Segundo Bloomfield estes aspectos seriam seis: extensão (breadth) do vocabulário, profundidade do vocabulário, uso de termos gerais ou específicos, uso de remissivas e referências - "ver" e "ver também", formato do índice, e inclusão de títulos ou outras frases qualificadoras.

Segundo Lancaster uma linguagem de índice completa deveria compor-se das seguintes partes:

1. Um conjunto de termos - índice;
2. Um conjunto de termos - código;
3. Um conjunto de termos - entrada;
4. Medidas (devices) usadas em associação com os termos-índice para ampliar ou restringir a definição da classe;
5. Regras para a utilização da linguagem de índice.

Entre as medidas (devices) mais comuns ele cita o controle da sinonímia, da quase-sinonímia, da forma da palavra, a ligação hierárquica, agrupamento (clumping) e aglomeração (cachos) (clustering), todos para permitir maior revocação (recall); cita também as medidas para precisão, que são a coordenação, o peso, o enganchamento (interlocking) e os indicadores de função (role indicators).

Outros aspectos são também estudados para a elaboração de índices, como o seu custo, a consistência dos diversos indexadores, o seu conhecimento ou não do assunto indexado, os diversos problemas da indexação automática, etc.

Pesquisas sobre todos estes aspectos contribuem, sempre de algum modo, por mais insignificante que seja, para o aperfeiçoamento dos diversos sistemas de indexação. Certos problemas, porém, persistem e são flagrantes, como a desatualização dos nossos sistemas de classificação (muitos estão por pouco fora do século dezenove). Co-

mo bem frisou Richmond, há necessidade de uma filosofia organizacional de maior realce para o fim do século vinte, que nos permita inserir em nossos sistemas dispositivos ou medidas para tornarem estes sistemas dinâmicos, permitindo um ajustamento e reação contínuos à mudança.

Será isto possível? Haverá uma maneira de descrever uma realidade objetiva em mudança contínua?

Mesmo no campo da indexação, outros aspectos nos parecem de difícil solução, como entre muitos outros as variações nos vocabulários: sinonímia, quase sinonímia, etc. Como controlá-los? É possível elaborar-se definições universalmente aceitas? Parece-nos, sem dúvida, uma arte.

## 1.2. Índices rotados\*: KWIC (Keyword-in-Context) e KWOC (Keyword out-of-Context).

Embora não tenha sido o primeiro a despertar para o problema da utilização dos títulos como indexadores do conteúdo das obras, através da sua rotação, ciclagem ou permutação, foi Hans Peter Luhn o primeiro a estudá-lo tendo acesso à tecnologia computacional e sem dúvida foi quem mais interesse despertou para este problema e mais adeptos de sua aplicação atraiu. Através de seu artigo "Keyword-in-Context Index for Technical Literature (KWIC Index)", em 1960, ele introduziu a idéia e o plano para um índice de rotações baseado em títulos e produzido por máquina.

Em 1962, esta sua idéia já se espalhara pelos Estados Unidos, havendo o índice KWIC sido adotado em mais de trinta instituições.

Anos mais tarde, uma necessidade de aperfeiçoamento fez com que muitos usuários do KWIC, salientassem o termo-índice, tirando-o fora do contexto, formando então um índice KWOC. Entre os usuários pioneiros, temos a biblioteca da "Douglas Missiles and Space", a

\* Terminologia utilizada por Kemp e seu grupo (ver citações bibliográficas).

"IBM - International Business Machines Corporation", a "NASA - National Space Aeronautics and Astronautics", o "Office of Technical Services", todos nos Estados Unidos.

O índice KWIC envolve uma rotação pelo computador das palavras do título e sua subsequente ordenação alfabética. O objetivo do programa de computador é rotar cada palavra "significativa" em um título, situando, uma por vez, numa mesma posição fixa no índice, "significativa" referindo-se a todas as palavras que não tenham sido incluídas na "stoplist" na memória da máquina. A "stoplist" consiste numa lista de palavras que são consideradas sem significação ou não significativas para propósitos de recuperação e indexação, as quais são excluídas do índice. Obviamente palavras não significativas dos títulos são rejeitadas juntamente com outras palavras julgadas não úteis para serem listadas numa coleção particular ou para um fim específico.

Cada palavra "significativa", mais conhecida como palavra-chave, em um título, constitui então um termo de indexação no índice KWIC, fazendo parte de uma entrada de uma linha. À medida que cada palavra é indexada, o título é mudado para a direita, uma palavra por vez, permitindo à próxima palavra-chave ocupar a posição de índice. O processo é repetido até que o título inteiro tenha sido rotado. Os termos-índice são arrumados em ordem alfabética e alinhados numa coluna da página, de modo que sua sequência possa ser facilmente observada. Havendo limite demarcado de exposição do título, exigido pelo formato, este às vezes, é truncado.

O índice KWOC difere do índice KWIC basicamente nesta disposição final. Nele, as palavras-chave são listadas alfabeticamente fora do contexto, ora como um cabeçalho, ora na margem esquerda. O título, então, vem logo abaixo do termo-índice na sua forma original e não no formato "giratório" do índice KWIC, onde cada palavra ocupa uma vez uma mesma posição (aquela que inicia a segunda coluna da página). No índice KWOC o título nunca é truncado pois nenhuma restrição é feita ao número de linhas para ele requeridas. Neste índice também, se desejado pode-se adicionar aos títulos todos os demais de

talhes bibliográficos evitando o problema da dupla-consulta exigida pelo índice KWIC (onde, necessitando-se de mais informações sobre a obra além do seu título, é preciso recorrer-se à listagem bibliográfica, sempre ligada à listagem de títulos por um código de referência, situado, geralmente, à direita da entrada). Logicamente, esta providência não é muito econômica, devendo ser seu uso previamente estudado. Em nossa amostra, por exemplo, havendo sido adicionados aos títulos cinco outros tipos de informação bibliográfica, podendo cada um deles ocupar mais de uma linha de impressão (Apêndice 2) optou-se pelas listagens separadas: uma de títulos, outra com todos os detalhes bibliográficos: autor, título, imprensa, etc.

### 1.3. CDU: o Sistema de Classificação Decimal Universal

A CDU foi elaborada em 1894, quando Paul Otlet e Henri La Fontaine, com o propósito de chegarem a uma mais precisa classificação de artigos de periódicos e descrições de patentes, iniciaram a tarefa de expandir para este fim a classificação decimal (desenvolvida em 1876 por Melvil Dewey).

A CDU é estruturada basicamente sobre dez divisões principais de assuntos, numeradas de 0 a 9, segundo o quadro exposto no apêndice 5. Dentro das divisões principais, enumeram-se as subdivisões, sempre numa sequência decimal, dentro das subdivisões, as suas ramificações, etc. Como medida para maior flexibilidade, ela apresenta dentro de cada classe - (sendo classe, segundo Veiga, um conjunto de assuntos que possuem uma qualidade ou propriedade em comum), subdivisões analíticas especiais, representadas por um hífen ou um zero à frente dos números que são subdivisões mais especializadas, próprias de cada classe, expressando conceitos que podem ser de primordial importância para certos grupos de especialistas. Outro artifício utilizado na CDU é a utilização dos "Auxiliares Comuns", feitas dos assuntos ou de suas subdivisões que podem ser comuns aos vários campos do conhecimento e que na CDU fazem-se representar por notações especiais, sinais e números que podem ser aplicados segundo a necessidade à todas as partes das tabelas principais. Estes auxiliares comuns são divididos nas seguintes tabelas:

língua, forma, lugar, raça e nacionalidade, tempo, ponto de vista , sendo também facultadas as subdivisões alfabéticas (nomes, letras) e as subdivisões de ordem numérica (não decimal), quando for necessário especificar nomes individuais ou números, devendo estes ser precedidos do número da classe da CDU.

Não visando aprofundarmo-nos nos princípios que regem a CDU, gostaríamos apenas de ressaltar o aspecto nela existente de ser uma classificação enumerativa, possuindo porém dispositivos que a aproximam das classificações analítico-sintéticas. Em outras palavras: embora haja sido estruturada para apenas listar conceitos, através da utilização do sinal de relação, os dois pontos, como um exemplo, a CDU pode também reunir blocos diversos formando assuntos complexos , que é a idéia básica das classificações analítico-sintéticas. Por esta razão, pode se unir a um índice em cadeia, pois, como veremos a seguir, foi este estruturado para indexar sistemas analítico-sintéticos.

#### 1.4. Índice em Cadeia.

O método de indexação em cadeia foi inventado há exatamente 40 anos por Ranganathan, pode-se dizer como consequência da necessidade de por ele sentida de indexar adequadamente a sua classificação dos dois pontos.

Foi, portanto, elaborado para funcionar como um índice alfabético de assuntos relativo a um sistema de classificação analítico-sintético, melhor dizendo, facetado e o próprio Ranganathan, conforme lembrou J. Mills, recomenda que ele somente seja aplicado à sistemas de classificação que tenham uma "estrutura expressiva", isto é , baseados no princípio "analítico-sintético", onde os símbolos de classificação expressem subordinação.

São dois os princípios fundamentais deste sistema: o primeiro - permitir nele a reunião dos termos correlatos dispersos (distributed relatives) no catálogo sistemático; o segundo é o de permitir o acesso a um assunto complexo em seus vários níveis de generalidade,

assegurando que um ponto de entrada será providenciado para um conceito, qualquer que seja o nível genérico no qual ele esteja sendo a bordado.

Segundo Bose, quando se faz referência ao Índice em Cadeia, duas básicas vantagens que este sistema oferece são sempre lembradas: (1) é um sistema relativamente mecânico e metódico (ordenado), onde todas as palavras-chave "procuradas" são consistentemente citadas como entradas no índice e (2) é econômico porque muitas das entradas do índice servem como canais aos assuntos que são inseridos na sequência classificada e não necessitam ser indexados.

Para aspecto econômico do índice em cadeia citamos um exemplo comparativo: - um assunto complexo, formado por tres conceitos - num índice em que se quisesse entrá-lo por estes 3 conceitos, teria **3 entradas:**

A B C

B C A

C A B

Se fosse pretendido, porém, ter acesso à combinação BAC, - por exemplo, deveríamos fazer a permutação completa, gerando então 6 entradas no índice. Se o assunto fosse formado por 4 conceitos e não 3, teríamos então que entrá-lo 24 vezes no índice. Transportando-se este problema ao âmbito de uma grande coleção, podemos imaginar o caos resultante.

A cadeia, através de seus princípios sistemáticos e rígidos, limita o número de entradas, evitando a proliferação de entradas no catálogo sistemático, permitindo ainda um acesso-múltiplo aos assuntos complexos. Ex.:

A B C

A B

A

Na Indexação em cadeia a sequência em que os conceitos são

citados é exatamente inversa àquela em que eles são dispostos no catálogo sistemático. Vários cabeçalhos vão sendo formados a partir - do destacamento em cada etapa de uma faceta do assunto complexo ou - composto, na ordem da progressão das generalidades (increasing generalness, como o denominou J. Mills). Logicamente, as entradas individuais são arquivadas alfabeticamente no índice, separadas. Deve - ser aqui lembrado, também, que uma vez tenha sido feita uma entrada no índice relativa a uma determinada notação, ela não mais necessita rã ser feita novamente. Sendo assim, à medida que o índice cresce , o número de entradas na cadeia a serem elaboradas para cada novo assunto decresce consideravelmente.

Os subcabeçalhos abertos na cadeia oriundos da ascensão na estrutura hierarquica do sistema de classificação, são adotados somente se constituírem "elos procurados" - (Segundo Veiga, aqueles - que possuem significação e que serão verbalizados pelo leitor), não sendo indexados os subcabeçalhos que constituam "elos falsos", "fundidos", e/ou "não procurados". (Sendo "elo falso" aquele que não é um número classificador, ou seja, não é uma concatenação de dígitos, não tem significação; "elo fundido" aquele que está dentro de uma - parte do símbolo de classificação, havendo sido obtido por meio de recursos auxiliares, como forma e tempo, e que, separadamente, perde o seu significado; e "elo não procurado" aquele que não será verbalizado pelo leitor), conforme os definiu Veiga e Jakobson.

### 1.5. Metodologia para Avaliação de Índices.

Quando se pensa em avaliação de índices o que primeiro ocorre a todo bibliotecário é o Projeto Cranfield, o experimento pioneiro nesta área. Ele iniciou a era dos testes de sistemas com bases realmente científicas.

De fato, um dos grandes méritos do Cranfield segundo Cleverdon, foi por fim a era das discussões infundáveis sobre valores deste ou daquela sistema de classificação ou indexação, as quais não se firmavam em nenhuma base científica, nenhum teste experimental, nenhum argumento comprovado, enfim, concretamente.

O 1º Projeto Cranfield teve início em 1957 quando C. W. Cleverdon recebeu uma verba da "National Science Foundation" para realizar um teste comparativo de vários sistemas de recuperação da informação, a ser dirigido pela ASLIB.

Neste Projeto introduziram-se como técnicas de avaliação de índices as medidas de performance, revocação (recall) e precisão, medidas estas que logo tiveram seu uso generalizado, sendo elas assim definidas:

$$Re = \frac{\text{número de documentos relevantes recuperados}}{\text{número de documentos relevantes da coleção}}$$

e

$$Pr = \frac{\text{número de documentos relevantes recuperados}}{\text{número total de documentos recuperados}}$$

Logicamente, estas medidas somente podem ser obtidas dados:

- (a) uma coleção de documentos,
- (b) um número de perguntas a serem usadas para pesquisa,
- (c) um acesso aos documentos da coleção para determinação de sua relevância para cada uma das perguntas feitas.

Vários parâmetros foram observados no Cranfield, como a experiência e consistência dos indexadores, o tempo de indexação, o assunto e forma da coleção, o fator "familiaridade com o assunto" do indexador, entre muitos outros.

As conclusões generalizáveis do Primeiro Projeto Cranfield, segundo o resumo elaborado posteriormente por Cleverdon, J. Mills e Keen, são as que se seguem:

1. Há um nível ótimo de exaustividade de indexação para uma coleção particular e dado um grupo específico de perguntas. Indexar além deste limite pouco contribuirá para aperfeiçoar o índice de revocação (recall rate) e enfraquecerá seriamente o índice de precisão.
2. Há uma relação inversa inevitável entre revocação e pre

cisão.

3. O resultado mais significativo do programa do teste principal foi que todos os quatro métodos de indexação operavam com aproximadamente o mesmo nível de performance de revocação.
4. Os fatores mais importantes a serem medidos na avaliação de sistemas de recuperação da informação são revocação e precisão.
5. A forma física do índice não surte efeito algum sobre a performance do sistema relativa à revocação e precisão.
6. Se várias linguagens de índice possuírem o mesmo nível de especificidade, serão potencialmente capazes de obter desempenho similar com respeito à revocação e precisão.
7. Quanto maior o número de dispositivos que a linguagem de índice incorpore, maiores as possibilidades para o pesquisador de alcançar mais altos níveis de performance relativos à revocação e precisão.
8. A revocação máxima depende da exaustividade da indexação, a máxima precisão depende da especificidade da linguagem de índice.

Fator não observado a tempo mas devidamente levado em conta no Segundo Projeto Cranfield foi a influência das perguntas artificiais, feitas para testar os sistemas, sobre os resultados.

Como conclusão principal do Segundo Projeto Cranfield segundo nos resumiu Lancaster em 1975, devemos citar o resultado obtido provando que uma linguagem de índice natural tem uma melhor performance do que uma linguagem controlada, podendo inclusive atingir ao mesmo tempo níveis igualmente altos de revocação e precisão se manuseada adequadamente (criando-se, por exemplo, na etapa de indexação, voca -

bulários de sinônimos, para serem utilizados tanto na etapa de indexação, como na estratégia de busca).

Um trabalho conjunto entre o grupo de Cranfield e o da Case Western Reserve University, em 1961, utilizou novamente as medidas de revocação e precisão, introduzindo porém, desta vez, o conceito do grau de relevância do documento recuperado, dividindo os resultados de performance em Relevância 2 ("tão útil como o documento fonte"), (considerando-se um "documento fonte" aquele que servira de base para elaboração de uma pergunta utilizada no teste), e Relevância 3 ("de menor relevância para a pergunta").

A Western Reserve, posteriormente, introduziu uma nova terminologia na área da "recuperação da informação", com as medidas de "sensitividade" (sensitivity) que equivale a "medida de revocação" - e especificidade (specificity) que não é o mesmo que precisão, pois

$$ES = \frac{\text{número de documentos não relevantes não recuperados}}{\text{número total de documentos não relevantes na coleção}}$$

Outras pesquisas fundamentais no campo da avaliação de índices são as desenvolvidas por Salton e seus colegas nas universidades de Harvard e Cornell. Por eles foram usadas, por exemplo, as medidas de revocação normalizada (normalized recall) e precisão normalizada numa pesquisa exaustiva, onde a coleção completa foi revocada e ordenada quanto à relevância para cada pergunta.

Lancaster, após Cranfield, voltou a utilizar as medidas para avaliação de relevância (revocação e precisão) no teste do MEDLARS (Medical Literature Analysis and Retrieval System), o sistema de informação da Biblioteca Nacional de Medicina dos Estados Unidos. Neste experimento, entretanto muitos outros parâmetros foram por ele observados, entre os quais:

1. a interação sistema-usuário. Foi reconhecida a enorme importância deste aspecto, analisando-se a apurando-se resultados obtidos na etapa da pesquisa do usuário mediante os vários tipos de interação, chegando-se, por e-

xemplo a apurar que as perguntas feitas pelo usuário ao MEDLARS com a assistência do operador do sistema recebiam valores inferiores de desempenho do que aquelas feitas pelo usuário diretamente ao sistema. Aliás, similar resultado obteve o grupo de Western Reserve em seu sistema de recuperação de informações ao analisar o aspecto "interação sistema-usuário".

2. as perguntas reais. Reconhecida a influência do tipo de pergunta sobre os resultados obtidos, foi este o primeiro estudo de um sistema onde se utilizou perguntas reais de usuários regulares do sistema. (As perguntas foram sendo colecionadas à medida que iam chegando ao MEDLARS).
3. a metodologia para calcular a medida de revocação. Introduziu-se neste teste uma maneira diferente para estimar-se a revocação, método este adequado a grandes sistemas de recuperação de informação, como era o caso do MEDLARS. A medida de revocação foi estimada sobre um sub-conjunto de documentos já reconhecidos anteriormente como relevantes para uma pergunta específica (estes documentos foram levantados em outras coleções afins ou após consulta verbal a especialistas).
4. a proporção da coleção realmente utilizada. Foi apurada a porcentagem de periódicos em função da porcentagem de citações recuperadas.
5. o efeito do tempo de resposta sobre o valor dos resultados da pesquisa.
6. o valor surpresa (serendipity) das pesquisas.
7. a "medida de novidade" (novelty ratio) das pesquisas.
8. consistência entre os indexadores.

9. a "medida de rejeição" (rejection rate) dos pedidos, etc.

Muita atenção também se dedicou aos fatores custo-benefício, análise das perguntas e estratégia de busca, etc, todos de suma importância na avaliação da performance de sistemas de recuperação da informação.

Enfim, estas medidas e estes parâmetros entre muitos outros aqui omitidos formam conjuntamente uma metodologia para a avaliação da performance de índices. Outros métodos porém existem para avaliar diversas linguagens de índices, onde não entra o fator usuário, focalizando-se somente o aspecto da relevância do sistema como representante de uma determinada coleção e esta avaliação é feita também comparativamente.

Entram aqui, então, entre muitos outros, os trabalhos de Kraft e Theodora Mills. O primeiro comparou o Índice KWIC (Keyword in-Context) com um Sistema de Classificação de Cabeçalhos de Assuntos (o índice feito para o "Index to Legal Periodicals (ILP)"). Para propósito de análise, os títulos foram agrupados em 5 tipos:

1. título contendo uma palavra igual ou em alguma forma de raiz ao cabeçalho de assunto do ILP.
2. um título que poderia ter sido indexado sob outro cabeçalho de assunto que também aparece no título.
3. um título contendo um sinônimo de seu cabeçalho de assunto do ILP.
4. um título não dos tipos 1, 2 ou 3, mas contendo palavras-chave que permitiriam a um pesquisador da área encontrá-lo de uma maneira óbvia num índice KWIC.
5. um título inexpressivo.

Os cabeçalhos do ILP foram considerados o padrão para definição daqueles documentos.

Kraft, ao nosso ver, foi bastante condescendente com o índice KWIC, pois julgou equivalentes um título e um cabeçalho de assunto, quando aquele continha apenas uma palavra semelhante ao cabeçalho de assunto. Ignorou, portanto, as demais entradas do índice KWIC (as restantes palavras do título) que poderiam ou não ser úteis ao usuário.

Julgamos também neste trabalho, muito perigoso, o enquadramento de títulos no tipo 4, pois é um critério totalmente subjetivo, como também pode ter sido o tipo 2, se uma definição exata de sinonímia não foi formulada.

Os resultados foram que 64,4% dos títulos equivaliam aos cabeçalhos de assunto (similaridade tipo 1); 5,3% pertenciam ao tipo 2; 6,4% ao tipo 3; 13,4% ao tipo 4 e apenas 10,5% ao tipo 5, isto é, representando a porcentagem de títulos totalmente inexpressivos.

Jã T. Mills definiu com mais precisão estas mesmas medidas - ao comparar o Índice KWIC ao índice feito para a "Bibliography of Agriculture". O tipo 1 de Kraft para ela equivaleu a medida 100%, o tipo 3 a medida 80%, o tipo 4 a 60% e o tipo 5 a medida 0%.

Ela não considerou, portanto o tipo 2 de Kraft. Para a sua escala de valores, no entanto atribuiu pesos

2 para os cabeçalhos principais

1 para os sub-cabeçalhos.

Justificou o seu critério de atribuir pesos mais altos aos cabeçalhos principais e pesos menores para cada sub-cabeçalho alegando a necessidade do usuário de seguir esta ordem ao pesquisar o índice. A média geral obtida foi 77,4%; entradas 100% totalizaram 21,1% da coleção; 80 a 100% = 37,0%; 60 a 80% = 22,8%; menos que 60%=19,1% da coleção.

Nosso trabalho segue, enfim, a linha destas duas últimas metodologias. Optamos pelo sistema da escala de valores 0% a 100% de Mills, abandonando porém o sistema de pesos, pois ele não se aplica -

ria ao sistema de indexação em cadeia. Abandonamos também o conceito do "sentido aproximado" (60%), após infinitas e inúteis tentativas de perseguirmos a imparcialidade e a objetividade. Modificamos outros parâmetros e introduzimos novos, o que explicaremos a seguir.

Gostaríamos aqui de agradecer a todas as pessoas que direta ou indiretamente, contribuíram para a realização deste trabalho.

Em particular, gostaríamos de agradecer ao IBBB - Instituto Brasileiro de Bibliografia e Documentação, através da Comissão Coordenadora do Curso de Pós-Graduação em Ciência da Informação, nos ter proporcionado participarmos de um curso de tão elevado nível; a Prof. Gilda Maria Braga, do IBBB, por ter possibilitado a concretização deste estudo através de seu apoio e segura orientação; ao Prof. Flávio Pereira de Sousa, do Rio Datacentro da Pontificia Universidade Católica do Rio de Janeiro, por sua colaboração, ao permitir que utilizássemos o computador do Rio Datacentro; ao Prof. F. Wilfred Lancaster, da Universidade de Illinois, pela atenção que nos dispensou possibilitando acrescentarmos ao nosso texto seus valiosos ensinamentos; ao Prof. Dr. Bert E. Boyce da Universidade de Missouri, por sua orientação no início desta pesquisa; ao Prof. Dr. Tefko Saracevic, da Case Western Reserve University, Ohio, por sua palavra amiga sempre presente; às bibliotecárias Maria Herbenia de Oliveira Praz, Glória Valadares e Suly Cambria Alves da Biblioteca do IBBB, por sua prestatividade e espírito de colaboração presentes desde o início de nosso curso.

Muito somos gratos também à CAPES-Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, por compreender a importância deste estudo, financiando-o; ao Prof. Tjerk Guus Franken, da FINEP-Financiadora de Estudos e Projetos e da FGV-Fundação Getúlio Vargas, por seu interesse e colaboração neste trabalho.

## 2. MATERIAL

Nêste estudo, utilizou-se a coleção de livros da biblioteca da FINEP - Financiadora de Estudos e Projetos. A FINEP, órgão da Secretaria do Planejamento da Presidência da República, instituído em 1967, desempenha atualmente as funções de administradora do FNDCT (Fundo Nacional de Desenvolvimento Científico e Tecnológico), do Programa de Apoio à Consultoria Nacional e do Programa de Desenvolvimento da Tecnologia Industrial, como também de outros programas de apoio financeiro a projetos e pesquisas. Como objetivo, ela financia os projetos e pesquisas que estão dentro dos planos de desenvolvimento do Governo Federal, possibilitando deste modo o desenvolvimento planejado do país.

Tendo em vista o cumprimento destes objetivos, técnicos em áreas diversas concentram-se em pesquisas para poder traçar e orientar estas diretrizes e em função de suas necessidades, que são sempre temporárias, é que se vai formando a coleção.

Uma amostra de 100 livros foi extraída através da consulta a uma lista de números aleatórios de três dígitos, elaborada para este fim no Rio Datacentro da PUC/RJ.

Esta amostra foi extraída da coleção de 806 livros, total de obras do mencionado acervo, em 27 de setembro de 1974.

Os índices nêste estudo comparados são os que foram elaborados em função desta coleção de 806 livros. Convém salientar que o índice em cadeia, no momento em que êstes dados foram coletados, já abrangia outras obras além destas 806 mencionadas, mas não se extraiu a amostra da coleção total indexada porque apenas os 806 livros constavam da listagem KWOC nesta época.

Por outro lado, a listagem KWOC abrangia também obras não indexadas ainda manualmente, inclusive alguns anuários e projetos, que não foram aqui considerados. Trabalhou-se então sobre o sub-conjunto da coleção onde ocorria a interseção entre os livros indexados pelo

KWOC e pelo índice em cadeia.

Foram retiradas do fichário de rubrica numérica de assuntos somente as fichas correspondentes aos cabeçalhos da cadeia indexadora do primeiro ao quinto nível.

Esta amostra aleatória, composta de 12,40% da coleção, representa-a com bastante exatidão. Na tabela 1 temos uma comparação entre a distribuição de frequência dos livros da amostra em função das classes gerais dos assuntos a que pertencem e a mesma distribuição de frequência na coleção total composta de 1011 livros. Em ambas as distribuições os dois assuntos de maior frequência são "Administração Pública" e "Economia", seguidos por "Engenharia/Tecnologia" e daí por uma representação pequena de vários outros assuntos.

Tabela 1. - Distribuição de frequência dos assuntos da coleção da FINEP: comparação entre o total (1011 livros) e o total da amostra (100 livros).

ASSUNTO E CLASSIFICAÇÃO	% coleção total*	% amostra**
1. Administração pública 35	27,20	30
2. Economia 33	24,04	30
3. Engenharia/Tecnologia em geral 62	6,63	8
4. Administração de Empresas 65	6,13	2
5. Fundamentos do conhecimento 00	4,95	4
6. Agricultura 63	4,55	7
7. Educação 37	3,86	5
8. Matemática 51	3,66	-
9. Indústria/Tecnologia química 66	3,17	1
10. Comércio 38	2,06	3
11. Estatística 31	1,38	-
12. Física 53	1,38	1
13. Geologia 55	0,99	3
14. Sociologia 30	0,99	-
15. Ciência política 32	0,79	1

ASSUNTO E CLASSIFICAÇÃO	% coleção total*	% amostra**
16. Indústrias (Borracha, Plástico, etc) 67	0,79	1
17. Direito 34	0,69	1
18. Ciência e Tecnologia 5/6	0,69	-
19. Química 54	0,69	-
20. Literatura 8	0,59	-
21. Construção de edifícios 69	0,49	-
22. Urbanização 71	0,49	1
23. Antropologia 57	0,40	-
24. Ciências Médicas 61	0,40	-
25. Indústrias especializadas 68	0,40	2
26. Geografia 91	0,40	-
27. Botânica 58	0,30	-
28. Divertimentos 79	0,30	-
29. Ciências aplicadas 6	0,20	-
30. Biologia 59	0,20	-
31. Arquitetura 72	0,20	-
32. Filosofia 1	0,20	-
33. Lógica - Metodologia 16	0,10	-
34. Ciências sociais 3	0,10	-
35. Assistência social 36	0,10	-
36. Filosofia - Linguística 4	0,10	-
37. Línguas Íbero-Românicas 46	0,10	-
38. Ciências puras 5	0,10	-
39. Astronomia 52	0,10	-
40. Biografia 92	0,10	-

\*\* Fonte: Pesquisa da autora na biblioteca da FINEP

\* Fonte: Memorando interno de 1-8-74 do Centro de Processamento de Dados - FINEP.

O índice KWOC foi utilizado pela biblioteca em caráter provisório, com o fim único de dar uma visão global da coleção aos usuários. A classificação das obras por assuntos era o objetivo final, já pressupondo-se que um índice por palavras-chave dos títulos não se

ria suficiente. Foi escolhido então o sistema de Classificação Decimal Universal, edição abreviada portuguesa, com extensões e correções à Classe 336 (Finanças) e à divisão geográfica do Brasil e optou-se também pelo índice em cadeia.

Para o índice KWOC utilizou-se a lista básica de palavras não-significativas (stopwords), acrescida apenas das expressões relativas a tempo: "período", "ano", etc. Esta lista foi assim elaborada visando-se manter como termos-chave o maior número possível de palavras que dentro de algum assunto do universo do conhecimento pudessem ter um significado próprio. Esta atitude foi tomada também em função da exaustividade da coleção. Desvantagens são óbvias, como o aparecimento como conceito-chave da frase "versando sobre matéria".

Foram codificados para dar entrada no índice KWOC os dados referentes a autores, títulos e sub-títulos, imprensa, cabeçalho de assunto do nível I da cadeia indexadora, dados relativos a aquisição e números de registro do livro na biblioteca.

Três listagens da amostra foram extraídas:

- a) listagem por palavras-chave dos títulos;
- b) listagem por frases-chave dos títulos;
- c) bibliografia.

(Apêndice 2).

Para o índice em cadeia, procurou-se a maior especificidade possível, (dentro de certos limites por motivos práticos e econômicos) para se definir os assuntos.

Não foram incluídos como termos-índice os conceitos indicativas de lugar. Partiu-se do princípio que seria mais útil ao usuário consultar o cabeçalho de assunto seguido da subdivisão de lugar do que ir ao nome do lugar e a partir dele procurar os assuntos. Assim,

como exemplo, o livro 1288, Apêndice 1, "Brasil 2002", possui a seguinte cadeia indexadora:

Níveis da Cadeia	Cabeçalhos	Número CDU
1	Desenvolvimento econômico - Brasil	330.19(81)
2	Desenvolvimento econômico	330.19
3	Princípios, leis e sistemas econômicos	330.1
4	Economia	33
5	Ciências Sociais	3

e não a que seria normalmente utilizada:

Níveis da Cadeia	Cabeçalhos	Número CDU
1	Brasil - Desenvolvimento econômico	330.19(81)
2	América do Sul - Desenvolvimento econômico	330.19(8)
3	Desenvolvimento econômico	330.19
4	Princípios, leis e sistemas econômicos	330.1
5	Economia	33

Consideramos, portanto as divisões de lugar da CDU elos não procurados pelo leitor. O mesmo tratamento demos às divisões de forma da CDU, que constaram dos cabeçalhos sempre após os assuntos. As divisões de ponto de vista, no entanto foram consideradas elos procurados e entradas, conseqüentemente, no índice.

É necessário observar-se, neste ponto, que estas divisões de ponto de vista tiveram uma função muito importante na elaboração deste índice em cadeia, razão pela qual foram tratadas como elos procurados. Faltando à tabela CDU utilizada termos recentes, altamente relevantes como agentes recuperadores da informação, foram estas acrescentadas junto a outros que se lhes estavam próximos na tabela de divisões auxiliares comuns de ponto de vista da CDU. Um exemplo é o do livro de nº 699, Apêndice 1, intitulado "A transferência de

tecnologia no desenvolvimento industrial do Brasil". Ele recebeu o número de classificação "67:62.001.8(81)" e a cadeia para ele elaborada foi a que se segue:

Nível	Cabeçalho	Número de Classificação
1	Transferência de tecnologia - Indústrias- Brasil	67:62.001.8(81)
2	Transferência de tecnologia - Indústrias	67:62.001.8
3	Tecnologia - Indústrias	67:62
4	Indústrias	67
5	Ciências Aplicadas	6

A divisão de ponto de vista ".001.8" equivale segundo a CDU, aos termos "Difusão"; "Generalização"; "Aplicabilidade"; "Valor"; "Importância"; "Utilidade". Sendo esses conceitos os que mais se aproximaram do conceito "Transferência", o incluímos neste grupo, possibilitando a formação do conceito-chave "Transferência de tecnologia", cuja relevância dispensa maiores comentários.

As diversas etapas dentro dos números das divisões auxiliares comuns foram consideradas e elas não procuradas e não foram indexadas. Como exemplo, da divisão auxiliar de lugar "Brasil (81)" nunca se separou a etapa "América do Sul (8)".

As divisões auxiliares comuns de língua, raça e nacionalidade e tempo não foram utilizadas nesta classificação para evitar a formação de números e cabeçalhos muito longos.

A inversão de termos dentro de um cabeçalho, que contrariaria o princípio básico da indexação em cadeia, não foi empregada em nenhuma hipótese.

O método de entrada única foi o adotado, isto é, jamais mais de um número de classificação foi atribuído a um mesmo livro e jamais um único número (embora relacionando dois ou três assuntos) gerou

mais de uma cadeia indexadora. Como exemplo citamos o caso acima ex posto de número "67:62.001.8(81)" - para ele foi feita apenas aquela cadeia e não uma para cada classe "Indústrias 67" e "Tecnologia 62". O intuito desta indexação foi economizar-se o máximo possível o número de entradas no Índice, e para tanto visou-se ater-se estritamente às regras básicas da indexação em cadeia.

Conforme preconizou Foskett, consideramos a faceta inicial aquela pela qual o documento seria armazenado e o critério para sua escolha foi o da conveniência para o usuário.

### 3. METODOLOGIA

Tendo em vista apurar-se o valor dos dois índices focalizados (KWOC e Indexação em Cadeia) como representantes da coleção anteriormente descrita, dois métodos foram elaborados e aplicados:

- 3.1. apuração do nível de similaridade entre as palavras e frases-chaves dos títulos e os cabeçalhos da cadeia indexadora e
- 3.2. apuração do nível de especificidade destes dois índices.

#### 3.1.1. Apuração do nível de similaridade entre as palavras-chave dos títulos e os cabeçalhos da cadeia indexadora.

Partiu-se do princípio que o cabeçalho de assunto de nível 1 da cadeia indexadora seria a representação-padrão da obra (considerando-se "representação-padrão" o termo indexador ou o conjunto de termos indexadores do conteúdo da obra que foram considerados "os melhores possíveis", dentro dos limites da CDU e do método de Indexação em Cadeia, por especialistas dos assuntos na FINEP - Financiadora de Estudos e Projetos) e que sendo os demais cabeçalhos (dos diversos níveis) também relacionados com o conteúdo da obra, seriam igualmente considerados.

Com o fim, então, de se avaliar o quanto o título se aproximava do significado do cabeçalho padrão (nível 1), ele foi a este comparado e valores foram atribuídos. Todas as palavras-chave do título (com exceção das duplicadas) foram comparadas a cada cabeçalho da cadeia (considerando-se palavra-chave aquela que não constasse da lista de "stopwords"), e a cada palavra em cada nível da cadeia indexadora foi atribuído um conceito de similaridade entre os cinco estabelecidos, correspondentes a uma escala caminhando de zero, setenta, oitenta, noventa, a cem por cento.

Os critérios para determinação destes valores são a seguir especificados:

100% (idênticos) - cabeçalho formado por uma única palavra igual a palavra-chave do título, ou contendo a mesma raiz (considerando-se palavras com a mesma raiz aquelas que coincidissem possuir as primeiras letras iguais e que tivessem o mesmo significado ou que fossem uma fusão do significado com um outro diferente. Ex: Agricultura X Agroindústria). Ex:

TÍTULO: Bancos de desenvolvimento - Modelo institucional.

PALAVRA-CHAVE: Bancos.

CABEÇALHO: Bancos.

90% (sinônimos) - cabeçalho formado por uma única palavra sinônima à palavra-chave do título ou à sua raiz (considerando-se sinônima, conforme Aurélio Buarque de Holanda, "a palavra ou locução que tem quase a mesma significação que outra"). Ex:

TÍTULO: Encargos trabalhistas e absorção de mão-de-obra.

PALAVRA-CHAVE: Trabalhistas.

CABEÇALHO: Emprego.

O exemplo acima também ilustra o fato da análise feita a nível de raiz da palavra.

Neste mesmo exemplo, aliás, juntamente com o cabeçalho "Emprego", foi especificado o cabeçalho "Trabalho", no mesmo nível da cadeia. Neste caso, similar a muitos outros, o título aos dois cabeçalhos foi comparado, escolhendo-se para representá-lo o maior valor apurado. Assim, entre

1 Emprego X Trabalhistas = 90% (sinônimos)

2 Trabalho X Trabalhistas = 100%(idênticos)

prevaleceu o 2º, valor 100%.

80% (outros termos sendo algum idêntico) - cabeçalho formado por uma ou mais palavras de terminologia idêntica à palavra-chave do título, entre outras de terminologia diferente da palavra-chave do título. Ex:

TÍTULO: Bibliografia brasileira de física.

PALAVRA-CHAVE: Bibliografia.

CABEÇALHO: Ciências puras - Bibliografia.

70% (outros termos sendo algum sinônimo) - cabeçalho formado por uma ou mais palavras sinônimas a palavra-chave do título e também - por outras de terminologia diferente da palavra-chave do título.

Ex:

TÍTULO: Indicadores econômicos conjunturais.

PALAVRA-CHAVE: Conjunturais.

CABEÇALHO: Situação econômica.

0% (inaceitáveis) - cabeçalho formado por palavras diferentes da palavra-chave do título. Ex:

TÍTULO: Programa educacional para clientes.

PALAVRAS-CHAVE: Programa; Educacional; Clientes.

CABEÇALHO: Computadores - Catálogos.

### 3.1.2. Apuração do nível de similaridade entre as palavras e frases-chave dos títulos e os cabeçalhos da cadeia indexadora.

Uma vez percebida a dificuldade de se comparar palavras-chave (títulos) com frases-chave (cabeçalhos), lançou-se a suposição - que se o título contivesse frases ao invés de palavras-chave poderia receber valores mais altos de similaridade e ao mesmo tempo aperfeiçoar a sua eficiência como índice (uma vez que quanto mais próximo da representação-padrão, melhor satisfaria o usuário). Note-se, como ilustração, que somente quando o cabeçalho possuía uma só palavra e esta coincidia com a palavra-chave do título é que esta última recebia o valor 100%.

Com esta hipótese em mente, reunimos o maior número possível de palavras dos títulos em frases-chave. O resultado foi o total de 99 frases-chave e 163 palavras-chave (Apêndice 4), contra o total anteriormente obtido de 316 palavras-chave (Apêndice 3). Convém observar-se neste ponto que nomes próprios geográficos sempre foram considerados uma única palavra-chave, mesmo que compostos.

Foram consideradas frases-chave aquelas que formadas pela conexão de duas palavras consecutivas do título, resultassem em um novo significado. Não se tentou formar frases-chave previamente esta-

belecidas em vocabulários ou thesauri especializados, bastando apenas que dois termos "combinassem", para serem reunidos. Logicamente, em casos de dúvidas, houve consulta a fontes apropriadas, como a tabela da CDU, a um thesaurus, a dicionários e a diversos especialistas.

O critério estabelecido para a formação de frases-chave foi o seguinte: - cada palavra foi analisada em função do seu significado naquela obra e em função das palavras que a antecederem ou se seguissem.

a) Se formasse com uma destas duas palavras uma frase-chave estabelecida, de uso generalizado, a união seria determinada com a palavra devida. Ex: Electronic Computers.

b) Se formasse com uma destas palavras uma frase-chave com algum significado, embora não de uso generalizado, e esta frase-chave não prejudicasse o potencial do índice para recuperação da informação, a união seria também realizada.

Ex: Decretos promulgados. Exemplo de união não realizada porque o segundo termo ficaria perdido, (sendo ele o mais relevante): Região agrícola.

c) Se não se enquadrasse em nenhum dos casos acima, não seria reunida em frase-chave.

Segundo este critério, obtivemos 59 frases-chave de uso generalizado, 40 frases-chave combinadas para esta coleção e 163 palavras-chave.

De posse destes novos dados, novas medidas foram calculadas, com base nos mesmos critérios expostos no método de "apuração do nível de similaridade entre as palavras-chave dos títulos e os cabeçalhos de assunto".

Como exemplo da diferença na apuração de nível de similaridade entre palavras-chave e cabeçalhos e frases-chave e cabeçalhos, vemos o livro nº 595 (Apêndice 1) abaixo:

TÍTULO: Reforma administrativa.

CABEÇALHO DO NÍVEL 3 DA CADEIA: Administração pública.

Na primeira apuração (palavras-chave X cabeçalhos) obteve-se os seguintes resultados:

"Reforma" X "Administração pública" = 0%

"Administrativa" X "Administração pública" = 80%

média = 40%

Na segunda apuração (frases-chave X cabeçalhos) os resultados modificaram-se, como se segue:

"Reforma administrativa X Administração pública" = 80%

média = 80%

### 3.2. Apuração do nível de especificidade entre as frases-chave dos títulos e dos cabeçalhos da cadeia indexadora.

Tendo novamente o cabeçalho de nível 1 da cadeia como representação padrão da obra, apurou-se o nível de especificidade, como se segue:

Frases - chave dos títulos e cabeçalhos dos vários níveis da cadeia indexadora foram analisadas, comparando-se o grau de especificidade e o número de frases especificadoras que cada um carregasse. (Considerando-se "frases especificadoras" aquelas que contribuissem para o maior detalhamento do assunto da obra).

Dados uma frase representativa de forma de apresentação física da obra e uma frase representativa de um assunto desta mesma obra, considerou-se mais específico o título ou cabeçalho que contivesse a última. Como exemplo, temos o livro nº 655 (Apêndice 1):

TÍTULO: Guia prático trabalhista.

CABEÇALHO EM NÍVEL 3 DA CADEIA: Legislação e fiscalização do trabalho.

Comparando-se a frase do título "guia prático" com a frase do cabeçalho "legislação e fiscalização", considerou-se a última de maior grau de especificidade para representar o conteúdo da obra.

Palavras não-significativas dos títulos (stopwords) não foram consideradas, pois no índice KWOC elas não estariam presentes - para a eventual coordenação de termos-chave. Como exemplo, temos o livro nº 1288 (Apêndice 1):

TÍTULO: Brasil 2002.

CABEÇALHO: Desenvolvimento econômico - Brasil.

O cabeçalho foi considerado aqui mais específico do que o título, pois naquele, o conceito de tempo, "2002", não entraria como índice, restando apenas a palavra "Brasil" que em nenhuma hipótese - pode ser mais específica do que ela mesma seguida de um outro conceito, como no caso acima.

Siglas conhecidas, quando constituíssem o tema central da obra, foram consideradas altamente relevantes como agentes especificadores. Se brasileiras, considerou-se também que já incluíam a idéia da localização "Brasil". Ex:

TÍTULO: Consolidação do FGTS.

CABEÇALHO: Legislação e fiscalização do trabalho - Brasil.

Neste caso, livro nº 112 (Apêndice 1), o título foi considerado mais específico do que o cabeçalho.

Necessário ressaltar que na designação do valor das frases - especificadoras, muitas vezes foi necessário o julgamento da autora sobre por exemplo a probabilidade de procura de um dado termo por parte do usuário. Embora baseando-se na experiência prévia e familiaridade com os usuários desta coleção, admitimos que pode haver interferido alguma subjetividade, embora uma vez conscientes do problema, tenhamos fervorosamente tentado evitá-lo.

## 4. RESULTADOS E COMENTÁRIOS

### 4.1. Nível de Similaridade

A média geral obtida considerando-se todos os níveis das cadeias indexadoras foi de 19,03% para os 100 títulos quando julgadas as suas palavras-chave, havendo esta média subido para 22,61% quando calculada posteriormente ao serem atribuídos valores às frases-chave formadas para os títulos.

Na comparação do índice KWOC original (formado por palavras-chave) com os cabeçalhos do primeiro nível das cadeias, a média obtida foi, no entanto, de 38,85% (Tabela 2). O valor da média obtida para os demais níveis foi, como se esperava, diminuindo, como se segue: segundo nível - 24,32% (Tabela 3); terceiro nível - 14,31% (Tabela 4); quarto nível - 8,03% (Tabela 5); quinto nível - 4,53% (Tabela 6).

Estes resultados não constituíram surpresa para a autora, uma vez conhecidos os princípios da Indexação em Cadeia - (partir-se do termo mais específico e ir desmembrando o cabeçalho, ascendendo-se numa estrutura hierárquica até chegar a termos gerais, por vezes exageradamente amplos para a obra indexada).

As médias dos diversos níveis da cadeia ao serem os cabeçalhos comparados às frases-chave dos títulos, no entanto, assim distribuíram-se: nível 1 - 47,41% (Tabela 7); nível 2 - 28,98% (Tabela 8); nível 3 - 17,20% (Tabela 9); nível 4 - 9,08% (Tabela 10); nível 5 - 3,76% (Tabela 11).

Tabela 2 - Distribuição de frequência dos valores de similaridade dos títulos versus cabeçalhos de nível 1 da cadeia indexadora (palavras-chave x cabeçalhos).

ORDEM DE SÉRIE	FREQUENCIA POR NÚMERO DE OBRAS	%	MÉDIA %	MÉDIA PARCIAL %
1	18		26,66	
2	12	52	80	36,92
	12		40	
3	10		0	
4	7		53,33	
5	5	16	60	50,08
6	4		32	
7	3		30	
	3		20	
	3		10	
8	2		100	
	2	21	76,66	37,99
	2		50	
	2		48	
	2		22,85	
	2		11,42	
9	1		77,5	
	1		64	
	1		46	
	1		35,55	
	1		34,28	
	1		32,85	33,33
	1	11	30,76	
	1		17,5	
	1		16	
	1		7,27	
	1		5	
MÉDIA			38,85	38,85
TOTAL	100	100		

Fonte: Pesquisa da autora na biblioteca da FINEP.

Tabela 3 - Distribuição de frequência dos valores de similaridade dos títulos versus cabeçalhos de nível 2 da cadeia indexadora (palavras-chave x cabeçalhos).

ORDEM DE SÉRIE	FREQUENCIA POR NÚMERO DE OBRAS	%	MÉDIA %	MÉDIA PARCIAL %
1	20		0	
2	15	49	13,33	15,50
3	14		40	
4	8		53,33	
5	6		20	
6	5	23	26,66	32,34
7	4		16	
8	3		11,42	
9	2		60	
	2		50	
	2	15	48	32,1
	2		25	
	2		22,85	
	2		17,77	
10	1		80	
	1		77,5	
	1		46	
	1		33,33	
	1		32	
	1		30	34,39
	1	13	28,57	
	1		24	
	1		23,33	
	1		22,5	
	1		18,46	
	1		17,5	
	1		14	
MÉDIA			24,32	24,32
TOTAL	100	100		

Fonte: Pesquisa da autora na biblioteca da FINEP.

Tabela 4 - Distribuição de frequência dos valores de similaridade dos títulos versus cabeçalhos de nível 3 da cadeia indexadora (palavras-chave x cabeçalhos).

ORDEM DE SÉRIE	FREQUÊNCIA POR NÚMERO DE OBRAS	%	MÉDIA %	MÉDIA PARCIAL %
1	39	39	0	0
2	14	14	13,33	13,33
3	7	7	26,66	
4	4	4	20	26,66
5	4	4	25	
6	3	3	33,33	
7	2	2	53,33	
8	2	2	11	26,36
9	2	2	14,28	
10	2	2	11,42	
11	1	1	62	
12	1	1	50	
13	1	1	48	
14	1	1	37,5	
15	1	1	32	
16	1	1	23,33	
17	1	1	14	26,26
18	1	1	17,5	
19	1	1	16,66	
20	1	1	14	
21	1	1	12,5	
22	1	1	12,3	
23	1	1	11,11	
24	1	1	8	
MÉDIA			14,31	14,31
TOTAL	100	100		

Fonte: Pesquisa da autora na biblioteca da FINEP.

Tabela 5 - Distribuição de frequência dos valores de similaridade dos títulos versus cabeçalhos de nível 4 da cadeia indexadora (palavras-chave x cabeçalhos).

ORDEM DE SÉRIE	FREQUÊNCIA POR NÚMERO DE OBPAS	%	MÉDIA %	MÉDIA PARCIAL %
1	56	57,73	0	0
2	13	13,40	13,33	13,33
3	5	9,27	20	22,96
4	4		26,66	
5	3		11,42	
6	2	7,21	40	23,46
	2		25	
7	1		62	
	1		32	
	1		23,33	
	1		17,5	
	1		16,66	
	1	12,37	16	19,62
	1		14,28	
	1		14	
	1		12,5	
	1		11,11	
	1		10	
	1		6,15	
MÉDIA			8,03	8,03
TOTAL	97	100		

Fonte: Pesquisa da autora na biblioteca da FINEP.

Tabela 6 - Distribuição de frequência dos valores de similaridade dos títulos versus cabeçalhos do nível 5 da cadeia indexadora (palavras-chave versus cabeçalho).

ORDEM DE SÉRIE	FREQUÊNCIA POR NÚMERO DE OBRAS	%	MÉDIA %	MÉDIA PARCIAL %
1	58	86,56	0	0
2	4	5,97	33,33	33,33
3	1		50	
	1		48	
	1	7,46	40	34,13
	1		25	
	1		7,69	
MÉDIA			4,53	4,53
TOTAL	67	100		

Fonte: Pesquisa da autora na biblioteca da FINEP.

Tabela 7 - Distribuição de frequência dos valores de similaridade das frases e palavras-chave dos títulos ver sus cabeçalho de nível 1 da cadeia indexadora

ORDEM DE SÉRIE	FREQUÊNCIA POR NÚMERO DE OBRAS	%	MÉDIA %	MÉDIA PARCIAL %
1	31	56	40	57,85
2	25		80	
3	10	16	0	19,99
4	6		53,33	
5	3		26,66	
	3		60	
	3		32	
	3	19	20	48,10
	3		16	
6	2		100	
	2		50	
7	1		48	
	1		45,71	
	1		38,33	
	1		37,5	
	1	9	35	29,78
	1		34,28	
	1		13,33	
	1		8,88	
	1		7	
MÉDIA			47,41	47,41
TOTAL	100	100		

Fonte: Pesquisa da autora na biblioteca da FINEP.

Tabela 8 - Distribuição de frequência dos valores de similaridade das frases e palavras-chave dos títulos ver sus cabeçalho do nível 2 da cadeia indexadora

ORDEM DE SÉRIE	FREQUÊNCIA POR NÚMERO DE OBRAS	%	MÉDIA %	MÉDIA PARCIAL %
1	20		40	
	20	60	0	20
	20		20	
2	7		80	
3	6	23	26,66	50,43
	6		53,33	
4	4		30	
5	3		33,33	
	3	8	50	35,24
6	2		16	
7	1		32	
	1		34,28	
	1		52	
	1		22,85	
	1	9	35	28,45
	1		17,5	
	1		16,66	
	1		23,33	
	1		22,5	
MÉDIA			28,98	28,98
TOTAL	100	100		

Fonte: Pesquisa da autora na biblioteca da FINEP.

Tabela 9 - Distribuição de freqüência dos valores de similaridade das frases e palavras-chave dos títulos versus cabeçalho do nível 3 da cadeia indexadora

ORDEM DE SÉRIE	FREQUÊNCIA POR NÚMERO DE OBRAS	%	MÉDIA %	MÉDIA PARCIAL %
1	40	40	0	0
2	23		20	
3	10	33	26,66	22,01
4	6	6	40	40
5	3		50	
6	2		80	
	2		33,33	
	2	15	25	38,46
	2		13,33	
	2		16	
	2		53,33	
7	1		45	
	1		17,5	
	1		35	
	1	6	10	27,05
	1		22,85	
	1		32	
MÉDIA			17,20	17,20
TOTAL	100	100		

Fonte: Pesquisa da autora na biblioteca da FINEP.

Tabela 10 - Distribuição de frequência dos valores de similaridade das frases e palavras-chave dos títulos versus cabeçalho do nível 4 da cadeia indexadora

ORDEM DE SÉRIE	FREQUÊNCIA POR NÚMEROS DE OBRAS	%	MÉDIA %	MÉDIA PARCIAL %
1	56	57,73	0	0
2	17	17,52	20	20
3	6	11,34	26,66	32,72
4	5		40	
5	2		16	
	2		25	
	2	10,30	80	16,6
	2		10	
	2		35	
6	1		13,33	
	1	3,09	11,42	14,08
	1		17,5	
MÉDIA			9,08	9,08
TOTAL	97	100		

Fonte: Pesquisa da autora na biblioteca da FINEP.

Tabela 11 - Distribuição de frequência dos valores de similaridade das frases e palavras-chave dos títulos versus cabeçalhos do nível 5 da cadeia indexadora

ORDEM DE SÉRIE	FREQUÊNCIA POR NÚMERO DE OBRAS	%	MÉDIA %	MÉDIA PARCIAL %
1	58	86,56	0	0
2	3	7,46	33,33	35,99
3	2		40	
4	4		86,66	
	1		10	
	1	5,97	75	49,16
	1		25	
MÉDIA			3,76	3,76
TOTAL	67	100		

Fonte: Pesquisa da autora na biblioteca da FINEP

Uma vez estabelecido ser o cabeçalho do primeiro nível, o padrão de excelência para representação do conteúdo da obra, concluiu-se ser inapropriada a análise dos resultados em termos gerais e dedicou-se maior atenção à análise das variáveis dos cabeçalhos do primeiro nível em comparação com as diversas palavras dos títulos.

A média geral de 19,03% (22,61% com frases-chave) é realmente muito baixa uma vez que os valores de similaridade variaram de 70% a 100%; ela indica o alto grau de desconcreto entre os cabeçalhos de assunto criados para a cadeia indexadora e as palavras e frases-chave dos títulos. Mesmo a média do cabeçalho nível 1, que se considerou mais relevante, cujo valor foi calculado em 38,85% (47,41% com frases-chave), ainda é baixa. Uma explicação, porém, se faz necessária: este método se restringiu a verificar o quanto o título era semelhante ao cabeçalho de assunto, (uma vez que este foi tido pelo usuário como a representação padrão do conteúdo da obra), mas não o quanto ele era pertinente como indexador da obra, independentemente do cabeçalho.

De fato, basta um rápido passar de olhos nas 199 palavras-chave que receberam o valor 0% no nível 1 (Apêndice 3), que constituíram 62,97% do total de palavras utilizadas nesta amostra, ou nas 142 frases e palavras-chave (Apêndice 4), representando 54,19% do total de frases e palavras-chave da coleção, as quais igualmente receberam o valor de similaridade 0% no primeiro nível da cadeia, para notarmos que a maior parte delas é relevante como termo de indexação da obra, como por exemplo, os nomes de países, siglas ou mesmo palavras e frases-chave de uso generalizado, como "Capital", "Mercado", "Encargos trabalhistas".

Estas palavras ou frases de valor 0%, em sua maioria, como pudemos constatar acima, estavam fortemente relacionadas com o assunto expresso no cabeçalho, mas não atingiram o nível de identidade ou sinonímia com algum termo do cabeçalho, recebendo assim o valor nulo.

Por esta razão, dividimos estas palavras-chave em três grupos distintos (Apêndice 3).

- a) palavras que receberam o valor 0% mas que eram realmente significativas no caso dado.
- b) palavras que receberam o valor 0% mas que seriam frases-chave quando coordenadas com outras palavras do mesmo título.
- c) palavras supérfluas, sem significação.

Consideramos então que somente as palavras-chave do grupo (c) poderiam ser consideradas realmente não-significativas como termo Índice das obras analisadas.

Sendo assim, obtivemos 85 palavras do grupo "a", 66 do grupo "b" e 48 do grupo "c".

Da mesma forma, dividimos as palavras e frases-chave (Apêndice 4) nestes três grupos, onde "a" resultou em 74 palavras e frases-chave, "b" resultou em 28 palavras-chave e "c" somou 40 palavras e frases-chave.

Assim, para as palavras-chave, podemos dizer que embora 199 sejam as palavras que receberam o valor de similaridade 0% em algum ou todos os títulos do nível 1 da cadeia, 48 apenas eram realmente não significativas, isto é, 24,12% do total de 316 palavras-chave dos títulos e para as frases e palavras-chave de nível 0%, que somaram 142, somente 40 eram realmente de valor nulo como termos índice, representando a parcela de 15,26% do total de 262 frases e palavras-chave do segundo índice KWOC.

Supomos por isto que se comparássemos os dois índices com a obra em si, sem apurarmos a sua similaridade nem o valor de cada um como indexador, mais justiça estaríamos fazendo aos títulos.

Veja-se, como exemplo, o caso do livro nº 194 (Apêndice 1); nele, a palavra-chave do título "trabalhistas" está fortemente relacionada ao cabeçalho "Assistência social - Mão-de-obra - Brasil".

Não chegando sequer, porém, nem ao nível de sinonímia com algumas destas palavras do cabeçalho, recebeu o conceito 0%. Como exemplo de caso em que o título foi mais específico que o cabeçalho, (o que não se pode a priori afirmar ser um melhor ou pior método), temos o que se segue: - livro nº 1140 (Apêndice 1)

TÍTULO: Oscilador a diodo tunel acoplado a uma microlinha de transmissão.

CABEÇALHO: Diodos - Osciladores.

Note-se aqui que as 4 últimas palavras-chave do título (sublinhadas), por não estarem contidas no cabeçalho, receberam o conceito 0%.

Outro fator que ocorreu para a delegação do conceito 0% a palavras do título foi a maneira adotada de se codificar para o índice KWOC conteúdos ou partes de obras quando necessário, para diferenciar títulos iguais, e ao mesmo tempo ressaltar aspectos importantes da obra. Nestes casos, que dentre os 100 títulos analisados totalizaram dois, (ver livros nº 556 e nº 1008 no Apêndice 1) o cabeçalho indexou a obra geral, pois segundo as regras de catalogação descritiva, a obra foi entrada uma só vez com indicação dos vários volumes na nota de conteúdo no corpo da ficha.

Assim, ao título, no caso do livro nº 1008, somou-se a palavra "Energia" ao livro nº 556, somou-se a palavra "Manga", as quais não poderiam constar do cabeçalho por serem próprias apenas de 2 dos vários volumes das duas coleções: -

1. "Plano de desenvolvimento econômico e social" e
2. "Contribuição ao desenvolvimento da agro-indústria".

No nível 1 (Tabela 2), notamos que a nossa amostra de 100 obras, numa ordem decrescente de número de frequência, foi dividida em 27 categorias de valores. (porcentagem das médias na coluna 4) variando de 0% a 100%. Aos 3 primeiros lugares com número de obras por valor 18, 12, 12 e 10 livros, representando 52% da coleção, couberam, respectivamente, as médias 26,66%; 80%; 40%; e 0%, o que dá a média ge

ral de 36,92%; por outro lado, do quarto ao sexto lugar, frequências de 7, 5 e 4 livros, representando 16% da coleção, obteve-se, respectivamente, as médias 53,33%; 60%; e 32%, resultando na média geral 50,08%; o terceiro grupo reuniu 21 obras com frequência de 2 a 3 livros por 9 categorias de valores, de 10% a 100%, com média geral 37,99%; o quarto e último grupo reuniu 11 obras, cada uma com um valor de similaridade diversa, variando de 5% a 77,5%, com média 33,33% para as 11 categorias. Estas 27 categorias de valores para as 100 obras é que nos deram a média geral para o cabeçalho nível 1 versus título de 38,85%.

A quantidade de categorias de valores por nível de indexação da cadeia foi diminuindo a média em que se progredia no nível (Tabelas 3, 4, 5 e 6), com respectivamente 27, 25, 19 e 7 categorias (médias diferentes obtidas).

Ao analisarmos mais minuciosamente estas tabelas, notamos a razão desta ocorrência. Na tabela 2, do nível 1, o grupo de médias de similaridade obtidas com maior frequência, reunindo 52% da coleção, dividiu-se em 4 médias diversas: 18 obras de média 26,66%; 12 obras de média 80%; 12 obras de média 40%; e 10 obras de valor 0%. Já no nível 2, tabela 3, este mesmo grupo de maior frequência, reunindo 49% da coleção, dividiu-se por 3 médias: 20 obras de valor 0; 15 obras de média 13,33% e 14 obras de valor 40%, numa média geral para este grupo de 15,50%, inferior, portanto, ao do grupo de nível 1, que foi de 36,92%. No nível 3, tabela 4, a diferença se acentua: uma média de similaridade, a de valor 0%, se destaca das demais em frequência de ocorrência, 39 obras (39%) contra 14 na segunda colocação, concentrando portanto maior número de obras na primeira ordem de frequência. O mesmo ocorre no nível 4, tabela 5, onde o número de obras com valor 0% acresce para 56 (57,73% do total), sendo este sozinho o 1º grupo de maior frequência e no nível 5, tabela 6, onde 58 (86,56%) são as obras de média 0%, restando ao segundo lugar, média 33,33% somente a frequência de 4 obras (5,97% da coleção).

Isto mostra claramente a ascensão da frequência das médias mais baixas e a maior concentração de obras de valor 0% à medida em

que se caminha para os níveis mais baixos da cadeia (2, 3, 4 e 5).

Comparando-se a tabela 2, de distribuição de frequência dos valores de similaridade entre as palavras-chave dos títulos e os cabeçalhos de nível 1 da cadeia indexadora, com a tabela 7, em que os valores são comparados aos títulos após a formação das frases-chave, constatamos que a diferença de resultados não foi muito marcante. Enquanto que na primeira tabela, as obras haviam sido divididas em 27 categorias de valores, na segunda elas reduziram-se a 20. Nesta, as duas maiores frequências, 31 e 25 obras por valor, representando 56% da coleção, tiveram a média 57,85%, e a terceira e quarta maiores frequências, 6 e 10 obras por valor, representando 16% da coleção, tiveram a média 19,99%. Continuamos assim com aproximadamente a mesma proporção (52 e 56%) da coleção aglomerada no primeiro grupo de maior frequência, mas enquanto o grupo de maior frequência na primeira medida atingia a média 36,92%, na segunda, após a formação das frases-chave, essa média aumentou para 57,85%. Mais precisamente: incluindo-se os restantes 48 e 44% da coleção para os dois grupos, obtivemos a média geral.

a) palavras-chave x cabeçalhos nível 1 = 38,85%

b) frases-chave x cabeçalhos nível 1 = 47,41%

sendo a sua diferença 8,56%, a qual se aproxima obviamente da média de diferença obtida na tabela 15, "8,88%."

As tabelas 7, 8, 9, 10 e 11, se comparadas as de 2, 3, 4, 5 e 6, nos dão o retrato do índice KWOC antes e depois da reunião das palavras em frases-chave.

A tabela 12, porém, melhor nos mostra os resultados da reforma do índice KWOC. Em ambos os índices (palavras e frases) notamos que a medida em que se descia nos níveis de indexação da cadeia, caminhando do termo mais específico para o mais geral, o número de categorias de valores para as obras diminuía, da mesma forma que o valor da média de similaridade para cada nível.

Comparando a performance dos dois índices KWOC frente ao índice em cadeia, vemos que o índice KWOC de frases-chave recebeu valores mais altos de similaridade em todos os níveis, com exceção do quinto, que teve sua média rebaixada de 4,53% para 3,76%. Este último aspecto não afeta, porém, o julgamento do índice KWOC reformado para frases por ser o quinto nível da cadeia de pouca significação.

De uma maneira geral, entretanto, a reforma aperfeiçoou os valores dos títulos, passando da média geral 19,03% para 22,61%.

Tabela 12 - Valores de similaridade dos títulos versus diversos níveis da cadeia indexadora antes e depois da formação de frases-chave

Nível da Cadeia	Número de Obras	Número de Categorias de Valores				Média %		
		Palavras	Frases	Palavras	Frases			
		Chave	Chave	Chave	Chave			
1	100	27	20	38,85	47,41			
2	100	27	19	24,32	28,98			
3	100	25	17	14,31	17,20			
4	97	19	12	8,03	9,08			
5	67	7	7	4,53	3,76			
Média Geral					19,03	22,61		

Convém salientarmos aqui que a diferença entre os valores atribuídos aos níveis 1 e 2 e mesmo entre os 3 primeiros níveis da cadeia, não é muito significativa. Basta analisarmos alguns títulos (A pên-dice 1) para nos armos a sua justificação. Em grande parte dos casos, a diferença entre os cabeçalhos de nível 1 e 2 é a extração no último do conceito indicativo de lugar e havendo cabeçalho com conceitos indicativos de lugar e forma de apresentação da obra (guias, manuais, etc), esta seria extraída no nível 2 e aquele no nível 3. Sendo assim, o cabeçalho permaneceria praticamente o mesmo nestes 3 níveis, não podendo, conseqüentemente acarretar uma diferença muito acen

tuada entre os valores médios de similaridade. Por outro lado, convém notar-se que é geralmente nos níveis 4 e 5 que os conceitos fortemente generalizados surgiam, trazendo muitos valores 0%. Estas observações se aplicam, entretanto, às obras que foram indexadas nos cinco níveis; para as de menor número de níveis, pode-se notar que o último cabeçalho em geral é o que apresenta a maior diferença de valor.

Como exemplo de obra indexada em 5 níveis, temos o livro nº 1302 do apêndice 1.

TÍTULO.	Média %
A previdência social brasileira interpretada; guia prático.	
CADEIA: 1. Previdência social-Legislação-Brasil-Guias	53,33
2. Previdência social-Legislação-Brasil	40
3. Previdência social-Legislação	26,66
4. Legislação governamental	0
5. Direito administrativo	0
Administração pública	

Na tabela 13, em que os valores de similaridade obtidos na comparação do índice KWOC original (palavras-chave) com os cabeçalhos do nível 1 da cadeia foram dispostos segundo o valor médio obtido por cada título, na ordem decrescente de valores, pode-se notar que apenas 17% dos títulos receberam os valores 70, 80, 90 e 100%, determinísticos de semelhança em algum nível entre o título e o cabeçalho; 10% dos títulos receberam o valor 0%, inaceitável tendo-se como padrão unicamente o cabeçalho de assunto nível 1; e 73% dos títulos variaram entre os limites do "inaceitável" e da "sinonímia", o que nos levou a confirmar mais uma vez que a grande maioria dos títulos, em geral, difere em terminologia dos cabeçalhos de assunto a eles atribuídos.

Tabela 13 - Valores de similaridade das palavras-chave dos títulos versus cabeçalhos do nível 1 da cadeia indexadora

Valor % Normal	Reforma do	Frequência	Número do Livro	Classe do assunto (CDU)	Número de Palavras-chave	Número de Palavras	de 0%
100	-	2	900	33	1	0	
	-		1296	62	1	0	
80	80	12	61	62	3	0	
	-		101	62	2	0	
	80		224	33	4	0	
	-		399	01	3	0	
	-		427	01	3	0	
	80		485	65	3	0	
	80		506	33	3	0	
	80		579	37	3	0	
	80		626	33	3	0	
	-		1000	65	2	0	
	-		1288	33	1	0	
	80		1457	33	3	0	
77,5	80	1	856	35	4	0	
76,66	80	2	97	37	3	0	
	80		565	38	3	0	
64	80	1	699	67	5	1	
60	53,33	5	208	33	4	1	
	60		1005	33	4	1	
	53,33		1139	33	4	1	
	80		1228	33	4	1	
	53,33		1383	01	4	1	
53,33	40	7	204	35	3	1	
	80		379	00	3	1	
	80		655	35	3	1	
	-		826	33	3	1	
	80		1130	33	3	1	
	60		1302	35	6	2	

Tabela 13 - Continuação (2)

Normal	Valor % Reforma do	Frequên cia	Número do Livro	Classe do assunto (CDU)	Número de Palavras-chave	de	Número de Palavras	de 0%
	60		1339	66	6		2	
50	80	2	31	55	2		1	
	-		979	37	3		1	
48	40	2	1008	33	5		2	
	53,33		1381	38	5		2	
46	80	1	88	35	5		2	
40	34,28	12	14	35	8		4	
	50		211	33	4		2	
	40		214	62	4		2	
	80		584	32	2		1	
	-		652	63	2		1	
	-		766	35	2		1	
	-		768	35	2		1	
	-		782	35	2		1	
	80		857	35	2		1	
	80		1023	34	2		1	
	-		1095	33	4		2	
	-		1939	33	4		2	
35,55	45,71	1	8	37	9		5	
34,28	40	1	460	63	7		4	
32,85	38,33	1	393	71	7		4	
32	40	4	67	63	10		6	
	40		1064	35	5		3	
	40		1173	63	5		3	
	53,33		1901	33	5		3	
30,76	40	1	118	33	13		8	
30	37,5	3	132	33	5		3	
	40		569	33	8		5	
	48		893	63	8		5	
26,66	40	18	80	35	3		2	
	26,66		149	62	9		6	
	32		192	63	6		4	

Tabela 13 - Continuação (3)

Valor % Normal	Reforma do	Frequên- cia	Número do Livro	Classe do assunto (CDU)	Número de Palavras-chave	Número de Palavras	0%
26,66	40	18	653	33	3	2	
	40		1140	62	6	4	
	40		925	35	6	4	
	40		850	35	6	4	
	40		916	35	6	4	
	40		945	35	6	4	
	40		1236	35	6	4	
	40		1053	35	6	4	
	40		1214	35	6	4	
	40		1231	35	6	4	
	40		1239	35	6	4	
	40		1240	35	6	4	
	40		1258	35	6	4	
	40		1269	35	6	4	
	40		1278	35	6	4	
22,85	32	2	833	35	6	4	
	32		2011	33	7	5	
20	26,66	3	4	68	7	5	
	26,66		353	38	4	3	
	-		556	63	4	3	
17,5	35	1	127	33	4	3	
16	-	1	511	33	4	3	
11,42	20	2	194	33	5	4	
	20		234	33	7	6	
10	16	3	577	62	7	6	
	16		609	62	8	7	
	13,33		879	33	8	6	
7,27	8,88	1	1443	33	8	7	
5	7	1	1076	35	11	10	
0	0	10	2	35	14	13	
	-		112	35	5	5	
0	0	10	230	33	2	2	

Tabela 13 - Continuação (4)

Valor % Normal Reforma	Frequên cia	Número do Livro	Classe do assunto (CDU)	Número de Palavras-chave	Número de Palavras 0%
0		235	37	3	3
0		500	33	4	4
-		595	35	3	3
0		986	55	2	2
0		1143	53	5	5
0		1167	55	7	7
0		1429	68	4	4
				3	3

Fonte. Pesquisa da autora na biblioteca da FINEP.

Esta proporção, 17%, difere consideravelmente daquela apresentada por Kraft, de 89,5% e da apurada por T. Mills, de 80,9%. Ambos adotaram, entretanto, critérios diferentes de julgamento dos títulos, sendo o que mais influenciou, o fato de não havermos considerado relevantes as palavras-chave de significado aproximado (mas não sinônimo), ao cabeçalho de assunto (valor 60% atribuído por Mills e tipo 4 de Kraft). O nosso critério de sinonímia foi também ao que parece muito mais rígido do que o adotado nos dois trabalhos anteriores, pois, embora eles não o tenham definido em seus trabalhos, pelos exemplos expostos, chegamos a conclusão que consideraram sinônimos dois termos com a mesma conotação no dado contexto, enquanto nós não consideramos o aspecto conotativo.

Outra diferença foi que atribuímos valor 70% a sinônimos enquanto Mills a eles atribuiu 80%. Além disto, fizemos a diferenciação 90%-70% para sinônimos e 100%-80% para termos idênticos, conforme o já explicado anteriormente na metodologia, capítulo 3.

Diferença que igualmente não deve ser ignorada é que coleções diferentes foram testadas, assim como índices diferentes, o que praticamente invalida toda e qualquer comparação.

Nesta mesma tabela 13, outros parâmetros foram observados, como as novas médias atribuídas aos títulos após a união das palavras em frases-chave, as classes de assuntos dos livros por valor recebido, o número de palavras-chave de cada título e o número de palavras que receberam valor 0% em todos os níveis.

Convém verificar-se os assuntos dos títulos analisados (coluna 5) para se concluir que eles de nada influenciaram na performance dos títulos em relação aos cabeçalhos. Em outras palavras, vê-se que as 16 classes de assunto que se apresentam em nossa amostra (Tabela 1) não oferecem particularidades próprias ou contribuem ou não para que certos títulos agrupem-se sob um determinado valor ou um determinado número de palavras-chave do título.

As colunas 6 e 7 nos auxiliam a diagnosticar as médias de similaridade atribuídas aos diversos títulos: os valores mais altos foram consequência de uma total ausência de valores nulos (0 %) atribuídos a quaisquer palavras-chave dos títulos. Entre eles estão também os pequenos títulos, de 1 a 4 palavras-chave, sendo 3 palavras a moda.

A medida em que os valores vão baixando os valores nulos vão surgindo e aumentando, e os títulos vão se alongando em média, chegando ao pico de 14 palavras-chave sendo 13 nulas, o que acarretou o valor de similaridade 5 %. A partir daí vem os valores 0 % onde ocorreram títulos de dimensão variando de 2 a 7 palavras-chave, sendo, obviamente, todas nulas.

Vê-se claramente aqui que os títulos mais longos distanciam-se mais dos cabeçalhos de assunto, pois sendo estes últimos o quanto possível sucintos, não podem chegar a conter detalhes próprios dos títulos longos.

A tabela 14 nos traz os níveis de similaridade após a formação de frases-chave em nível 1, dispostos em ordem decrescente de valores de similaridade, mostrando os mesmos parâmetros da tabela 13, da média para palavras-chave.

Na tabela 14, vê-se que os valores 70, 80, 90 e 100%, couberam à 27% da coleção, enquanto que anteriormente abrangiam a faixa dos 17%; novamente 10% dos títulos receberam o conceito 0% e os restantes 63% não atingiram os níveis de similaridade aos cabeçalhos, mas não chegaram também a serem totalmente desconsiderados, pelo mesmo fator anteriormente explicado, da relatividade do valor 0% atribuído às palavras e frases-chave.

O parâmetro "Classes de assunto" continuou com as mesmas particularidades (ou ausência delas), dispensando maiores comentários.

Tabela 14 - Valores de similaridade dos títulos após a formação de frases-chave versus cabeçalhos do nível 1 da cadeia indexadora; influência dos valores 0% sobre a média apresentada

Valor %	Frequência	Número do Livro	Classe do assunto (CDU)	Número de Palavras e Conceitos - chave	Número de Palavras e Conceitos 0%
100	2	900	33	1	0
		1296	62	1	0
80	25	61	62	2	0
		101	62	2	0
		224	33	3	0
		399	01	3	0
		427	01	3	0
		485	65	2	0
		506	33	2	0
		579	37	2	0
		626	33	2	0
		1000	65	2	0
		1288	33	1	0
		1457	33	2	0
		856	35	2	0
		97	37	2	0
565	38	2	0		
699	67	3	0		

Tabela 14 - Continuação (2)

Valor %	Frequência	Número do Livro	Classe do assunto (CDU)	Número de Palavras e Conceitos-chave	Número de Palavras e Conceitos 0%
		1228	33	3	0
		379	00	2	0
		655	35	2	0
		1130	33	2	0
		31	55	1	0
		88	35	3	0
		584	32	1	0
		857	35	1	0
		1023	34	1	0
60	3	1005	33	3	1
		1302	35	4	1
		1339	66	4	1
53,33	6	208	33	3	1
		1139	33	3	1
		1383	01	3	1
		1381	38	3	1
		1901	33	3	1
		826	33	3	1
50	2	211	33	2	1
		979	37	3	1
48	1	893	63	5	2
45,71	1	8	37	7	3
40	31	204	35	2	1
		1008	33	4	2
		214	62	2	1
		652	63	2	1
		766	35	2	1
		768	35	2	1
		782	35	2	1
		1095	33	4	2
		1939	33	4	2
		460	63	6	3

Tabela 14 - Continuação (3)

Valor %	Frequência	Número do Livro	Classe do assunto (CDU)	Número de Palavras e Conceitos-chave	Número de Palavras e Conceitos-chave	Número de Palavras e Conceitos-chave	0%
		67	63	8	4		
		1064	35	4	2		
		1173	63	4	2		
		118	33	8	4		
		569	33	6	3		
		80	35	2	1		
		653	33	2	1		
		1140	62	4	2		
		925	35	4	2		
		850	35	4	2		
		916	35	4	2		
		945	35	4	2		
40	31	1236	35	4	2		
		1053	35	4	2		
		1214	35	4	2		
		1231	35	4	2		
		1239	35	4	2		
		1240	35	4	2		
		1258	35	4	2		
		1269	35	4	2		
		1278	35	4	2		
38,33	1	393	71	6	3		
37,5	1	132	33	4	2		
35	1	127	33	2	1		
34,28	1	14	35	7	4		
32	3	192	63	5	3		
		833	35	5	3		
		2011	33	5	3		
26,66	3	149	62	6	4		
		4	68	3	2		
		353	38	3	2		
20	3	194	33	4	3		

Tabela 14 - Continuação (4)

Valor %	Frequência	Número do Livro	Classe do assunto (CDU)	Número de Palavras e Conceitos-chave	Número de Palavras e Conceitos 0%
		234	33	4	3
		556	63	4	3
16	3	511	33	5	4
		577	62	5	4
		609	62	5	4
13,33	1	879	33	6	5
8,88	1	1443	33	9	8
7	1	1076	35	10	9
0	10	2	35	4	4
		112	35	2	2
		230	33	2	2
0	10	235	37	3	3
		500	33	2	2
		595	35	2	2
		986	55	4	4
		1143	53	6	6
		1167	55	3	3
		1429	68	2	2

Fonte: Pesquisa da autora na biblioteca da FINEP.

Como resultado da tabela 13 verificamos que para 81 títulos frases-chave puderam ser formadas. (Ver lista de frases e palavras-chave resultantes no apêndice 4). Entre estes 81 títulos, (tabela 15) 58 tiveram seus valores de similaridade aumentados numa média de 13,10% ; 5 títulos tiveram seus valores diminuídos numa média de 8,07%; e os restantes 18 títulos tiveram seus valores inalterados após a formação de frases-chave. Assim, embora a maioria dos títulos com frases-chave (71,60%) haja sido beneficiada, (ainda que com a baixa média 13,10% ), com a formação destas frases, deve-se levar em conta que este número significa pouco mais da metade (58%) do total da coleção analisada, que

os demais 37% nada receberam como vantagem e que 5% dos títulos ainda sofreram desvantagem.

Um fato deve ser aqui ressaltado, entretanto: a baixa média geral (8,88%) de melhoria de valores de similaridade por si só não muito bem pode representar a significação ou não da formação destas frases-chave. Um estudo econômico da formação destas frases tem que ser levado em conta, assim como um estudo de sua relevância feito com usuários do sistema. Muitos estudos existem sobre a interação KWOC - Usuário e não é ignorada a relutância do usuário em não atribuir-lhe elevado conceito.

Tabela 15 - Diferença entre os valores de similaridade atribuídas às palavras-chave dos títulos versus cabeçalhos do nível I da cadeia indexadora e às frases e palavras-chave do título em comparação com estes mesmos cabeçalhos.

%		%		%		Títulos
Aumento		Diminuição		Inalteração		Sem conceito
6,66	3,34	-	5,72	0		101
10,16	10	-	13,33	0		112
30	6	-	6,67	0		399
8	40	-	8	0		427
13,34	6	-	6,67	0		511
34	13,34			0		556
3,34	26,67			0		595
9,24	16			0		652
17,5	9,15			0		766
7,5	2,5			0		768
5,34	40			0		782
8,58	3,33			0		826
10	18			0		900
8,58	40			0		979
6,66	8			0		1000
26,67	2			0		1095
5,48	26,67			0		1288
5,72	6,67			0		1296
						1939

Tabela 15 - Continuação (2)

% Aumento		% Diminuição		Inalteração	Títulos Sem conceito
13,34	13,34				
8	13,34				
13,34	13,34				
13,34	13,34				
13,34	20				
13,34	6,67				
13,34	6,67				
13,34	5,33				
13,34	1,61				
13,34	21,33				
13,34	9,15				
Média	13,10		8,07		
TOTAL	38		5	18	19

Fonte: Pesquisa da autora na biblioteca da FINEP.

Vickery por exemplo lembra muito bem que os dois mais graves defeitos dos índices KWIC ou KWOC são a dificuldade de coordenação dos termos na etapa da pesquisa para recuperação da informação e o número infinito de formas pelas quais um mesmo conceito pode estar indexado, devendo todas estas formas (sinônimas, quase-sinônimas, outras línguas, até mesmo erros de grafia) serem lembradas pelo usuário para que possa garantir um bom resultado para sua pesquisa.

Ora, é evidente que ao reunirmos as palavras-chave em frases de duas palavras (como foi por nós experimentado neste trabalho) ou mesmo em frases de mais palavras, estamos diminuindo a probabilidade de perda da informação pelo usuário e estamos aumentando a possibilidade de maior precisão na recuperação de informações, embora isto implique um mais baixo índice de revocação.

O apêndice 2 nos mostra as saídas dos índices KWOC com pala

avras-chave e com frases-chave, o que nos permite avaliá-los sob o aspecto físico. A listagem por palavras-chave é composta de 36 páginas com 1407 linhas impressas. Foram lidos pela máquina 679 cartões e o tempo total de execução foi de 4,60 minutos. A listagem por palavras e frases-chave, por sua vez, compõe-se de 27 páginas com 1096 linhas impressas, havendo sido lidos os mesmos 679 cartões e havendo sido gasto para sua execução o tempo de 4,23 minutos. O preço de execução da primeira foi de Cr\$ 126,02 (cento e vinte e seis cruzeiros e dois centavos) em 13 de junho de 1975 e o da segunda somou Cr\$ 98,41 (noventa e oito cruzeiros e quarenta e um centavos) em 17 de junho de 1975.

Estes últimos parâmetros nos levam a creditar méritos ao índice KWOC com formação de frases, mas um estudo do tempo gasto em formá-los e da reação a ele do usuário seriam indispensáveis para concluirmos ser ele realmente superior ao KWOC original.

Pela tabela 16, onde apresentamos a distribuição de frequência do número de palavras-chave dos títulos comparada à distribuição de frequência para cada grupo de palavras-chave das frases-chave delas originadas, com indicação da diferença de valor de similaridade dos títulos no nível 1 da cadeia indexadora após a formação de frases-chave, tentamos verificar se haveria alguma regularidade de comportamento das palavras-chave dos títulos em função das frases que formaram, o que talvez possibilitasse propormos um método mais mecânico e objetivo de formação de frases-chave. Concluimos, entretanto, que o valor e a quantidade de frases formadas independe do número de palavras-chave dos títulos. Para títulos de todos os tamanhos, indiscriminadamente, foram formadas em geral 1 ou 2 frases-chave; apenas 5 títulos (de 5, 7, 8 (2) e 9 palavras-chave) possibilitaram a formação de 3 frases-chave e apenas 2 títulos (de 8 e 14 palavras-chave) resultaram em 4 frases. Entretanto, títulos de 5, 7, 8 ou 9 palavras também formaram o padrão de 1 ou 2 frases cada.

Tabela 16 - Comparação entre as palavras-chave dos títulos e o número de frases-chave formadas com indicação da diferença entre os dois valores de similaridade obtidos no cabeçalho de nível 1 da cadeia indexadora

Ordem de Série	Frequência por Número de Obras	Palavras Chave	Frequência	Conceitos Chave Formados	Diferença no valor %
1	21	3	9	1	0
			4	-	-
			3	1	26,67
			2	1	13,34
			2	1	3,34
			1	1	-13,33
2	17	6	14	2	13,34
			2	2	6,67
			1	1	5,34
			4	4	0
			3	-	-
			2	1	- 6,67
			1	1	6,66
			1	1	6,67
			1	1	20
			2	2	2,5
			2	2	17,5
			2	2	10
3	12	2	8	-	-
3			1	40	
1			1	30	
4	11	5	2	1	0
				1	8
			1	-	-
				2	16
				1	- 8

Tabela 16 - Continuação (2)

Ordem de Série	Frequência por Número de Obras	Palavras Chave	Frequência	Conceitos Chave Formados	Diferença no valor %
				1	7,5
				3	34
				2	5,33
				2	21,33
5	7	7	2	2	9,15
			1	1	0
				1	5,72
				2	5,48
				2	8,58
				3	8,58
6	6	8	2	3	6
			1	1	- 5,72
				4	18
				2	3,33
				2	10
7	3	1	3	-	-
8	2	9	1	2	10,16
				3	0
9	1	10	1	2	8
10		11	1	2	1,61
11		13	1	5	9,24
12		14	1	4	2
MÉDIA		316		99	8,99

Fonte: Pesquisa da autora na biblioteca da FINEP.

As únicas regularidades observadas são em número pequeno e não cremos podê-las considerar altamente relevantes. Entre os títulos formados por 3 palavras-chave há uma concentração de 9 que formaram 1 frase-chave cada o que em nada alterou o valor de similaridade do título (tabela 16). São os títulos nº 61, 230, 485, 500, 506, 579, 626,

1429 e 1457 (apêndice 1). Isto se explica pela sua observação: a formação de frases-chave em todas essas vezes não alterou o valor da palavra, mas apenas reduziu o número total de palavras, o que nestes casos não provocou mudanças na média. Em outros casos a média foi diminuída, como no título nº 379 (apêndice 1).

Entre os títulos formados por 6 palavras-chave, 14 se aglomeram dentro da faixa de 2 frases-chave com diferença de valor 13,34 %. Isto se explica porque 13 dos 14 títulos pertencem a uma mesma coleção tendo os títulos iguais variado somente o conceito indicativo de lugar, o qual varia igualmente no cabeçalho, proporcionando para os 13 títulos sempre o mesmo valor de similaridade. (Ver títulos nº 850, 925, 916, 945, 1053, 1214, 1231, 1236, 1239, 1240, 1258, 1269 e 1278 no apêndice 1).

A tabela 17 nos dá uma visão global da distribuição dos valores de similaridade entre as diversas palavras dos títulos segundo a sua colocação (1º ao 14º lugar no título) e nos diversos níveis das cadeias indexadoras (1 a 5). Nota-se que em maior frequência e para qualquer nível, as palavras-chave dos títulos obtiveram em sua totalidade o valor 0%, isto é, não corresponderam aos conceitos dos cabeçalhos. Apenas a 6ª e a 12ª palavras no nível 1 equipararam a moda do valor 0% com a do valor 80% nas frequências 18 e 1 respectivamente, sendo que a 12ª palavra repetiu estas modas até o 4º nível, tendo no quinto nível equiparado ainda em primeira ordem de frequência os valores 0 e 100%.

Isto se explica facilmente pois apenas 2 títulos continham 12 palavras; (ver livros nº 118 e 1076 no apêndice 1), razões temos para supor que numa amostra maior, em que maior número de títulos longos ocorresse, também eles trariam suas últimas palavras ao nível 0% em maior frequência. Em segunda ordem de frequência temos uma constância de valor 80% para todos os níveis com ligeiras variações: os títulos mais longos (11, 12, 13 e 14 palavras) não passam de uma só ordem de frequência e de uma maneira geral pode-se dizer que da 1ª até a 6ª palavra do título o valor 80% numa segunda ordem de frequência é uma

constante.

Em terceira ordem há uma constância do valor 70%, sendo que do 3º nível para baixo ele cede lugar ao valor 100% - trata-se do reflexo da generalização nos cabeçalhos de assunto.

As conclusões que podemos tirar desta comparação são várias, como se segue:

- a) as palavras-chave dos títulos, não possuem maior peso pela colocação que ocupam no título. Vemos, como ilustração que a 1ª, a 7ª, ou a 14ª palavra-chave do título são igualmente e de um modo geral insuficientes índices, segundo a metodologia aqui aplicada; receberam todas em 1ª ordem de frequência o valor nulo em comparação com qualquer nível da cadeia.
- b) a baixa porcentagem de palavras-chave sinônimas mostram que a sinonímia pura não é fator comum entre palavras-chave do título e cabeçalhos de assunto nesta coleção. Ou os termos são iguais (80 e 100%) ou expressam o significado dos livros de forma diversa (valores 0%), ou não o trazem consigo (também expresso pelo conceito 0%).
- c) o título tende a ser específico em sua quase-totalidade, aproximando-se do cabeçalho nível 1, ou ainda sendo mais detalhado que ele. O aumento da diferença entre valores 0% e outros atribuídos às palavras à medida em que se desce nos níveis da cadeia, é um atestado do afirmado. Veja-se a 1ª palavra-chave do título: no nível 1 o valor 0% foi atribuído a 54% delas, no nível 2 a 73%; no nível 3 a 83%; no nível 4 a 89,6% (total 97 títulos); e no nível 5 a 97,01% (total 67 títulos).

Tabela 17 - Moda dos valores de similaridade das palavras-chave dos títulos com os diversos níveis da cadeia indexadora

Palavra-chave	1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	9ª	10ª	11ª	12ª	13ª	14ª
Ordem de Série														
1º	0 (54)*	0 (58)	0 (46)	0 (42)	0 (26)	0e80 (18)	0 (13)	0 (9)	0 (5)	0 (3)	0 (3)	0e80 (1)	0 (2)	0 (1)
2º	80 (41)	80 (35)	80 (38)	80 (22)	80 (20)	-	80 (5)	80 (3)	80 (1)	80 (1)	-	-	-	-
3º	100 (3)	70 (4)	70 (1)	-	70 (1)	-	70 (1)	-	-	-	-	-	-	-
4º	70 (2)	-	-	-	-	-	-	-	-	-	-	-	-	-
1º	0 (73)	0 (67)	0 (55)	0 (48)	0 (32)	0 (32)	0 (18)	0 (9)	0 (5)	0 (4)	0 (3)	0e80 (1)	0 (2)	0 (1)
2º	80 (23)	80 (27)	80 (26)	80 (16)	80 (14)	80 (4)	100 (1)	80 (3)	80 (1)	-	-	-	-	-
3º	70e100 (2)	70 (2)	100 (2)	-	70 (1)	-	-	-	-	-	-	-	-	-
4º	-	100 (1)	90e70 (1)	-	-	-	-	-	-	-	-	-	-	-
1º	0 (83)	0 (81)	0 (70)	0 (38)	0 (33)	0 (33)	0 (19)	0 (12)	0 (6)	0 (4)	0 (3)	0e80 (1)	0 (2)	0 (1)
2º	80 (10)	80 (12)	80 (12)	80 (8)	80 (13)	80 (3)	-	-	-	-	-	-	-	-
3º	100 (4)	100 (4)	100 (2)	100 (5)	70 (1)	-	-	-	-	-	-	-	-	-
4º	70 (3)	-	70 (1)	-	-	-	-	-	-	-	-	-	-	-

Tabela 17 - Continuação (2)

Palavra-chave	1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	9ª	10ª	11ª	12ª	13ª	14ª	
Ordem de Série	Nível 4	1ª	0	0	0	0	0	0	0	0	0	0	0	0	
			(87)	(87)	(77)	(58)	(31)	(33)	(18)	(12)	(6)	(4)	(3)	(1)	
			80	80e100	80	80	80	80e100	-	-	-	-	-	-	-
		2ª	(6)	(3)	(5)	(3)	(14)	(1)	-	-	-	-	-	-	-
			70e100	70	100	100	70	-	-	-	-	-	-	-	-
			(2)	(1)	(1)	(2)	(1)	-	-	-	-	-	-	-	-
	Nível 5	1ª	0	0	0	0	0	0	0	0	0	0	0e80	0	0
			(65)	(62)	(59)	(47)	(36)	(28)	(13)	(8)	(5)	(4)	(3)	(1)	(2)
			100	80e100	80	80	80	-	-	-	-	-	-	-	-
		2ª	(2)	(2)	(2)	(1)	(1)	-	-	-	-	-	-	-	-
			-	-	70e100	-	-	-	-	-	-	-	-	-	-
			-	-	(1)	-	-	-	-	-	-	-	-	-	-
3ª	-	-	-	-	-	-	-	-	-	-	-	-	-		
	-	-	-	-	-	-	-	-	-	-	-	-	-		
	-	-	-	-	-	-	-	-	-	-	-	-	-		

\* O número entre parênteses equivale ao total de títulos naquele valor, ou a frequência do valor

Fonte: Pesquisa da autora na biblioteca da FINEP.

#### 4.2. - Nível de Especificidade

Tabela 18 - Comparação entre o número de obras com maior nível de especificidade nos títulos ou nos cabeçalhos dos diversos níveis da cadeia indexadora, ou em ambos

Nível da Cadeia	CABEÇALHOS	TÍTULOS	EMPATE	TOTAL de Obras
	Mais específicos	Mais específicos		
1	45	26	29	100
2	19	72	9	100
3	8	83	9	100
4	2	92	3	97
5	-	65	2	67

Pela tabela 18 vemos o resultado da avaliação da especificidade dos índices analisados. Concluimos que de um modo geral o cabeçalho de assunto elaborado para o nível 1 da cadeia indexadora é mais específico que o título com todas as suas frases-chave. Entre 100 livros 45 deles foram melhor descritos nos referidos cabeçalhos, 26 foram pelos títulos e 29 foram igualmente bem descritos tanto pelo cabeçalho de assunto, como pelo título.

A partir do nível 2, entretanto, o título se sobressai, sendo que a partir do nível 3 a diferença é realmente grande.

Os resultados obtidos na comparação dos títulos com os níveis 2, 3, 4 e 5 da cadeia indexadora eram já esperados, pela já explicada filosofia do método de indexação em cadeia, de se criar cabeçalhos na ordem da progressão das generalidades de um dado esquema de classificação.

Surpreendeu-nos, no entanto, a comprovação da superioridade da indexação em cadeia sobre os títulos sob o aspecto da especificidade em seus cabeçalhos do nível 1, cabeçalhos estes já tidos como repre-

sentação padrão das obras da coleção.

Já é conhecida há muito a maior potencialidade de especificidade que uma linguagem de índice natural apresenta sobre uma linguagem controlada. Conforme já citamos anteriormente neste trabalho, já foi isto comprovado no 2º Projeto Cranfield. Sendo, então, os títulos portadores de uma linguagem natural, representando em sua maior parte uma síntese do conteúdo da obra feita pelos seus próprios autores, (que melhor do que ninguém devem saber o que escreveram), e os cabeçalhos de assunto aqui tratados estruturados segundo as normas rígidas do sistema de indexação em cadeia e segundo a existência e disposição dos assuntos na tabela do sistema de Classificação Decimal Universal demonstrando utilizarem uma linguagem controlada, obviamente seríamos levados a supor que tais títulos em geral fossem mais específicos do que os cabeçalhos.

Como explicação deste resultado, porém, convém lembrarmos da filosofia adotada pela Biblioteca da FINEP, para elaboração da cadeia indexadora, já explicada no capítulo 2 (material).

Termos novos, isto é, que não constavam da tabela da CDU, embora pudessem estar contidos dentro do significado de certas classes ou sub-classes, ou mesmo divisões auxiliares lá existentes, foram introduzidos dentro destas classes. Assim, utilizou-se no vocabulário da indexação os conceitos "Inovação tecnológica", "Mudança tecnológica", "Avaliação e elaboração de índices", etc., o que contribuiu para dar maior especificidade aos cabeçalhos da cadeia, aumentando consequentemente o seu potencial para maior precisão na recuperação das informações.

Outros termos que foram igualmente introduzidos no vocabulário da indexação e também contribuíam para aperfeiçoar nos cabeçalhos da cadeia o nível de especificidade foram nomes de instituições, formados mediante um agrupamento de classes e dispositivos da CDU. Assim, o livro nº 900 (Apêndice 1), por exemplo:

TÍTULO: SUDENE 10 ANOS

CABEÇALHO: SUDENE ou

Instituições oficiais para o desenvolvimento regional-Nordeste 338.984.2:061.1

Este último tipo de inovação no vocabulário convém ressaltar, aumenta por certo a especificidade do cabeçalho, mas não aumenta a potencialidade de maior precisão para a recuperação desta informação no cabeçalho, pois serão recuperados pelo código que agrupa não só obras sobre a SUDENE, como também quaisquer outras instituições oficiais para o desenvolvimento do Nordeste, enquanto que títulos que contenham a palavra-chave SUDENE, estarão certamente tratando desta instituição. Esta falha, podemos dizer, do índice em cadeia, poderia, aliás ter sido facilmente sanada, se ao número geral de assunto se houvesse acrescentado um código único para cada instituição (não se julgou, porém, este artifício necessário no caso, tendo em vista o tamanho reduzido da coleção da FINEP).

## 5. CONCLUSÕES

Podemos resumir em dois os objetivos que motivaram esta nossa pesquisa:

1. Testar esta metodologia de avaliação de sistemas.
2. Verificar qual dos dois índices analisados (KWOC e Indexação em Cadeia feito para a CDU) seria o melhor para ser utilizado na biblioteca da FINEP.

Concluimos, a respeito da metodologia, que ela é válida, mas não pode ser aplicada individualmente. Nenhuma conclusão pode ser tida como definitiva antes de levantarmos a opinião do usuário e o parâmetro custo-benefício. O conhecimento do resultado destas outras metodologias, nos permitem, inclusive, adaptarmos o sistema escolhido, mesmo que ele haja sido qualificado como o melhor, para melhor ainda servir ao usuário.

Na metodologia aqui aplicada, podemos saber qual o índice mais específico, mas não podemos avaliar, por exemplo, qual o nível de especificidade exigido em média pelos usuários.

Uma crítica, entretanto, convém se fazer aos dados aqui utilizados: comparamos um sistema de indexação a outro, tido como padrão, o que certamente provocou distorções de resultados, como as assinaladas nos apêndices 3 e 4. Mais preciso seria comparar-se os diversos índices ao livro em si, determinando para ele uma indexação livre e completa, que não tivesse que ser restrita, como em nosso caso, aos limites da CDU e da indexação em cadeia.

O Índice em cadeia aqui utilizado e reconhecido como o padrão de excelência, não pode no entanto ser considerado perfeito. Ao longo desta pesquisa, muitas pequenas falhas na classificação e indexação foram notadas, que provocaram, por certo, distorção no resultado.

Também a coleção é muito exaustiva (40 diferentes assuntos, os mais variados possíveis (Tabela 1)), o que não permite de modo algum que generalizemos os resultados obtidos.

De um modo geral, entretanto, pudemos traçar características relevantes sobre os dois sistemas avaliados, embora, conforme explicamos acima, não possamos concluir resultados definitivos.

Através da apuração dos níveis de similaridade, vimos que os títulos mereceram valores extremamente baixos; isto não é suficiente entretanto para desaboná-los como agentes indexadores, conforme os provam as listas de palavras e frases chave (Apêndices 3 e 4) que receberam o valor 0%, entre as quais tão alta proporção era realmente valorosa para indexação do documento. Vale observar-se aqui que grande parte das palavras de valor 0% que eram realmente "stopwords", permaneceram como palavras-chave devido a generalidade da lista de "stopwords", assim construída frente às características abrangentes da coleção. Este fato, por certo, prejudicou a performance do índice KWOC.

Por outro lado, se a reunião das palavras-chave em frases não melhorou em muito o valor de similaridade dos títulos perante os cabeçalhos da cadeia, lançou a idéia de se aprofundar nesta linha de pesquisa, uma vez que a listagem das frases-chave foi de preço inferior à das palavras-chave. Estudos de preferência do usuário, tempo gasto pelo indexador ao reunir conceitos, diferença entre tempo de pesquisa do usuário na primeira e segunda listagens, assim como a reunião de mais de duas palavras em frases, etc., deveriam ser levados a efeito.

Também cremos ser apropriado sugerir que estudos de outros tipos de coordenação das palavras-chave dos títulos poderiam ser realizados, o que cremos poderia não tornar as listagens mais econômicas, mas aperfeiçoaria em muito a performance do índice KWOC. Como exemplo de coordenação de palavras-chave, temos a sua completa permutação ou um método que carrega consigo as vantagens do método de permutação, sendo no entanto mais econômico: trata-se do SLIC (Selected

Listing of Indexing Combination), idealizado por J. Sharp. Este programa combina automaticamente os diversos termos-índice de todas as formas possíveis, numa rígida ordem alfabética, e o número total de entradas no índice nesta fase é representado pela fórmula  $2^n - 1$  (onde  $n$  é o total de termos de um cabeçalho de assuntos, ou num título, o total de suas palavras-chave). Assim, um título composto de 4 palavras-chave teria o total de 15 entradas no índice combinado, assim distribuídas (termos A, B, C e D):

ABCD	AB	A
ABC	AC	B
ABD	AD	C
ACD	BC	D
BCD	BD	
	CD	

Após esta primeira distribuição das palavras-chave, o SLIC e elimina as entradas desnecessárias do índice, e o número total final de entradas no índice é então representado pela fórmula  $2^{n-1}$  (onde  $n$  é novamente o número total de palavras-chave). Assim, para os 4 termos: A, B, C, D teremos no índice o total de 8 entradas, resultantes da eliminação de 7 das 15 combinações anteriormente elaboradas. Assim:

ABCD	<del>AB</del>	<del>A</del>	
<del>ABC</del>	<del>AC</del>	<del>B</del>	
ABD	AD	<del>C</del>	(15)
ACD	<del>BC</del>	D	
BCD	BD		
	CD		

resulta em

ABCD	AD	
ABD	BD	(8)
ACD	CD	
BCD	D	

Sugerimos ainda que a metodologia aqui utilizada seja aplicada em outras áreas do conhecimento, pois nesta coleção, em que mais pesaram a Economia e a Administração, ficou provado que as palavras do título pela sua colocação não coincidem em valor como índice, que

os assuntos não apresentam particularidades nas terminologias empregadas nos títulos, sendo umas melhor definidoras do que outras, que os títulos não apresentam uma constância de tamanho (dentro de cada ciência, medidos pelo número de palavras-chave) e que a variação destes tamanhos não é significativa. Serão estas conclusões válidas para outras ciências?

Quanto à apuração dos níveis de especificidade dos dois índices, constatamos, com surpresa, que um índice em cadeia (como o aqui analisado) pode ser mais específico do que um índice KWOC, com o que contrariamos conclusões anteriormente estabelecidas.

Ressaltamos, porém, que além de aqui estarmos lidando com títulos de livros (reconhecidamente mais gerais do que títulos de artigos de periódicos ou relatórios), testamos neste estudo um índice em cadeia elaborado sobre uma linguagem de índice híbrida, isto é, em que se misturou os princípios das linguagens controlada e natural, a qual permitiu descer este índice a maiores níveis de especificidade.

Isto nos leva a sugerir que qualquer linguagem de índice pode ser manuseada (em maior ou menor grau, obviamente), para servir às necessidades de uma dada coleção frente aos requisitos de seus usuários. Aliás, como bem lembrou Rubin, "muitos esforços tem sido feitos no sentido de mudar a terminologia dos sistemas de classificação" e cita Foskett... "enquanto o uso da terminologia tradicional persistir, ... isto limita o pensamento à categorias tradicionais". Aliás, convém ressaltar, o estudo da linguagem e suas propriedades de relações, categorias e sistemas sobre os esquemas de classificação, índices e thesauri, deveria ser realmente considerado por todo bibliotecário ou cientista da informação que se lançasse ao campo de representação dos documentos.

## 6. CITAÇÕES BIBLIOGRÁFICAS

- BLOOMFIELD, Masse. Evaluation of indexing; 5. Discussion and summary. Special Libraries, 62 (2):94-9, Feb. 1971.
- BOSE, H. Chain indexing and Universal Decimal Classification. Indian Librarian, 22 (4):235-9, Mar. 1968.
- CLEVERDON, Cyril; MILLS, Jack; KEEN, Michael. ASLIB Cranfield research project; factors determining the performance of indexing systems; volume 1; Design. Cranfield, NSF, 1966. part. 1, p.1-120.
- FEINBERG, Hilda. Title derivative indexing techniques; a comparative study. Metuchen, N.J., Scarecrow, 1973. 297 p.
- FERREIRA, Aurélio Buarque de Holanda. Novo dicionário da língua portuguesa. Rio de Janeiro, Nova Fronteira, 1975. 1517 p.
- FID. Classificação decimal universal; edição abreviada portuguesa. Lisboa, Instituto de Alta Cultura, 1961. 214 p.
- FISCHER, Marguerite. The KWIC index concept; a retrospective view. American Documentation, 17 (2):57-70, Apr. 1966.
- FOSKETT, D. J. Serviço de informação em bibliotecas. São Paulo, Polígono, 1969. 159 p.
- KEMP, D. A.; SIMPSON, I. S.; WILSON, T. D. Indexing; permuted, rotated, or cycled. Journal of Documentation, 28 (1):67-68, Mar. 1972.
- KRAFT, Donald 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.
- LANCASTER, F. Wilfred. Comunicação informal, 1975.

- LANCASTER, F. Wilfred. Information retrieval systems; characteristics, testing, and evaluation. New York, Wiley, 1968. 222p.
- LUHN, Hans Peter. Keyword-in-Context index for technical literature (KWIC index). American Documentation, 11(4):288-95, Oct. 1960.
- MILLS, J. Chain indexing and the classified catalogue. Library Association Record, 57 (5):141-8, Apr. 1955.
- MILLS, Theodora. Testing a Keyword-in-Context (KWIC) index on certain materials in the Bibliography of Agriculture. Washington, DC., 1965. 94 p. (A dissertation submitted to the Faculty of the Graduate School of Arts and Sciences of the Catholic University of America in partial fulfillment of the requirements for the Degree of Master of Arts in Library Science).
- MOSS, R. Minimum vocabularies in information indexing. Journal of Documentation, 23 (3):179-95, Sept. 1967.
- NOVO Michaelis dicionário ilustrado. 14 ed. São Paulo, Melhoramentos, 1973. v.1-2, português-inglês, v.3-4, inglês-português.
- OCDE. Macrothesaurus OCDE; lista dos principais descritores referentes ao desenvolvimento econômico e social. Ed. brasileira. Rio de Janeiro, IBBD, 1974. 427 p.
- RICHMOND, Phyllis A. Document description and representation. Annual Review of Information Science and Technology, 7:73-102, 1972.
- RUBIN, Judith G. Language, classification and indexing. Special Libraries, 62 (4):189-94, Apr. 1971.
- SARACEVIC, Tefko. Selected results from an inquiry into testing of information retrieval systems. Journal of the American Society for Information Science, 22 (2) Mar./Apr. 1971.

SMITH, Patricia P. Increasing recall of keyword indexes by automatically indexing root character strings. In: AMERICAN SOCIETY FOR INFORMATION SCIENCE. Annual meeting, 33, Philadelphia, 1970. The information conscious society; proceedings. Philadelphia, ASIS, 1970. v.7, p.275-7.

VEIGA, Evangelina de Azevedo & JAKOBSON, Sara Roitman. Índice em cadeia. Ciência da Informação, 3(1):69-78, 1974.

VICKERY, B. C. Techniques of information retrieval. London, Butterworths, 1970. 262 p.



## TÍTULO: Prejulgados na íntegra e súmulas do TST e STF

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Legislação e fiscalização do trabalho - Brasil	351.83(81)
2. Legislação e fiscalização do trabalho	351.83
3. Legislação governamental	351
4. Direito administrativo Administração pública	35
5. Ciências sociais	3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	% Média
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
Media %	0	0	0	0	0	0

## Frases-chave

Níveis da cadeia	1	2	3	4	% Média
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
Media %	0	0	0	0	0

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Termos do título "Prejulgados na íntegra" e "Súmulas" praticamente nulos. Termos "TST" e "STF" pouco procuráveis. Cabeçalhos dos níveis 3 e 4 tão vagos como "TST" e "STF". Quinto cabeçalho mais vago do que o título.
2		X	
3	X	X	
4	X	X	
5	X		

TÍTULO: Electronic computers; gaps in technology

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Política científica - Bens de capital - Computadores	681.142:338.91:001.89
2. Bens de capital - Computadores	681.142:338.91
3. Produção econômica - Computadores Situação econômica - Computadores	681.142:338
4. Computadores	681.142
5. Indústrias especializadas	68

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	80	0	0	20
2	0	80	0	0	20
3	0	80	0	0	20
4	0	100	0	0	25
5	0	0	0	0	0
Media %	0	68	0	0	17

## Frases-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	0	0	26,66
2	80	0	0	26,66
3	80	0	0	26,66
4	80	0	0	26,66
5	0	0	0	0
Media %	64	0	0	21,32

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	
2	X	X	
3	X	X	
4	X		
5	X		

TÍTULO: Alternative educational futures in the United States and Europe methods, issues and policy relevance

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Política educacional - Europa e Estados Unidos	379.3(4:73)
2. Política educacional	379.3
3. Escola e as autoridades públicas	379
4. Educação Ensino	37
5. Ciências sociais	3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	9	% Média
1	0	80	0	80	80	0	0	80	0	35,55
2	0	80	0	0	0	0	0	80	0	17,77
3	0	0	0	0	0	0	0	0	0	0
4	0	100	0	0	0	0	0	0	0	11,11
5	0	0	0	0	0	0	0	0	0	0
Media %	0	52	0	16	16	0	0	32	0	12,88

## Frases-chave

Níveis da cadeia	1	2	3	4	5	6	7	% Média
1	0	80	80	80	0	0	80	45,71
2	0	80	0	0	0	0	80	22,85
3	0	0	0	0	0	0	0	0
4	0	80	0	0	0	0	0	11,42
5	0	0	0	0	0	0	0	0
Media %	0	48	16	16	0	0	32	15,99

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		
5	X		
			1) Conceitos significativos ausentes do cabeçalho: "futures", "methods" e "issues"

TÍTULO: A empresa pública no direito brasileiro; princípios constitucionais, definição legal

ÍNDICE EM CADEIA CAPEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Empresas públicas - Direito administrativo - Brasil	351(81)
2. Empresas públicas - Direito administrativo	351
Empresas públicas - Administração	
Serviços públicos - Direito administrativo	
Serviços públicos - Administração	
Legislação governamental	
3. Administração pública	35
Direito administrativo	
4. Ciências sociais	3

SIMILARIDADE  
Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	% Média
1	80	80	80	80	0	0	0	0	40
2	80	80	80	0	0	0	0	80	40
3	0	80	80	0	0	0	0	0	20
4	0	0	0	0	0	0	0	0	0
Média %	40	60	60	20	0	0	0	20	25

Frases-chave

Níveis da cadeia	1	2	3	4	5	6	7	% Média
1	80	80	80	0	0	0	0	34,28
2	80	80	0	0	0	0	80	34,28
3	80	80	0	0	0	0	0	22,85
4	0	0	0	0	0	0	0	0
Média %	60	60	20	0	0	0	20	22,85

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	No primeiro nível o cabeçalho introduz o conceito "direito administrativo" e o título especifica "princípios constitucionais", definição legal.
2	X		
3	X		
4	X		

TÍTULO: Oceanografia física

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1.	Oceanografia	551.46
2.	Geografia física	551.4
3.	Geologia	55
4.	Ciências Puras	5

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	% Média
1	100	0	50
2	0	80	40
3	0	0	0
4	0	0	0
Media %	25	20	22,5

## Frases-chave

Níveis da cadeia	1	% Média
1	80	80
2	80	80
3	0	0
4	0	0
Media %	40	40

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		

TÍTULO: The economics of technological change

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Economia - Mudança tecnológica	62.001.7:33
2. Mudança tecnológica	62.001.7
3. Tecnologia Engenharia	62
4. Ciências aplicadas	6

SIMILARIDADE

Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	80	80
2	0	80	80	53,33
3	0	100	0	33,33
4	0	0	0	0
Media %	20	65	40	41,66

Frases-chave

Níveis da cadeia				%
	1	2		Média
1	80	80		80
2	0	100		50
3	0	0		0
4	0	0		0
Media %	20	45		32,5

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		

TÍTULO: Estudo do sistema radicular do cacauero em alguns tipos de solos da região cacauera do sul da Bahia.

ÍNDICE EM CADEIA CABEÇALHO	NÚMERO DO ASSUNTO (CDU)
1. Cacau - Solos - Bahia	631.4:633.74(814.2)
2. Cacau - Solos	631.4:633.74
3. Culturas arvenses - Solos	631.4:633
4. Solos Ciência do solo	631.4
5. Agricultura Agronomia	631

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	9	10	% Média
1	0	0	0	80	0	80	0	80	0	80	32
2	0	0	0	80	0	80	0	80	0	0	24
3	0	0	0	0	0	80	0	0	0	0	0
4	0	0	0	0	0	100	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0
Media %	0	0	0	32	0	68	0	32	0	16	14,8

## Frases-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	% Média
1	0	0	80	80	0	80	0	80	40
2	0	0	80	80	0	80	0	0	30
3	0	0	0	80	0	0	0	0	10
4	0	0	0	80	0	0	0	0	10
5	0	0	0	0	0	0	0	0	0
Media %	0	0	32	64	0	32	0	16	18

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		1) Conceitos significativos ausentes do cabeçalho: "sistema radicular", "tipos de solos", "região cacauera do sul...".
2	X		
3	X		
4	X		
5	X		

TÍTULO: O município na organização nacional

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1.	Administração municipal-Brasil	352(81)
2.	Administração municipal Administração local	352
3.	Administração pública Direito administrativo	35
4.	Ciências Sociais	3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	% Média
1	80	0	0	26,66
2	80	0	0	26,66
3	0	0	0	0
4	0	0	0	0
Media %	40	0	0	13,33

## Frases-chave

Níveis da cadeia	1	2	% Média
1	80	0	40
2	80	0	40
3	0	0	0
4	0	0	0
Media %	40	0	20

## ESPECIFICIDADE

Níveis da cadeia	KWCC	CADEIA	OBSERVAÇÕES
1		X	No segundo nível a imprecisão de locali- zação se estende aos dois índices.
2	X	X	
3	X		
4	X		

TÍTULO: Estudo dos orçamentos das unidades familiares do Distrito Federal

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Direitos - Funcionários públicos - Brasil	35.08:342.7(81)
2. Direitos - Funcionários públicos Garantias- Funcionários públicos	35.08:342.7
3. Direito constitucional - Funcionários públicos	35.08:342
4. Funcionários públicos - Direito administrativo Funcionários públicos - Administração	35.08
5. Administração pública Direito administrativo	35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia						%
	1	2	3	4	5	Média
1	70	80	80	0	0	46
2	70	80	80	0	0	46
3	70	80	80	80	0	62
4	70	80	80	0	80	62
5	0	80	80	0	80	48
Média %	56	80	80	16	32	52,8

## Frases-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	80	80
2	80	80	80	80
3	80	80	80	80
4	80	80	80	80
5	80	80	100	86,66
Média %	80	80	84	81,33

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		Título com duas especificações de assunto não presentes no cabeçalho. Este, no nível 1, com um conceito apenas ausente do título: "Brasil".
2	X		
3	X		
4	X		
5	X		

TÍTULO: Technical education in the United Kingdom

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Ensino técnico superior-Grã Bretanha	378.962(410)
2. Ensino técnico superior Institutos superiores técnicos	378.962
3. Faculdades	378.9
4. Ensino superior	378
5. Educação Ensino	37

## SIMILARIDADE

Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	0	80	76,66
2	80	70	0	50
3	0	0	0	0
4	0	70	0	23,33
5	0	100	0	33,33
Media %	32	62	16	36,66

Frases-chave

Níveis da cadeia			%
	1	2	Média
1	80	80	80
2	80	0	40
3	0	0	0
4	70	0	35
5	80	0	40
Media %	62	16	39

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	
2	X	X	
3	X		
4	X		
5	X		

TÍTULO: L'Europe technologique

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Tecnologia - Europa	62.(4)
2. Engenharia Tecnologia	62
3. Ciências aplicadas	6

SIMILARIDADE

Palavras-chave

Níveis da cadeia	%		Média
	1	2	
1	80	80	80
2	0	100	50
3	0	0	0
Media %	26,66	60	43,33

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		

TÍTULO: Consolidação do FGTS

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Legislação e fiscalização do trabalho-Brasil	351.83(81)
2. Legislação e fiscalização do trabalho	351.83
3. Legislação governamental	351
4. Direito administrativo Administração pública	35
5. Ciências sociais	3

## SIMILARIDADE

Palavras-chave

Níveis da cadeia			%
	1	2	Média
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
Media %	0	0	0

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		A sigla "FGTS" foi consi
2	X		derada de maior especifi
3	X		cidade que todos os ca
4	X		beçalhos da cadeia
5	X		

TÍTULO: Analytical report; comparisons between member countries in education, research & development, technological innovation, international economic exchanges

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Pesquisa e desenvolvimento Tecnologia-Educação- Situação econômica	338:37.62:001.89
2. Tecnologia-Educação Situação econômica	338:37:62
3. Educação - Situação econômica	338:37
4. Situação econômica	338
5. Economia	33

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	9	10	11	12	13	% Média
1	0	0	0	0	0	80	80	80	80	0	0	80	0	30,76
2	0	0	0	0	0	80	0	0	80	0	0	80	0	18,46
3	0	0	0	0	0	80	0	0	0	0	0	80	0	12,30
4	0	0	0	0	0	0	0	0	0	0	0	80	0	6,15
5	0	0	0	0	0	0	0	0	0	0	0	100	0	7,69
Média %	0	0	0	0	0	48	16	16	32	0	0	84	0	15,07

## Frases-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	% Média
1	0	0	0	80	80	80	0	80	40
2	0	0	0	80	0	80	0	80	30
3	0	0	0	80	0	0	0	80	20
4	0	0	0	0	0	0	0	80	10
5	0	0	0	0	0	0	0	80	10
Média %	0	0	0	48	16	32	0	80	22

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		Titulo contem varios termos especificado- res ausentes do cabe- çalho
2	X		
3	X		
4	X		
5	X		

TÍTULO: O programa de ação e as reformas de base

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Planos de desenvolvimento econômico nacionais-Brasil	338.984.3(81)
2. Planos de desenvolvimento econômico nacionais	338.984.3
3. Planos de desenvolvimento econômico	338.984
4. Planificação econômica e controle	338.98
5. Fatores da produção	338.9

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	70	0	0	0	17,5
2	70	0	0	0	17,5
3	70	0	0	0	17,5
4	70	0	0	0	17,5
5	0	0	0	0	0
Media %	56	0	0	0	14

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	70	0	35
2	70	0	35
3	70	0	35
4	70	0	35
5	0	0	0
Media %	56	0	28

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	●BSERVAÇÕES
1		X	Título de signi-
2		X	ficado muito va
3		X	go.
4		X	Cabeçalho do ní
5	X		vel 5 mais vago
			que o título

TÍTULO: Brazil country background paper; programming period  
1973-76

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Planos de desenvolvimento econômico nacionais Brasil		338.984.3(81)
2. Planos de desenvolvimento econômico nacionais		338.984.3
3. Planos de desenvolvimento econômico		338.984
4. Planificação econômica e controle		338.98
5. Fatores da produção		338.9

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia						%
	1	2	3	4	5	Média
1	80	0	0	0	70	30
2	0	0	0	0	70	14
3	0	0	0	0	70	14
4	0	0	0	0	70	14
5	0	0	0	0	0	0
Media %	16	0	0	0	56	14,4

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	80	0	0	70	37,5
2	0	0	0	70	17,5
3	0	0	0	70	17,5
4	0	0	0	70	17,5
5	0	0	0	0	0
Media %	16	0	0	56	18

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Níveis 3 e 4 "Brazil programming" equiparado aos cabeçalhos por não conterem estes o conceito indicativo de lugar. Cabeçalho do nível 5 muito vago para a obra
2		X	
3	X	X	
4	X	X	
5	X		

TÍTULO: Terceiro programa nacional de cooperação técnica da República Federativa do Brasil; ano fiscal 1971-1972 para a OEA

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Cooperação tecnológica - Brasil e Américas		62.009.01(81:7/8)
2. Cooperação tecnológica		62.009.01
3. Tecnologia Engenharia		62
4. Ciências aplicadas		6

SIMILARIDADE  
Palavras-chave

Níveis da cadeia										%
	1	2	3	4	5	6	7	8	9	Média
1	0	0	80	80	0	0	80	0	0	26,66
2	0	0	80	80	0	0	0	0	0	17,77
3	0	0	0	100	0	0	0	0	0	11,11
4	0	0	0	0	0	0	0	0	0	0
Media %	0	0	40	65	0	0	20	0	0	13,88

Frases-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	80	0	80	0	0	26,66
2	0	100	0	0	0	0	16,66
3	0	80	0	0	0	0	13,33
4	0	0	0	0	0	0	0
Media %	0	65	0	20	0	0	14,16

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		1) "OEA" mais es- pecífico que "Américas"
2	X		
3	X		
4	X		

TÍTULO: Estudos sobre uma região agrícola; zona da mata de Minas Gerais

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Pesquisa e desenvolvimento - Agricultura - Minas Gerais	631:001.89(815.1)
2. Pesquisa e desenvolvimento - Agricultura	631:001.89
3. Agricultura Agronomia	631
4. Ciências aplicadas	6

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	80	0	0	80	26,66
2	0	0	80	0	0	0	13,33
3	0	0	100	0	0	0	16,66
4	0	0	0	0	0	0	0
Media %	0	0	65	0	0	20	14,16

## Frases-chave

Níveis da cadeia							%
	1	2	3	4	5		Média
1	0	0	80	0	80		32
2	0	0	80	0	0		16
3	0	0	100	0	0		20
4	0	0	0	0	0		0
Media %	0	0	65	0	20		17

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		No título há maior especificação de lugar "zona da Mata"
2	X		
3	X		
4	X		

TÍTULO: Encargos trabalhistas e absorção de mão-de-obra uma interpretação do problema e seu debate

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Assistência social - Mão-de-obra - Brasil		331.024:362(81)
2. Assistência social - Mão-de-obra		331.024:362
3. Mão-de-obra		331.024
4. Trabalho Emprego		331
5. Economia		33

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia								%
	1	2	3	4	5	6	7	Média
1	0	0	0	80	0	0	0	11,42
2	0	0	0	80	0	0	0	11,42
3	0	0	0	100	0	0	0	14,28
4	0	100	0	0	0	0	0	14,28
5	0	0	0	0	0	0	0	0
Media %	0	20	0	52	0	0	0	10,28

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	80	0	0	20
2	0	80	0	0	20
3	0	80	0	0	20
4	80	0	0	0	20
5	0	0	0	0	0
Media	16	48	0	0	16

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		No cabeçalho do nível 1, o termo "assistência social" é mais geral que o termo do título "encargos trabalhistas". Restam então dois conceitos-extra ao título "encargos trabalhistas" e "absorção" contra um conceito-extra do cabeçalho "Brasil", o que torna o título mais específico
2	X		
3	X		
4	X		
5	X		

## TÍTULO: Fundações e empresas públicas

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Empresas públicas - Direito administrativo Legislação governamental Serviços públicos - Direito administrativo		351
2. Direito administrativo Administração pública		35
3. Ciências sociais		3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	0	80	80	53,33
2	0	0	80	26,66
3	0	0	0	0
Média %	0	26,66	53,33	26,66

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	0	80	40
2	0	80	40
3	0	0	0
Média %	0	53,33	26,66

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		Termo "funda- ções" ausente do cabeçalho
2	X		
3	X		

TÍTULO: Organização administrativa para o desenvolvimento econômico

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Administração pública-Desenvolvimento econômico Direito administrativo-Desenvolvimento econômico	330.19:35
2. Desenvolvimento econômico	330.19
3. Economia	33
4. Ciências Sociais	3

SIMILARIDADE  
Palavras-chave

Níveis da cadeia	1	2	3	4	% Média
1	0	80	80	80	60
2	0	0	80	80	40
3	0	0	0	100	25
4	0	0	0	0	0
Media %	0	20	40	65	31,25

Frases-chave

Níveis da cadeia	1	2	3	% Média
1	0	80	80	53,33
2	0	0	100	33,33
3	0	0	80	26,66
4	0	0	0	0
Media %	0	20	65	28,33

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		

TÍTULO: Bancos de desenvolvimento: modelo institucional

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Bancos de desenvolvimento	336,711
2. Bancos	336,71
3. Economia	33
4. Ciências Sociais	3

SIMILARIDADE  
Palavras-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	80	80	0	0	40
2	100	0	0	0	25
3	0	0	0	0	0
4	0	0	0	0	0
Media %	45	20	0	0	16,25

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	100	0	50
2	80	0	40
3	0	0	0
4	0	0	0
Media %	45	0	22,5

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		

TÍTULO: The conditions for success in technological innovation

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Política científica - Industrias e inovação tecnológica	62.001.7+67:001.89
2. Indústrias e inovação tecnológica	62.001.7+67
3. Inovação tecnológica	62.001.7
4. Tecnologia Engenharia	62
5. Ciências sociais	6

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	% Média
1	0	0	80	80	40
2	0	0	80	80	40
3	0	0	80	80	40
4	0	0	100	0	25
5	0	0	0	0	0
Media %	0	0	68	48	29

## Frases-chave

Níveis da cadeia	1	2	% Média
1	0	80	40
2	0	80	40
3	0	100	50
4	0	80	40
5	0	0	0
Media %	0	68	34

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	
2		X	
3	X	X	
4	X		
5	X		

TÍTULO: Manual de projetos de desenvolvimento econômico

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Projetos - Desenvolvimento econômico - Manuais		330.19:65.012.2(021)
2. Projetos - Desenvolvimento econômico		330.19:65.012.2
3. Métodos de administração - Desenvolvimento econômico		330.19:65.012
4. Administração - Desenvolvimento econômico		330.19:65
5. Desenvolvimento econômico		330.19

SIMILARIDADE

Palavras-chave

Níveis da cadeia	1	2	3	4	% Média
1	80	80	80	80	80
2	0	80	80	80	60
3	0	0	80	80	40
4	0	0	80	80	40
5	0	0	80	80	40
Media %	16	32	80	80	52

Frases-chave

Níveis da cadeia	1	2	3	% Média
1	80	80	80	80
2	0	80	80	53,33
3	0	0	80	26,66
4	0	0	80	26,66
5	0	0	100	33,33
Media %	16	32	84	27,99

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		
5	X		

TÍTULO: Metas e bases para a ação 1972 - 1974.

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Planos de desenvolvimento econômico regionais Pará	338.984.2(811.5)
2. Planos de desenvolvimento econômico regionais	338.984.2
3. Planos de desenvolvimento econômico	338.984
4. Planificação econômica e controle	338.98
5. Fatores da produção	338.9

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
Media %	0	0	0	0

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
Media %	0	0	0

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Título inexpressivo.
2		X	Cabeçalho do 5º nível
3		X	tão vago quanto o título.
4		X	
5	X	X	

TÍTULO: Plano de aplicação dos fundos de participação dos estados e especial (exercício de 1970)

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Planos de desenvolvimento econômico regionalis Pernambuco	338.984.2	(813.4)
2. Planos de desenvolvimento econômico regionalis	338.984.2	
3. Planos de desenvolvimento econômico	338.984	
4. Planificação econômica e controle	338.98	
5. Fatores da produção	338.9	

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia								%
	1	2	3	4	5	6	7	Média
1	80	0	0	0	0	0	0	11,42
2	80	0	0	0	0	0	0	11,42
3	80	0	0	0	0	0	0	11,42
4	80	0	0	0	0	0	0	11,42
5	0	0	0	0	0	0	0	0
Media %	64	0	0	0	0	0	0	9,13

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	80	0	0	0	20
2	80	0	0	0	20
3	80	0	0	0	20
4	80	0	0	0	20
5	0	0	0	0	0
Media %	64	0	0	0	16

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	1) Assunto mais espe-
2	X		cífico no título
3	X		versus idéia de lu
4	X		gar no cabeçalho
5	X		

TÍTULO: Centro Integrado do Rio Doce CIRU

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Pesquisa tecnológica - Política educacional Minas Gerais	379.3:62.001.5(815.1)
2. Pesquisa tecnológica - Política educacional	379.3:62.001.5
3. Tecnologia - Política educacional	379.3:62
4. Política educacional	379.3
5. Escola e as autoridades públicas	379

SIMILARIDADE

Palavras-chave

Níveis da cadeia	1	2	3	4	% Média
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
Media %	0	0	0	0	0

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		
5	X		

TÍTULO: Exportação de produtos primários não-tradicionais

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Exportação - Brasil	382.6(81)
2. Exportação	382.6
3. Comércio externo	382
4. Comércio	38
5. Ciências sociais	3

SIMILARIDADE  
Palavras-chave

Níveis da cadeia	1	2	3	4	% Média
1	80	0	0	0	20
2	100	0	0	0	25
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
Media %	36	0	0	0	9

Frases-chave

Níveis da cadeia	1	2	3	% Média
1	80	0	0	26,66
2	100	0	0	33,33
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
Media %	36	0	0	11,99

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		
5	X		

TÍTULO: Pesquisas em processo no Brasil

ÍNDICE EM CADEIA  
CABEÇALHOSNÚMERO DO  
ASSUNTO (CDU)

1. Pesquisa científica - Brasil	001.891(81)
2. Pesquisa científica	001.891
3. Organização do trabalho científico	001.89
4. Metodologia	001.8
5. Ciência Conhecimento	001

## SIMILARIDADE

Palavras-chave

Níveis da cadeia	1	2	3	% Média
1	80	0	80	53,33
2	80	0	0	26,66
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
Media %	32	0	16	15,99

Frases-chave

Níveis da cadeia	1	2	% Média
1	80	80	80
2	80	0	40
3	0	0	0
4	0	0	0
5	0	0	0
Media %	32	16	24

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	O termo-chave "Pesquisas em processo" foi considerado equivalente à "Pesquisa científica" conceito de tempo versus conceito de assunto.
2	X		
3	X		
4	X		
5	X		

TÍTULO: Anais do 2º curso de planejamento urbano e local; planejamento metropolitano, São Paulo, 29 de junho a 28 de julho de 1972

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Planejamento urbano - Documentos de seminários	711(063)
2. Planejamento urbano	711
3. Urbanização	71
4. Belas Artes	7

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia								%
	1	2	3	4	5	6	7	Média
1	0	70	80	80	0	0	0	32,85
2	0	0	80	80	0	0	0	22,85
3	0	0	0	100	0	0	0	14,28
4	0	0	0	0	0	0	0	0
Media %	0	17,5	40	65	0	0	0	17,49

## Frases-chave

Níveis da cadeia								%
	1	2	3	4	5	6		Média
1	0	70	80	0	80	0		38,33
2	0	0	100	0	80	0		30
3	0	0	80	0	0	0		13,33
4	0	0	0	0	0	0		0
Media %	0	17,5	65	0	40	0		20,41

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		

TÍTULO: Bibliografia brasileira de zoologia

ÍNDICE EM CADEIA  
CABEÇALHOSNÚMERO DO  
ASSUNTO (CDU)

1. Zoologia - Bibliografias - Brasil	016:59(81)
2. Zoologia - Bibliografias	016:59
3. Ciências Puras - Bibliografias	016:5
4. Bibliografias especializadas	016
5. Bibliografia	01

## SIMILARIDADE

Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	80	80
2	80	0	80	53,33
3	80	0	0	26,66
4	80	0	0	26,66
5	100	0	0	33,33
Media %	84	16	32	43,99

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		
5	X		

TÍTULO: Bibliografia brasileira de Física

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1.	Física - Bibliografias - Brasil	016:53(81)
2.	Física - Bibliografias	016:53
3.	Ciências Puras - Bibliografias	016:5
4.	Bibliografias especializadas	016
5.	Bibliografia	01

SIMILARIDADE

Palavras-chave

Níveis da cadeia	1	2	3	% Média
1	80	80	80	80
2	80	0	80	53,33
3	80	0	0	26,66
4	80	0	0	26,66
5	100	0	0	33,33
Media %	84	16	32	43,99

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		
5	X		

TÍTULO: Dirección de explotaciones agrícolas; administración, organización y técnica en la agricultura

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Administração agrícola	631.1
2. Agricultura Agronomia	631
3. Ciências aplicadas	6

SIMILARIDADE  
Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	% Média
1	0	0	80	80	0	0	80	34,28
2	0	0	100	0	0	0	100	28,57
3	0	0	0	0	0	0	0	0
Media %	0	0	60	26,66	0	0	60	20,95

Frases-chave

Níveis da cadeia	1	2	3	4	5	6	% Média
1	0	80	80	0	0	80	40
2	0	80	0	0	0	100	30
3	0	0	0	0	0	0	0
Media %	0	53,33	26,66	0	0	60	23,33

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		

## TÍTULO: Elaboração e avaliação de projetos

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Avaliação e elaboração de projetos	65.012.2.001.4/.8
2. Planejamento Planificação Programas Projetos	65.012.2
3. Métodos administrativos	65.012
4. Administração e organização da indústria, comércio e transportes	65
5. Ciências aplicadas	6

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	80	80
2	0	0	100	33,33
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
Média %	16	16	36	22,66

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	80	80	80
2	0	80	40
3	0	0	0
4	0	0	0
5	0	0	0
Média %	16	32	24

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		
5	X		

## TÍTULO: Indicadores econômicos conjunturais

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Custo de vida - Distrito Federal	338.585.3(817.4)
2. Custo de vida	338.585.3
3. Custos	338.58
4. Preço e valor	338.5
5. Situação econômica	338

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	% Média
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	80	70	50
Media %	0	16	14	10

## Frases-chave

Níveis da cadeia	1	2	% Média
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	80	70	75
Media %	16	14	15

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	O cabeçalho de nível 2, a pesar de perder o conceito indicativo de lugar, expri me melhor o assunto da obra que o livro.
2		X	
3	X		
4	X		
5	X		

TÍTULO: Primo piano nazionale di sviluppo 1972/74

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1.	Planos de desenvolvimento econômico nacionais-Brasil	338.984.3(81)
2.	Planos de desenvolvimento econômico nacionais	338.984.3
3.	Planos de desenvolvimento econômico	338.984
4.	Planificação econômico e controle	338.98
5.	Fatores de produção	338.9

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	80	80
2	80	80	80	80
3	80	0	80	53,33
4	80	0	0	26,66
5	0	0	0	0
Media %	64	32	48	47,99

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	80	80	80
2	80	80	80
3	80	80	80
4	80	0	40
5	0	0	0
Media %	64	48	56

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	
2		X	
3		X	
4		X	
5		X	

TÍTULO: Estudo dos orçamentos das unidades familiares do Distrito Federal.

ÍNDICE ■ CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1 Distribuição de renda - Desenvolvimento regional - Distrito Federal	338.984.2:339(817.4)
2. Distribuição de renda - Desenvolvimento regional	338.984.2:339
3. Desenvolvimento regional Planos de desenvolvimento econômico regionais	338.984.2
4. Planos de desenvolvimento econômico	338.984
5. Planificação econômica e controle	338.98

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia						%
	1	2	3	4	5	Média
1	0	0	0	0	80	16
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
Média %	0	0	0	0	16	3,2

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		Conceito "Unidades familiares" ausente do cabeçalho.
2	X		
3	X		
4	X		
5	X		

TÍTULO: Contribuição ao desenvolvimento da agroindústria Manga

ÍNDICE EM CADEIA  
CABEÇALHOSNÚMERO DO  
ASSUNTO (CDU)

1. Industrialização-Agricultura - Brasil	631:338.924(81)
2. Industrialização-Agricultura	631:338.924
3. Situação econômica - Agricultura	631:338
4. Economia - Agricultura	631:33
5. Agricultura Agronomia	631

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	0	20
2	0	0	90	0	22,5
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	100	0	25
Média	0	0	84	0	21,5

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		
5	X		

TÍTULO: Export directory of Brazil 1965

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1.	Exportação - Brasil - Guias	382.6 (81) (021)
2.	Exportação - Brasil	382.6 (81)
3.	Exportação	382.6
4.	Comércio externo	382
5.	Comércio	38

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	70	80	76,66
2	80	0	80	53,33
3	100	0	0	33,33
4	0	0	0	0
5	0	0	0	0
Média %	52	14	32	32,66

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	80	80	80
2	80	80	80
3	80	0	40
4	0	0	0
5	0	0	0
Média %	48	32	40

## ESPECIFICIDADE

Níveis da cadeia	KWOC	cadeia	Observações
1	X	X	O conceito indicativo de tempo do título "1965", não foi considerado palavra chave
2	X		
3	X		
4	X		
5	X		

TÍTULO: Banco Interamericano de desenvolvimento: manual de treinamento: estrutura, recursos, operações

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Bancos de desenvolvimento - Américas - Manuais	336.711(7/8)(021)
2. Bancos de desenvolvimento - Américas	336.711(7/8)
3. Bancos de desenvolvimento	336.711
4. Bancos	336.71
5. Economia	33

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia									%
	1	2	3	4	5	6	7	8	Média
1	80	0	80	80	0	0	0	0	30
2	80	0	80	0	0	0	0	0	20
3	80	0	80	0	0	0	0	0	20
4	100	0	0	0	0	0	0	0	12,5
5	0	0	0	0	0	0	0	0	0
Media %	68	0	48	16	0	0	0	0	16,5

## Frases-chave

Níveis da cadeia								%
	1	2	3	4	5	6	Média	
1	80	80	80	0	0	0	40	
2	80	80	0	0	0	0	26,66	
3	80	80	0	0	0	0	26,66	
4	80	0	0	0	0	0	13,33	
5	0	0	0	0	0	0	0	
Media %	64	32	48	0	0	0	21,33	

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		O nome do banco, elemento em questão, especificado no título e não no cabeçalho
2	X		
3	X		
4	X		
5	X		

TÍTULO: Leis e decretos, promulgados no período 1965 (janeiro) 1967  
(dezembro) versando sobre matéria de interesse específico do  
DNER

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Estradas de rodagem - Brasil - Legislação	625.711.3(81)(094)
2. Estradas de rodagem - Brasil	625.711.3(81)
3. Estradas de rodagem	625.711.3
4. Estradas	625.711
5. Engenharia de estradas	625.7

SIMILARIDADE  
Palavras-chave

Níveis da cadeia									%
	1	2	3	4	5	6	7	8	Média
1	80	0	0	0	0	0	0	0	10
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
Media %	16	0	0	0	0	0	0	0	2

Frases-chave

Níveis da cadeia						%
	1	2	3	4	5	Média
1	80	0	0	0	0	16
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
Media %	16	0	0	0	0	3,2

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	1) "DNER" engloba os con-
2	X		ceitos do cabeçalho
3	X		"Estradas de rodagem
4	X		no Brasil"
5	X		

TÍTULO: Legislação do ensino industrial

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Ensino industrial-Brasil-Legislação	373.6(81)(094)
2. Ensino industrial-Brasil	373.6(81)
3. Ensino industrial Institutos comerciais, industriais e profis- sionais.	373.6
4. Ensino secundário	373
5. Educação Ensino	37

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	80	80
2	0	80	80	53,33
3	0	80	80	53,33
4	0	80	0	26,66
5	0	100	0	33,33
Média %	16	84	48	49,33

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	80	80	80
2	0	80	40
3	0	100	50
4	0	80	40
5	0	80	40
Média %	16	84	50

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	
2	X	X	
3	X		
4	X		
5	X		

TÍTULO: A sociedade afluyente

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Classes sociais		323.3
Grupos sociais		
Fenômenos sociais		
2. Política interna		323
3. Ciência política		32
Política		
4. Ciências sociais		3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	% Média
1	80	0	40
2	0	0	0
3	0	0	0
4	80	0	40
Média %	40	0	20

## Frases-chave

Níveis da cadeia	1	% Média
1	80	80
2	0	0
3	0	0
4	80	80
Média %	40	40

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		

TÍTULO: Reforma administrativa

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Legislação governamental - Brasil	351(81)
2. Legislação governamental	351
Empresas públicas - Direito administrativo	
Empresas públicas - Administração	
Serviços públicos - Direito administrativo	
Serviços públicos - Administração	
3. Administração pública	35
Direito administrativo	
4. Ciências sociais	3

SIMILARIDADE

Palavras-chave

Níveis da cadeia	1	2	3	% Média
1	0	0	0	0
2	0	80	0	40
3	0	80	0	40
4	0	0	0	0
Media %	0	40	0	20

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Termos da cadeia mais próximos do que os termos do título
2		X	
3		X	
4	X		

TÍTULO: Leis e decretos promulgados no período 1962 (janeiro) 1964 (dezembro) versando sobre matéria de interesse orgânico do DNER

ÍNDICE EM CADEIA CABECALHOS	NÚMERO DO ASSUNTO (CDU)
1. Estradas de rodagem - Brasil - Legislação	625.711.3(81)(094)
2. Estradas de rodagem - Brasil	625.711.3(81)
3. Estradas de rodagem	625.711.3
4. Estradas	625.711
5. Engenharia de estradas	625.7

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	% Média
1	80	0	0	0	0	0	0	0	10
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
Média %	16	0	0	0	0	0	0	0	2

## Frases-chave

Níveis da cadeia	1	2	3	4	5	% Média
1	80	0	0	0	0	16
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
Média %	16	0	0	0	0	3,2

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	O conceito "DNER"
2	X		contém o signifi-
3	X		cado "Estradas de
4	X		rodagem - Brasil"
5	X		

TÍTULO: Government and technical innovation

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1.	Inovação tecnológica - Incentivos governamentais	338.983:62.001.7
2.	Tecnologia - Incentivos governamentais	338.983:62
3.	Incentivos governamentais	338.983
4.	Política industrial Planificação econômica e controle	338.98
5.	Fatores da produção	338.9

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	80	80
2	80	80	0	53,33
3	80	0	0	26,66
4	0	0	0	0
5	0	0	0	0
Media %	48	32	16	31,99

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	80	80	80
2	80	80	80
3	80	0	40
4	0	0	0
5	0	0	0
Media %	48	32	40

## ESPECIFICIDADE

Níveis da cadeia	RWOC	CADEIA	OBSERVAÇÕES
1		X	
2	X	X	
3	X		
4	X		
5	X		

TÍTULO: Agricultura e desenvolvimento

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Administração econômica - Agricultura	631.16
2. Administração agrícola	631.1
3. Agricultura Agronomia	63
4. Ciências aplicadas	6

SIMILARIDADE  
Palavras-chave

Níveis da cadeia	1	2	% Média
1	80	0	40
2	80	0	40
3	100	0	50
4	0	0	0
Media %	65	0	32,5

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	O título "Agricultura e desenvolvimento" não exprime exatamente a idéia da obra que é a "Administração econômica na Agricultura"
2		X	
3	X		
4	X		

TÍTULO: Problemas sobre teoria dos preços

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Preços e valor	338.5
2. Produção	338
3. Economia	33
4. Ciências sociais	3

**SIMILARIDADE**  
**Palavras-chave**

Níveis da cadeia				%
	1	2	3	Média
1	0	0	80	26,66
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
Media %	0	0	20	6,66

**Frases-chave**

Níveis da cadeia			%
	1	2	Média
1	0	80	40
2	0	0	0
3	0	0	0
4	0	0	0
Media %	0	20	10

**ESPECIFICIDADE**

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		

TÍTULO: Guia prático trabalhista

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Legislação e fiscalização do trabalho-Brasil-Guias	351.83(81)(021)
2. Legislação e fiscalização do trabalho-Brasil	351.83(81)
3. Legislação e fiscalização do trabalho	351.83
4. Legislação governamental	351
5. Direito administrativo	35
Administração publica	35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	0	80	53,33
2	0	0	80	26,66
3	0	0	80	26,66
4	0	0	0	0
5	0	0	0	0
Media %	16	0	48	21,33

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	80	80	80
2	0	80	40
3	0	80	40
4	0	0	0
5	0	0	0
Media %	16	48	32

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Nos níveis 2 e 3 considerou-se a especificidade do assunto mais importante que o conceito indicativo de forma contido no título.
2		X	
3		X	
4	X		
5	X		

TÍTULO: Coleção das leis de 1967

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1.	Legislação governamental - Brasil	351(81)
2.	Legislação governamental	351
	Empresas públicas - Direito administrativo	
	Empresas públicas - Administração	
	Serviços públicos - Direito administrativo	
	Serviços públicos - Administração	
3.	Administração pública	35
	Direito administrativo	
4.	Ciências sociais	3

SIMILARIDADE

Palavras-chave

Níveis da cadeia			% Média
	1	2	
1	0	80	40
2	0	80	40
3	0	0	0
4	0	0	0
<b>Média %</b>	<b>0</b>	<b>40</b>	<b>20</b>

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	
2		X	
3		X	
4	X		

TÍTULO: A transferência de tecnologia no desenvolvimento industrial do Brasil

ÍNDICE EM CADEIA  
CABEÇALHOS

NÚMERO DO  
ASSUNTO (CDU)

1. Transferência de tecnologia - Industrias - Brasil	67:62.001.8(81)
2. Transferência de tecnologia - Industrias	67:62.001.8
3. Tecnologia industrial	67:62
4. Indústrias	67
5. Ciências aplicadas	6

SIMILARIDADE

Palavras-chave

Níveis da cadeia						%
	1	2	3	4	5	Média
1	80	80	0	80	80	64
2	80	80	0	80	0	48
3	0	80	0	80	0	32
4	0	0	0	100	0	20
5	0	0	0	0	0	0
Média %	32	48	0	68	16	32,8

Frases-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	80	80
2	80	80	0	53,33
3	80	80	0	53,33
4	0	80	0	26,66
5	0	0	0	0
Média %	48	64	16	42,66

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		
5	X		

TÍTULO: Coleção das leis de 1946

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Legislação governamental - Brasil	351(81)
2. Legislação governamental	351
Empresas públicas - Direito administrativo	
Empresas públicas - Administração	
Serviços públicos - Direito administrativo	
Serviços públicos - Administração	
3. Administração pública	35
Direito administrativo	
4. Ciências sociais	3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	% Média
1	0	80	40
2	0	80	40
3	0	0	0
4	0	0	0
Média %	0	40	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	
2		X	
3		X	
4	X		

TÍTULO: Coleção das leis de 1948

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1.	Legislação governamental - Brasil	351(81)
2.	Legislação governamental	351
	Empresas públicas - Direito administrativo	
	Empresas públicas - Administração	
	Serviços públicos - Direito administrativo	
	Serviços públicos - Administração	
3.	Administração pública	35
	Direito administrativo	
4.	Ciências sociais	3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia			%
	1	2	Média
1	0	80	40
2	0	80	40
3	0	0	0
4	0	0	0
Media %	0	40	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	
2		X	
3		X	
4	X		

TÍTULO: La infraestructura en la planificación del desarrollo

ÍNDICE EM CADEIA  
CABEÇALHOS

NÚMERO DO ASSUNTO (CDU)

1. Planos de desenvolvimento econômico - América Latina	338.984 (8=6)
2. Planos de desenvolvimento econômico	338.984
3. Planificação econômica e controle	338.98
4. Fatores da produção	338.9
5. Situação econômica	338

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	0	80	80	53,33
2	0	80	80	53,33
3	0	80	0	26,66
4	0	0	0	0
5	0	0	0	0
Média %	0	48	32	26,66

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	"Planificación" = "Planos"
2	X		"desarrollo" = "desenvolvimento econômico"
3	X		"infraestructura" equivalente a "América Latina"
4	X		
5	X		

TÍTULO: Relatório preliminar de desenvolvimento local integrado do município de Grão Mongol

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos - Administração municipal Grão Mongol		352:351(815.12)
2. Serviços públicos - Administração municipal		352:351
3. Administração pública - Municipal		352:35
4. Administração municipal		352
5. Administração pública		35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia								%
	1	2	3	4	5	6	7	Média
1	0	0	0	0	0	80	80	22,85
2	0	0	0	0	0	80	0	11,42
3	0	0	0	0	0	80	0	11,42
4	0	0	0	0	0	80	0	11,42
5	0	0	0	0	0	0	0	0
Média %	0	0	0	0	0	64	16	11,42

## Frases-chave

Níveis da cadeia						%
	1	2	3	4	5	Média
1	0	0	0	80	80	32
2	0	0	0	80	0	16
3	0	0	0	80	0	16
4	0	0	0	80	0	16
5	0	0	0	0	0	0
Média %	0	0	0	64	16	16

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	1) Assunto mais expli- citos no cabeçalho
2	X		
3	X		
4	X		
5	X		

TÍTULO: Relatório preliminar de desenvolvimento integrado do município de Codô.

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos - Administração municipal - Codô	352:351(812.12)
2. Serviços públicos - Administração municipal	352:351
3. Administração pública - Municipal	352:35
4. Administração municipal	352
5. Administração pública	35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Média %	0	0	0	0	64	16	13,33

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Média %	0	0	64	16	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a cadeia foi mais específica porque assuntos foram considerados mais importantes do que foram de apresentação.
2	X		
3	X		
4	X		
5	X		

## TÍTULO: Classificação de cargos no serviço público

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Classificação de cargos - Funcionários públicos-Brasil	35.081 (81)
2. Classificação de cargos - Funcionários públicos	35.081
3. Funcionários públicos - Direito administrativo Funcionários públicos - Administração	35.08
4. Administração pública Direito administrativo	35
5. Ciências sociais	3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	80	80	70	80	77,5
2	80	80	70	80	77,5
3	0	0	70	80	37,5
4	0	0	0	80	20
5	0	0	0	0	0
Média %	32	32	42	64	42,5

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	80	80	80
2	80	80	80
3	0	80	40
4	0	80	40
5	0	0	0
Média %	32	64	48

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	
2	X		
3	X		
4	X		
5	X		

TÍTULO: Estatutos sociais

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Companhia Estadual de Gas - Guanabara - Estatutos		352.9(81)(094.7)
2. Companhia Estadual de Gas - Guanabara		352.9(81)
3. Companhia Estadual de Gas Autarquias locais		352.9
4. Administração local Administração municipal		352
5. Administração pública Direito administrativo		35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	% Média
1	80	0	40
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
Média %	16	0	8

## Frases-chave

Níveis da cadeia	1	% Média
1	80	80
2	0	0
3	0	0
4	0	0
5	0	0
Média %	16	16

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Cabeçalhos do 4º
2		X	e 5º níveis tão
3		X	vagos quanto o
4	X	X	título
5	X	X	

TÍTULO: Estratégia industrial e empresas internacionais: posição re  
lativa da América Latina e do Brasil

ÍNDICE EM CADEIA  
CABEÇALHOS

NÚMERO DO  
ASSUNTO (CDU)

1. Investimentos internacionais	336.745.3(100)
2. Investimentos estrangeiros	336.745.3
3. Economia	33
4. Ciências sociais	3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	% Média
1	0	0	0	80	0	0	0	0	10
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
Média %	0	0	0	20	0	0	0	0	2,5

## Frases-chave

Níveis da cadeia	1	2	3	4	5	6	% Média
1	0	0	80	0	0	0	13,33
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
Média %	0	0	20	0	0	0	3,33

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		

TÍTULO: Produção agrícola; produção dos municípios, discriminada se-  
gundo os produtos cultivados, culturas permanentes

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Produção agrícola - São Paulo		631.002.2(816.1)
2. Produção agrícola		631.002.2
3. Agricultura Agronomia		631
4. Ciências aplicadas		6

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	% Média
1	80	80	0	0	80	0	0	0	30
2	80	80	0	0	80	0	0	0	30
3	0	100	0	0	0	0	0	0	12,5
4	0	0	0	0	0	0	0	0	0
Média %	40	65	0	0	40	0	0	0	18,12

## Frases-chave

Níveis da cadeia	1	2	3	4	5	% Média
1	80	80	0	80	0	48
2	100	80	0	80	0	52
3	80	0	0	0	0	16
4	0	0	0	0	0	0
Média %	65	40	0	40	0	29

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		Há melhor especificação no título sobre o tipo de produção
2	X		
3	X		
4	X		

TÍTULO: Sudene dez anos

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. SUDENE	338.984.2:061.1(812/814)
Instituições oficiais para o desenvolvimen to regional - Nordeste	
2. Instituições oficiais para o desenvolvimen regional	338.984.2:061.1
3. Planos de desenvolvimento econômico regio nais	338.984.2
4. Planos de desenvolvimento econômico	338.984
5. Planificação econômica e controle	338.98

SIMILARIDADE

Palavras-chave

Níveis da cadeia	l	% Média
1	100	100
2	0	0
3	0	0
4	0	0
5	0	0
Média %	20	20

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		
5	X		

TÍTULO: Relatório preliminar de desenvolvimento integrado do município de Soure.

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos - Administração municipal - Soure	352.351(811.52)
2. Serviços públicos - Administração municipal	352.351
3. Administração pública - Municipal	352.35
4. Administração municipal	352
5. Administração pública	35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Media %	0	0	0	0	64	16	13,33

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Media %	0	0	64	16	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a cadeia foi mais específica porque assuntos foram considerados mais importantes do que formas de apresentação.
2	X		
3	X		
4	X		
5	X		

TÍTULO: Relatório preliminar de desenvolvimento integrado do município de Parintins

ÍNDICE EM CADEIA  
CABEÇALHOS

NÚMERO DO  
ASSUNTO (CDU)

1. Serviços públicos - Administração municipal Parintins	352:351(811.32)
2. Serviços públicos - Administração municipal	352:351
3. Administração pública - Municipal	352:35
4. Administração municipal	352
5. Administração pública	35

SIMILARIDADE

Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Media %	0	0	0	0	64	16	13,33

Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Media %	0	0	64	16	20

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a cadeia
2	X		foi mais específica porque
3	X		assuntos foram considerados
4	X		mais importantes do que for
5	X		mas de apresentação.

TÍTULO: Relatório preliminar de desenvolvimento integrado do município de Araguaína.

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos - Administração municipal Araguaína	352:351(817.32)
2. Serviços públicos - Administração municipal	352:351
3. Administração pública - Municipal	352:35
4. Administração municipal	352
5. Administração pública	35

SIMILARIDADE  
Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Média %	0	0	0	0	64	16	13,33

Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Média %	0	0	64	16	20

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a cadeia foi mais específica porque assuntos foram considerados mais importantes do que foram de apresentação
2	X		
3	X		
4	X		
5	X		

TÍTULO: A tecnologia na educação do 1º grau

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Tecnologia - Ensino primário e secundário	372/373:62
2. Ensino primário e secundário Ensino médio	372/373
3. Educação Ensino	37
4. Ciências sociais	3

SIMILARIDADE

Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	70	0	50
2	0	70	0	23,33
3	0	100	0	33,33
4	0	0	0	0
Media %	20	60	0	26,66

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		

TÍTULO: Geologia das quadrículas de Porto Seguro Nordeste e Magiquiçaba

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Agrogeologia-Bahia	55:631.4(814.2)	
2. Agrogeologia Solos-Agricultura-Geologia	55:631.4	
3. Geologia	55	
4. Ciências Puras	5	

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	% Média
1	0	0	0	0	0	0
2	80	0	0	0	0	16
3	100	0	0	0	0	20
4	0	0	0	0	0	0
Media %	45	0	0	0	0	9

## Frases-chave

Níveis da cadeia	1	2	3	4	% Média
1	0	0	0	0	0
2	80	0	0	0	20
3	100	0	0	0	25
4	0	0	0	0	0
Media %	45	0	0	0	11,25

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	1) O prefixo "agro" está ausente do título, mas este em compensação especifica melhor o conceito do lugar, apresentando dois termos a mais - contra um.
2	X		
3	X		
4	X		

TÍTULO: Manual de patrimônio.

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Patrimônio - Administração - Manuais	658.787.6(021)
2. Patrimônio - Administração Inventários - Administração Controle de stocks - Administração	658.787.6
3. Administração de armazéns Expedição - Administração	658.78
4. Organização industrial e comercial	658
5. Administração e organização da indústria, comércio e transportes	65

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	% Média
1	80	80	80
2	0	80	40
3	0	0	0
4	0	0	0
5	0	0	0
Média %	16	32	24

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Sendo o assunto mais importante que a forma "Administração" foi considerado mais específico que "Manual".
2		X	
3	X		
4	X		
5	X		

TÍTULO: Teoria e política do desenvolvimento econômico

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Desenvolvimento econômico Política econômica	330.19
2. Princípios leis e sistemas econômicos	330.1
3. Economia	33
4. Ciências Sociais	3

SIMILARIDADE  
Palavras-chave

Níveis da cadeia					% Média
	1	2	3	4	
1	0	80	80	80	60
2	0	0	0	80	20
3	0	0	0	100	25
4	0	0	0	0	0
Media %	0	20	20	65	26,25

Frases-chave

Níveis da cadeia				% Média
	1	2	3	
1	0	80	100	60
2	0	0	80	26,66
3	0	0	80	26,66
4	0	0	0	0
Media %	0	20	65	28,33

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		

TÍTULO: Plano decenal de desenvolvimento econômico e social; Energia

ÍNDICE EM CADEIA  
CABEÇALHOS

NÚMERO DO  
ASSUNTO (CDU)

1. Planos de desenvolvimento econômico nacionais - Brasil	338.984.3(81)
2. Planos de desenvolvimento econômico nacionais	338.984.3
3. Planos de desenvolvimento econômico	338.984
4. Planificação econômica e controle	338.98
5. Fatores de produção	338.9

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	% Média
1	80	80	80	0	0	48
2	80	80	80	0	0	48
3	80	80	80	0	0	48
4	80	0	80	0	0	32
5	0	0	0	0	0	0
Média %	64	48	64	0	0	35,2

## Frases-chave

Níveis da cadeia	1	2	3	4	% Média
1	80	80	0	0	40
2	80	80	0	0	40
3	80	80	0	0	40
4	80	80	0	0	40
5	0	0	0	0	0
Média %	64	64	0	0	32

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		
5	X		

TÍTULO: Mar territorial

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Mar - Direito internacional	341.225
2. Direito internacional	341
3. Direito Legislação Jurisprudência	34
4. Ciências sociais	3

SIMILARIDADE  
Palavras-chave

Níveis da cadeia	1	2	% Média
1	80	0	40
2	0	0	0
3	0	0	0
4	0	0	0
Media %	20	0	10

Frases-chave

Níveis da cadeia	1	% Média
1	80	80
2	0	0
3	0	0
4	0	0
Media	20	20

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Assunto melhor
2	X		especificado
3	X		no 1º nível da
4	X		cadeia

TÍTULO: Relatório preliminar de desenvolvimento integrado do município de Esplanada

ÍNDICE EM CADEIA  
CABEÇALHOS

NÚMERO DO  
ASSUNTO (CDU)

1. Serviços públicos - Administração municipal Esplanada	352:351(814.22)
2. Serviços públicos - Administração municipal	352:351
3. Administração pública - Municipal	352:35
4. Administração municipal	352
5. Administração pública	35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Média %	0	0	0	0	64	16	13,33

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Média %	0	0	64	16	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a cadeia
2	X		foi mais específica porque
3	X		assuntos foram considerados
4	X		mais importantes do que for
5	X		mas de apresentação

TÍTULO: Plano de desenvolvimento integrado do município de Taubaté

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos - Administração municipal - Taubaté	352:351(815.52)
2. Serviços públicos - Administração municipal	352:351
3. Administração pública - Municipal	352:35
4. Administração municipal	352
5. Administração pública	35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia						%
	1	2	3	4	5	Média
1	0	0	0	80	80	32
2	0	0	0	80	0	16
3	0	0	0	80	0	16
4	0	0	0	80	0	16
5	0	0	0	0	0	0
Média %	0	0	0	64	16	16

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Média %	0	0	64	16	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Apesar de ser o título pouco expressivo, do segundo nível em diante ele contém o conceito de lugar "Taubaté" ausente a partir de então do cabeçalho
2	X		
3	X		
4	X		
5	X		

TÍTULO: Projeto especial força tarefa 1, diagnóstico preliminar das áreas urbanas do estado do Acre e dos territórios de Amapá, Roraima e Rondônia

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos - Administração municipal - Região Norte	352:351(811)
2. Serviços públicos - Administração municipal	352:351
3. Administração pública - Municipal	352:35
4. Administração Municipal	352
5. Administração pública	35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	% Média
1	0	0	0	0	0	0	70	0	0	0	0	0	0	0	5
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Média %	0	0	0	0	0	0	14	0	0	0	0	0	0	0	1

## Frases-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	9	10	% Média
1	0	0	0	70	0	0	0	0	0	0	7
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0
Média %	0	0	0	14	0	0	0	0	0	0	1,4

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	1) palavras-chave do título não-si- gnificativas até a palavra "áreas" inclusive. "Áreas urbanas" = "Muni- cipal", "Acre, Amapá, etc" = "Regi- ão Norte"
2	X		
3	X		
4	X		
5	X		
			2) idéia de lugar no título mais espe- cifica que assunto do cabeçalho.

TÍTULO: O modelo político brasileiro e outros ensaios

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Política econômica - Brasil	330.19(81)
2. Política econômica Desenvolvimento econômico	330.19
3. Princípios, leis e sistemas econômicos	330.1
4. Economia	33
5. Ciências sociais	3

SIMILARIDADE  
Palavras-chave

Níveis da cadeia	1	2	3	4	% Média
1	0	80	80	0	40
2	0	80	0	0	20
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
Média %	0	32	16	0	12

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		
5	X		

TÍTULO: Estatística básica para planificação

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1.	Estatística - Planificação econômica e controle	338.98:31
2.	Planificação econômica e controle Política industrial	338.98
3.	Fatores da produção	338.9
4.	Situação econômica	338
5.	Economia	33

## SIMILARIDADE

Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	0	80	53,33
2	0	0	80	26,66
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
Média %	16	0	32	15,99

## Frases-chave

Níveis da cadeia			%
	1	2	Média
1	80	80	80
2	0	80	40
3	0	0	0
4	0	0	0
5	0	0	0
Média %	16	32	24

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		
5	X		

TÍTULO: Exercícios de história econômica do Brasil

ÍNDICES EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Economia - Brasil - História	33(81)(091)
2. Economia - Brasil	33(81)
3. Economia	33
4. Ciências sociais	3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	80	80	80	60
2	0	0	80	80	40
3	0	0	100	0	25
4	0	0	0	0	0
Media %	0	20	65	40	31,25

## Frases-chave

Níveis da cadeia				%
	1	2	3	Média
1	0	80	80	53,33
2	0	80	80	53,33
3	0	80	0	26,66
4	0	0	0	0
Media %	0	60	40	33,33

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		Título contém a
2	X		palavra-chave
3	X		"exercícios",
4	X		conceito de for
			ma de apresenta
			ção ausente do
			cabeçalho

TÍTULO: Oscilador a diodo tunel acoplado a uma microlinha de transmissão

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Diodos - Osciladores	621.373:621.385
2. Eletrônica - Osciladores	621.373:621.38
3. Engenharia eletrotécnica - Osciladores	621.373:621.3
4. Osciladores	621.373
5. Ondas elétricas e eletromagnéticas	621.37

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	80	80	0	0	0	0	26,66
2	80	0	0	0	0	0	13,33
3	80	0	0	0	0	0	13,33
4	100	0	0	0	0	0	16,66
5	0	0	0	0	0	0	0
Media %	68	16	0	0	0	0	13,99

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	80	80	0	0	40
2	80	0	0	0	20
3	80	0	0	0	20
4	100	0	0	0	25
5	0	0	0	0	0
Media %	68	16	0	0	21

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		
5	X		

TÍTULO: Influência das tensões mecânicas biaxiais na foto-corrente de heterojunções de Ge-Si

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Semicondutores - Eletricidade - Física Campo magnético - Física	537.312
2. Transmissão de eletricidade - Física	537.31
3. Correntes elétricas - Física Eletrocincemática - Física	537.3
4. Eletricidade - Física	537
5. Física	53

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	% Média
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
Média	0	0	0	0	0	0	0	0

## Frases-chave

Níveis da cadeia	1	2	3	4	5	6	% Média
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
Média %	0	0	0	0	0	0	0

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		
5	X		

TÍTULO: Geologia da folha de Mascote Noroeste

ÍNDICE EM CADEIA  
CABEÇALHOSNÚMERO DO  
ASSUNTO (CDU)

1. Agrogeologia - Bahia	55:631.4(814.2)
2. Agrogeologia Solos - Agricultura - Geologia	55:631.4
3. Geologia	55
4. Ciências Puras	5

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	0	0	0
2	80	0	0	0	20
3	100	0	0	0	25
4	0	0	0	0	0
Média %	45	0	0	0	11,25

## Frases-chave

Níveis da cadeia				%
	1	2	3	Média
1	0	0	0	0
2	80	0	0	26,66
3	100	0	0	33,33
4	0	0	0	0
Média %	45	0	0	14,99

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		1) "Agrogeologia" é mais específico do que "Geologia" mas "Mascote Noroeste" é mais específico do que "Bahia" e a idéia de "fôlha" está ausente do cabeçalho
2	X		
3	X		
4	X		

TÍTULO: Elementos para planejamento das atividades da zootecnia no Brasil

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Pesquisa científica - Zootecnia - Brasil	636:001.891(81)
2. Pesquisa científica - Zootecnia	636:001.891
3. Política científica - Zootecnia	636:001.89
4. Zootecnia	636
5. Ciências aplicadas	6

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	% Média
1	0	0	0	80	80	32
2	0	0	0	80	0	16
3	0	0	0	80	0	16
4	0	0	0	100	0	20
5	0	0	0	0	0	0
Media %	0	0	0	68	16	16,8

## Frases-chave

Níveis da cadeia	1	2	3	4	% Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	100	0	25
5	0	0	0	0	0
Media %	0	0	68	16	21

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		
5	X		

TÍTULO: Relatório preliminar de desenvolvimento integrado do município de Vigia

ÍNDICE EM CADEIA  
CABEÇALHOS

NÚMERO DO  
ASSUNTO (CDU)

1. Serviços públicos - Administração municipal - Vigia	352:351(811.52)
2. Serviços públicos - Administração municipal	352:351
3. Administração pública - Municipal	352:35
4. Administração municipal	352
5. Administração pública.	35

SIMILARIDADE

Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Média %	0	0	0	0	64	16	13,33

Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Média %	0	0	64	16	20

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a cadeia
2	X		foi mais específica porque
3	X		assuntos foram considerados
4	X		mais importantes do que for
5	X		mas de apresentação

## TÍTULO: Integração regional e desenvolvimento do Nordeste

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Planos de desenvolvimento econômico regionais - Região Nordeste - Documentos de seminários		338.984 2(812/814) (063)
2. Planos de desenvolvimento econômico regionais - Região Nordeste		338.984 2(812/814)
3. Planos de desenvolvimento econômico regionais		338.984.2
4. Planos de desenvolvimento econômico		338.984
5. Planificação econômica e controle		338.98

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	80	80	80	60
2	0	80	80	80	60
3	0	80	80	0	40
4	0	0	80	0	20
5	0	0	0	0	0
Media %	0	48	64	32	36

## Frases-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	80	80
2	80	80	80	80
3	80	80	0	53,33
4	0	80	0	26,66
5	0	0	0	0
Media %	48	64	32	47,99

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	
2		X	
3	X	X	
4	X		
5	X		

TÍTULO: Relatório preliminar de desenvolvimento integrado do município de Coari

	ÍNDICE EM CADEIA CAREÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos - Administração municipal - Coari		352:351(811.32)
2. Serviços públicos - Administração municipal		352:351
3. Administração pública - Municipal		352:35
4. Administração municipal		352
5. Administração pública		35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Media %	0	0	0	0	64	16	13,33

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Media %	0	0	64	16	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a cadeia
2	X		foi mais específica porque
3	X		assuntos foram considerados
4	X		mais importantes do que for
5	X		mas de apresentação

TÍTULO: Relatório preliminar de desenvolvimento integrado do município de Camocim

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos - Administração municipal - Camocim	352:351(813.12)
2. Serviços públicos - Administração municipal	352:351
3. Administração pública - Municipal	352:35
4. Administração municipal	352
5. Administração pública	35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Média %	0	0	0	0	64	16	13,33

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Média %	0	0	64	16	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a cadeia
2	X		foi mais específica porque
3	X		assuntos foram considerados
4	X		mais importantes do que for
5	X		mas de apresentação

TÍTULO: Relatório preliminar de desenvolvimento integrado do município de Brumado

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos-Administração municipal-Brumado		352:351(814.22)
2. Serviços públicos-Administração municipal		352:351
3. Administração pública-Municipal		352:35
4. Administração municipal		352
5. Administração pública		35

SIMILARIDADE  
Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Media %	0	0	0	0	64	16	13,33

Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Media %	0	0	64	16	20

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a cadeia foi mais específica porque assuntos foram considerados mais importantes do que formas de apresentação.
2	X		
3	X		
4	X		
5	X		

TÍTULO: Relatório preliminar de desenvolvimento integrado do município de Lagarto.

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos-Administração municipal-Lagarto	352:351 (814.12)
2. Serviços públicos-Administração municipal	352:351
3. Administração pública-Municipal	352:35
4. Administração municipal	352
5. Administração pública	35

SIMILARIDADE

Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	% Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Média %	0	0	0	0	64	16	13,33

Frases-chave

Níveis da cadeia	1	2	3	4	% Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Média %	0	0	64	16	20

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a
2	X		cadeia foi mais es-
3	X		pecífica porque as-
4	X		suntos foram consi-
5	X		derados mais impor-
			taentes do que formas
			de apresentação.

TÍTULO: Relatório preliminar de desenvolvimento  
integrado do município de Chapadinha

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos-Administração municipal-Chapadinha	352:351(812.12)
2. Serviços públicos-Administração municipal	352:351
3. Administração pública-Municipal	352:35
4. Administração municipal	352
5. Administração pública	35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Média %	0	0	0	0	64	16	13,33

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	30	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Média %	0	0	64	16	20

## ESPECIFICIDADE

Níveis da cadeia	KWCC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a cadeia foi mais específica porque assuntos foram considerados mais importantes do que formas de apresentação.
2	X		
3	X		
4	X		
5	X		

TÍTULO: Relatório preliminar de desenvolvimento  
integrado do município de Tianguá

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Serviços públicos-Administração municipal-Tianguá	352:351(813.12)
2. Serviços públicos-Administração municipal	352:351
3. Administração pública-Municipal	352:35
4. Administração municipal	352
5. Administração pública	35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	% Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Média %	0	0	0	0	64	16	13,33

## Frases-chave

Níveis da cadeia	1	2	3	4	% Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Média %	0	0	64	16	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a
2	X		cadeia foi mais es-
3	X		pecífica porque as-
4	X		suntos foram consi-
5	X		derados mais impor-
			tantes do que formas
			de apresentação.

TÍTULO: Relatório preliminar de desenvolvimento  
integrado do município de São Raimundo Nonato

ÍNDICE EM CADEIA  
CABEÇALHOS

NÚMERO DO  
ASSUNTO (CDU)

1. Serviços públicos-Administração municipal- São Raimundo Nonato	352:351(812.22)
2. Serviços públicos-Administração municipal	352:351
3. Administração pública-Municipal	352:35
4. Administração municipal	352
5. Administração pública	35

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	0	0	0	0	80	80	26,66
2	0	0	0	0	80	0	13,33
3	0	0	0	0	80	0	13,33
4	0	0	0	0	80	0	13,33
5	0	0	0	0	0	0	0
Média %	0	0	0	0	64	16	13,33

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	0	0	80	80	40
2	0	0	80	0	20
3	0	0	80	0	20
4	0	0	80	0	20
5	0	0	0	0	0
Média %	0	0	64	16	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	No primeiro nível a cadeia foi mais específica porque assuntos foram considerados mais importantes do que formas de apresentação.
2	X		
3	X		
4	X		
5	X		

TÍTULO: Brasil 2002

	ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1.	Desenvolvimento econômico - Brasil	330.19 (81)
2.	Desenvolvimento econômico Política econômica	330.19
3.	Economia	33
4.	Ciências Sociais	3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	% Média
1	80	80
2	0	0
3	0	0
4	0	0
Média %	20	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Nos níveis 2 e 3 a palavra-chave do tí- tulo "Brasil" foi considerada de mes- mo grau de especifi- cidade que os cabe- çalhos destes níveis
2	X	X	
3	X	X	
4	X		

TÍTULO: Bridges

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Pontes	624.21
2. Engenharia civil	624
3. Tecnologia Engenharia	62
4. Ciências aplicadas	6

SIMILARIDADE

Palavras-chave

Níveis da cadeia	1	% Média
1	100	100
2	0	0
3	0	0
4	0	0
Media %	25	25

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		

TÍTULO: A previdência social brasileira interpretada, guia prático

ÍNDICE EM CADEIA CABEÇALHO	NÚMERO DO ASSUNTO (CDU)
1. Previdência social - Legislação - Brasil - Guias	351.84(81)(021)
2. Previdência social - Legislação - Brasil	351.81(81)
3. Previdência social - Legislação	351.81
4. Legislação governamental	351
5. Direito administrativo Administração pública	35

## SIMILARIDADE

Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	% Média
1	80	80	80	0	80	0	53,33
2	80	80	80	0	0	0	40
3	80	80	0	0	0	0	26,66
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
Media %	48	48	32	0	16	0	23,99

## Frases-chave

Níveis da cadeia	1	2	3	4	% Média
1	80	80	0	80	60
2	80	80	0	0	40
3	80	0	0	0	20
4	0	0	0	0	0
5	0	0	0	0	0
Media %	48	32	0	16	24

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	
2	X	X	
3	X		
4	X		
5	X		

TÍTULO: A indústria química brasileira; limitações e perspectivas  
ao seu desenvolvimento

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Desenvolvimento - Indústrias químicas - Brasil	66.001.6(81)
2. Desenvolvimento - Indústrias químicas	66.001.6
3. Indústrias químicas Tecnologia química	66
4. Ciências aplicadas	6

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia							%
	1	2	3	4	5	6	Média
1	80	80	80	0	0	80	53,33
2	80	80	0	0	0	80	40
3	80	80	0	0	0	0	26,66
4	0	0	0	0	0	0	0
Média %	60	60	20	0	0	40	29,99

## Frases-chave

Níveis da cadeia					%
	1	2	3	4	Média
1	80	80	0	80	60
2	80	0	0	80	40
3	100	0	0	0	25
4	0	0	0	0	0
Média %	65	20	0	40	31,25

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		

TÍTULO: Estatística de comércio exterior, conceitos e definições

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO ASSUNTO (CDU)	DC
1. Estatística - Comércio externo	382:31	
2. Comércio externo	382	
3. Comércio	38	
4. Ciências sociais	3	

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia						%
	1	2	3	4	5	Média
1	80	80	80	0	0	48
2	0	80	80	0	0	32
3	0	100	0	0	0	20
4	0	0	0	0	0	0
Média %	20	65	40	0	0	25

## Frases-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	0	53,33
2	0	100	0	33,33
3	0	80	0	26,66
4	0	0	0	0
Média %	20	65	0	28,33

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		

TÍTULO: Amazônia brasileira: catálogo da exposição

ÍNDICE EM CADEIA  
CABEÇALHOSNÚMERO DO  
ASSUNTO (CDU)

- |    |                                                   |                 |
|----|---------------------------------------------------|-----------------|
| 1. | Bibliografias- Amazônia - Catálogos de exposições | 016(811)(061.4) |
| 2. | Bibliografias- Amazônia                           | 016(811)        |
| 3. | Bibliografias especializadas                      | 016             |
| 4. | Bibliografia                                      | 01              |

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	% Média
1	80	0	80	80	60
2	80	0	0	0	20
3	0	0	0	0	0
4	0	0	0	0	0
Média %	40	0	20	20	20

## Frases-chave

Níveis da cadeia	1	2	3	% Média
1	80	0	80	53,33
2	80	0	0	26,66
3	0	0	0	0
4	0	0	0	0
Média %	40	0	20	19,99

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Cadeia carrega o conceito de "bibliografia" ausente do título. No nível 2, a cadeia perde, entretanto, a frase "catálogos de exposições", equiparando-se ao título
2	X	X	
3	X		
4	X		

TÍTULO: Programa educacional para clientes

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Computadores - Catálogos	681.142(085)
2. Computadores	681.142
3. Indústrias especializadas	68
4. Ciências aplicadas	6

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	% Média
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
Média %	0	0	0	0

## Frases-chave

Níveis da cadeia	1	2	% Média
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
Média %	0	0	0

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	Título inexpressivo.
2		X	Cabeçalho do 4º nível
3		X	mais inexpressivo do
4	X		que o título.

TÍTULO: Programa de apoio à infra-estrutura dos grandes centros urbanos do Nordeste BNB, BNH, CEF, FINEP.

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Financiamento - Desenvolvimento econômico Nordeste	330.19:336.77(812/814)
2. Financiamento - Desenvolvimento econômico	330.19:336.77
3. Desenvolvimento econômico	330.19
4. Economia	33
5. Ciências sociais	3

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	9	10	11	% Média
1	0	0	0	0	0	0	80	0	0	0	0	7,27
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
Média %	0	0	0	0	0	0	16	0	0	0	0	1,45

## Frases-chave

Níveis da cadeia	1	2	3	4	5	6	7	8	9	% Média
1	0	0	0	0	80	0	0	0	0	8,88
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
Média %	0	0	0	0	16	0	0	0	0	1,77

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X		
2	X		
3	X		
4	X		
5	X		
			A terminologia do título é mais específica. Há também os termos "BNB", "BNH", "CEF", "FINEP", ausentes do cabeçalho

TÍTULO: El mercado comun latinoamericano

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Mercado comum - América Latina	337.91 (8=6)
2. Mercado comum	337.91
3. Comércio livre Alfândegas	337
4. Economia	33
5. Ciências Sociais	3

SIMILARIDADE

Palavras-chave

Níveis da cadeia				%
	1	2	3	Média
1	30	80	80	80
2	30	80	0	53,33
3	70	0	0	23,33
4	0	0	0	0
5	0	0	0	0
Média %	46	32	16	31,33

Frases-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	80	80
2	100	0	0	50
3	90	0	0	45
4	0	0	0	0
5	0	0	0	0
Média %	54	16	35	

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		
5	X		

TÍTULO: Guia Interinvest: o Brasil e o capital internacional

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Investimentos estrangeiros-Brasil-Guias	336.745.3(81)(021)
2. Investimentos estrangeiros-Brasil	336.745.3(81)
3. Investimentos estrangeiros	336.745.3
4. Finanças privadas	336.7
5. Economia	33

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia						%
	1	2	3	4	5	Média
1	80	0	80	0	0	32
2	0	0	80	0	0	16
3	0	0	80	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
Media %	16	0	32	0	0	9,6

## Frases-chave

Níveis da cadeia				%
	1	2	3	Média
1	80	80	0	53,33
2	0	80	0	26,66
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
Media %	16	32	0	15,99

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	
2	X		
3	X		
4	X		
5	X		

TÍTULO: Diagnóstico das desigualdades de renda no Brasil

ÍNDICE EM CADEIA  
CABEÇALHOSNÚMERO DO  
ASSUNTO (CDU)

1. Distribuição de renda-Desenvolvimento regional - Brasil - Documentos de seminários	338.984.2:339(81)(063)
2. Distribuição de renda-Desenvolvimento regional - Brasil	338.984.2:339(81)
3. Distribuição de renda-Desenvolvimento regional	338.984.2:339
4. Desenvolvimento regional	338.984.2
5. Planos de desenvolvimento econômico	338.984

## SIMILARIDADE

## Palavras-chave

Níveis da cadeia	1	2	3	4	% Média
1	0	0	80	80	40
2	0	0	80	80	40
3	0	0	80	0	20
4	0	0	0	0	0
5	0	0	0	0	0
Media %	0	0	48	32	20

## ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1		X	A partir do nível 3
2		X	o cabeçalho perde
3	X		o conceito indicati
4	X		vo de lugar "Brasil"
5	X		e torna-se menos es
			pecifico do qu o
			título

TÍTULO: Ajuda ao desenvolvimento industrial regional nos países do Mercado Comum

ÍNDICE EM CADEIA CABEÇALHOS	NÚMERO DO ASSUNTO (CDU)
1. Planos de desenvolvimento econômico regionais Europa - Documentos de seminários	338.984.2(4)(063)
2. Planos de desenvolvimento econômico regionais Europa	338.984.2(4)
3. Planos de desenvolvimento econômico regionais	338.984.2
4. Planos de desenvolvimento econômico	338.984
5. Planificação econômica e controle	338.98

SIMILARIDADE  
Palavras-chave

Níveis da cadeia	1	2	3	4	5	6	7	% Média
1	0	80	0	80	0	0	0	22,85
2	0	80	0	80	0	0	0	22,85
3	0	80	0	80	0	0	0	22,85
4	0	80	0	80	0	0	0	11,42
5	0	0	0	0	0	0	0	0
Média %	0	64	0	48	0	0	0	15,99

Frases-chave

Níveis da cadeia	1	2	3	4	5	% Média
1	0	80	80	0	0	32
2	0	80	80	0	0	32
3	0	80	80	0	0	32
4	0	80	0	0	0	16
5	0	0	0	0	0	0
Média	0	64	48	0	0	22,4

ESPECIFICIDADE

Níveis da cadeia	KWOC	CADEIA	OBSERVAÇÕES
1	X	X	1) "Mercado Comum" é mais específico do que "Europa", mas cabeçalho contém o conceito "documentos de seminários", ausente do título.
2	X		
3	X		
4	X		
5	X		

1.1. Bibliografias das 100 obras

BRASIL: O PAIS E O GOVERNO	101
BRASIL: O PAIS E O GOVERNO	102
BRASIL: O PAIS E O GOVERNO	103
BRASIL: O PAIS E O GOVERNO	104
BRASIL: O PAIS E O GOVERNO	105
BRASIL: O PAIS E O GOVERNO	106
BRASIL: O PAIS E O GOVERNO	107
BRASIL: O PAIS E O GOVERNO	108
BRASIL: O PAIS E O GOVERNO	109
BRASIL: O PAIS E O GOVERNO	110
BRASIL: O PAIS E O GOVERNO	111
BRASIL: O PAIS E O GOVERNO	112
BRASIL: O PAIS E O GOVERNO	113
BRASIL: O PAIS E O GOVERNO	114
BRASIL: O PAIS E O GOVERNO	115
BRASIL: O PAIS E O GOVERNO	116
BRASIL: O PAIS E O GOVERNO	117
BRASIL: O PAIS E O GOVERNO	118
BRASIL: O PAIS E O GOVERNO	119
BRASIL: O PAIS E O GOVERNO	120
BRASIL: O PAIS E O GOVERNO	121
BRASIL: O PAIS E O GOVERNO	122
BRASIL: O PAIS E O GOVERNO	123
BRASIL: O PAIS E O GOVERNO	124
BRASIL: O PAIS E O GOVERNO	125
BRASIL: O PAIS E O GOVERNO	126
BRASIL: O PAIS E O GOVERNO	127
BRASIL: O PAIS E O GOVERNO	128
BRASIL: O PAIS E O GOVERNO	129
BRASIL: O PAIS E O GOVERNO	130
BRASIL: O PAIS E O GOVERNO	131
BRASIL: O PAIS E O GOVERNO	132
BRASIL: O PAIS E O GOVERNO	133
BRASIL: O PAIS E O GOVERNO	134
BRASIL: O PAIS E O GOVERNO	135
BRASIL: O PAIS E O GOVERNO	136
BRASIL: O PAIS E O GOVERNO	137
BRASIL: O PAIS E O GOVERNO	138
BRASIL: O PAIS E O GOVERNO	139
BRASIL: O PAIS E O GOVERNO	140
BRASIL: O PAIS E O GOVERNO	141
BRASIL: O PAIS E O GOVERNO	142
BRASIL: O PAIS E O GOVERNO	143
BRASIL: O PAIS E O GOVERNO	144
BRASIL: O PAIS E O GOVERNO	145
BRASIL: O PAIS E O GOVERNO	146
BRASIL: O PAIS E O GOVERNO	147
BRASIL: O PAIS E O GOVERNO	148
BRASIL: O PAIS E O GOVERNO	149
BRASIL: O PAIS E O GOVERNO	150

APÊNDICE 2

Amostra das listagens do Índice KWOC

## 2.1. Bibliografias dos 100 livros.

- 652 OOACAD.  
49/73.
- 653 ALLEN, C L. MORGNER, A. STROTZ, R H.  
PROBLEMAS SOBRE TEORIA DOS PREÇOS.  
SAO PAULO, PIDNEIRA, 1973.  
PREÇOS E VALOR.  
COMPRA.  
4/73.
- 655 MALTA, C P T. PIRAGIBE, F M.  
GUIA PRATICO TRABALHISTA (EX 2).  
3 ED RIO DE JANEIRO, EDICDES TRABALHISTAS, 1971.  
LEGISLAÇÃO E FISCALIZAÇÃO DO TRABALHO-BRASIL-GUIAS.  
OO  
312/73.
- 699 FIGUEIREDO N F. IPEA INPES.  
A TRANSFERENCIA DE TECNOLOGIA NO DESENVOLVIMENTO INDUSTRIAL  
DO BRASIL.  
RIO DE JANEIRO, 1972.  
TRANSFERENCIA DE TECNOLOGIA-INDUSTRIAS-BRASIL.  
DOACAD.  
434/73EX1 670/73EX2.
- 766 BRASIL LEIS DECRETOS ETC.  
COLEÇÃO DAS LEIS DE 1946.  
RIO DE JANEIRO, IMPRENSA NACIONAL, 1946.  
LEGISLAÇÃO GOVERNAMENTAL-BRASIL.  
COMPRA  
1873/74V1 1113/74V2 1874/74V3 1114/74V4 1876/74V6 1115/74V8  
1116/74V9.
- 768 BRASIL LEIS E DECRETOS ETC.  
COLEÇÃO DAS LEIS DE 1948.  
RIO DE JANEIRO, IMPRENSA NACIONAL, 1948.  
LEGISLAÇÃO GOVERNAMENTAL-BRASIL.  
COMPRA.  
1125/74V1 1126/74V2 1877/74V3 1127/74V4 1128/74V5 1129/74V6  
1878/74V7 1130/74V8.
- 782 BRASIL LEIS, DECRETOS, ETC.  
COLEÇÃO DAS LEIS DE 1967.  
RIO DE JANEIRO, IMPRENSA NACIONAL, 1967.  
LEGISLAÇÃO GOVERNAMENTAL-BRASIL.  
OO.  
1891/74V1 1892/74V2 1893/74V3 1894/74V4 1895/74V5 1236/74V6  
1237/74V7 1549/74V8.
- 826 CIBOTTI, R.  
LA INFRAESTRUCTURA EN LA PLANIFICACION DEL DESARROLLO.  
SANTIAGO, CEPAL/ILPES, 1971.  
PLANOS DE DESENVOLVIMENTO ECONOMICO-AMERICA LATINA.  
COMPRA.  
1105/74.

## 2.2. Palavras-chave dos títulos.

### DESENVOLVIMENTO

ORGANIZACAO ADMINISTRATIVA PARA O DESENVOLVIMENTO ECONOMICO.	208
BANCOS DE DESENVOLVIMENTO (MODELO INSTITUCIONAL).	211
MANUAL DE PROJETOS DE DESENVOLVIMENTO ECONOMICO.	224
CONTRIBUICAO AO DESENVOLVIMENTO DA AGROINDUSTRIA V 7:MANGA.	556
BANCO INTERAMERICANO DE DESENVOLVIMENTO (MANUAL DE TREINAMENTO ESTRUTURA, RECURSOS, OPERACOES).	569
AGRICULTURA E DESENVOLVIMENTO.	652
A TRANSFERENCIA DE TECNOLOGIA NO DESENVOLVIMENTO INDUSTRIAL DO BRASIL.	699
RELATORIO PRELIMINAR DE DESENVOLVIMENTO LOCAL INTEGRADO DO MUNICIPIO DE GRAO MONGOL.	833
RELATORIO PRELIMINAR DE DESENVOLVIMENTO INTEGRADO DO MUNICIPIO DE CODO.	850
RELATORIO PRELIMINAR DE DESENVOLVIMENTO INTEGRADO DO MUNICIPIO DE SOURE.	916
RELATORIO PRELIMINAR DE DESENVOLVIMENTO INTEGRADO DO MUNICIPIO DE PARINTINS.	925
RELATORIO PRELIMINAR DE DESENVOLVIMENTO INTEGRADO DO MUNICIPIO DE ARAGUAINA.	945
TEORIA E POLITICA DO DESENVOLVIMENTO ECONOMICO N 24.	1005
PLANO DECENAL DE DESENVOLVIMENTO ECONOMICO E SOCIAL ENERGIA V 1.	1008
RELATORIO PRELIMINAR DE DESENVOLVIMENTO INTEGRADO DO MUNICIPIO DE ESPLANADA.	1053
PLANO DE DESENVOLVIMENTO INTEGRADO DO MUNICIPIO DE TAUBATE.	1064
RELATORIO PRELIMINAR DE DESENVOLVIMENTO INTEGRADO DO MUNICIPIO DE VIGIA.	1214
INTEGRACAO REGIONAL E DESENVOLVIMENTO DO NORDESTE.	1228
RELATORIO PRELIMINAR DE DESENVOLVIMENTO INTEGRADO DO MUNICIPIO DE COARI.	1231
RELATORIO PRELIMINAR DE DESENVOLVIMENTO INTEGRADO DO MUNICIPIO DE CAMOCIM.	1236

### 2.3. Frases-chave dos títulos.

#### DEBATE

ENCARGOS TRABALHISTAS E ABSORCAO DE MAO-DE-OBRA (UMA INTERPRETACAO DO PROBLEMA E SEU DEBATE). 194

#### DECRETOS PROMULGADOS

LEIS E DECRETOS PROMULGADOS NO PERIODO 1965 (JANEIRO) 1967 (DEZEMBRO) VERSANDO SOBRE MATERIA DE INTERESSE ESPECIFICO DO DNER. 577

LEIS E DECRETOS PROMULGADOS NO PERIODO 1962 JANEIRO 1964 DEZEMBRO VERSANDO SOBRE MATERIA DE INTERESSE ORGANICO DO DNER. 609

#### DEFINICAO.

A EMPRESA PUBLICA NO DIREITO BRASILEIRO (PRINCIPIOS CONSTITUCIONAIS, DEFINICAO LEGAL). 14

#### DESARROLLO

LA INFRAESTRUCTURA EN LA PLANIFICACION DEL DESARROLLO. 826

#### DESENVOLVIMENTO

CONTRIBUICAO AO DESENVOLVIMENTO DA AGROINDUSTRIA V 7:MANGA. 556

BANCO INTERAMERICANO DE DESENVOLVIMENTO (MANUAL DE TREINAMENTO ESTRUTURA, RECURSOS, OPERACOES). 569

AGRICULTURA E DESENVOLVIMENTO. 652

INTEGRACAO REGIONAL E DESENVOLVIMENTO DO NORDESTE. 1228

A INDUSTRIA QUIMICA BRASILEIRA (LIMITACOES E PERSPECTIVAS AO SEU DESENVOLVIMENTO). 1339

#### DESENVOLVIMENTO ECONOMICO

ORGANIZACAO ADMINISTRATIVA PARA O DESENVOLVIMENTO ECONOMICO. 208

MANUAL DE PROJETOS DE DESENVOLVIMENTO ECONOMICO. 224

TEORIA E POLITICA DO DESENVOLVIMENTO ECONOMICO N 24. 1005

PLANO DECENAL DE DESENVOLVIMENTO ECONOMICO E SOCIAL ENERGIA V 1. 1008

#### DESENVOLVIMENTO INDUSTRIAL

A TRANSFERENCIA DE TECNOLOGIA NO DESENVOLVIMENTO INDUSTRIAL DO BRASIL. 699

APÊNDICE 3

Palavras-chave dos títulos (316)

Absorção 194*	b	BNH 1443*	a
Ação 127*230*	c	Brasil 149,379,699,879*,1139,1173,	
Acoplado 1140*	b	1288,1901,1939	a
Acre 1076*	a	Brasileira 399,427,1302,1339,1383*	a
Administracion 460		Brasileiro 14,1095	
Administrativa 208,595*	a	Brazil 565,132	
Administrativo 88*	a	Bridges 1296	
Afluyente 584*	b	Brumado 1239	
Agrícola 192,893		Cacaueira 67	
Agrícolas 460		Cacaueiro 67	
Agricultura 652,460		Camocim 1236	
Agroindustria 556		Capital 1901*	a
Ajuda 2011*	b	Cargos 856	
Alternative 8*	c	Catálogo 1383	
Amapá 1076*	a	CEF 1443*	a
Amazônia 1383		Centro 235*	b
América Latina 879*	a	Centros 1443*	b
Anais 393*	a	Change 61	
Analytical 118*	b	Chapadinha 1258	
Aplicação 234*	b	CIRD 235*	a
Apoio 1443*	b	Classificação 856	
Araguaina 945		Clientes 1429*	a
Areas 1076		Coari 1231	
Atividades 1173*	c	Codo 850	
Avaliação 485		Coleção 766*,768*,782*	c
Background 132*	b	Comercio 1381	
Bahia 67		Comparisons 118*	c
Banco 569		Computers 4	
Bancos 211		Comum 2011*	b
Base 127*	b	Comum 1457	
Bases 230*	c	Conceitos 1381*	c
Basica 1130*	b	Conditions 214*	c
Biaxiais 1143*	b	Conjunturais 500*	b

Bibliografia 399,427		Consolidação 112*	b
BNB 1443*	a	Constitucionais 14*	a
Constitucional 88*	a	Elaboração 485	
Contribuição 556*	c	Electronic 4*	a
Cooperação 149		Elementos 1173*	c
Countries 118*	c	Empresa 14	
Country 132*	c	Empresas 204,879*	a
Cultivados 893*	b	Encargos 194*	a
Culturas 893*	a	Energia 1008*	a
Curso 393		Ensaaios 1095*	c
Debate 194*	c	Ensino 579	
Decretos 577* a, 609* a		Especial 234*, 1076*	b
Definição 14*	c	Específico 577*	c
Definições 1381*	c	Esplanada 1053	
Desarrollo 826		Estado 1076*	a
Desenvolvimento 208,211,224, 556*,569 ,652*,699*,833* , 850*,916*,925*,945*,1005 , 1008,1053*,1064*,1214* , 1228,1231*,1236*,1239* , 1240*,1258*,1269*,1278* , 1339,2011	a	Estados 234*	b
		Estatística 1130,1381	
		Estatutos 857	
		Estratégia 879*	b
		Estrutura 569*	b
		Estudo 67*,511*	c
		Estudos 192*	c
Desigualdades 1939*	b	Europe 8,101	
Development 118		Exchanges 118*	a
Diagnostico 1076*,1939*	c	Exercício 234*	c
Diodo 1140		Exercícios 1139*	c
Dirección 460*	c	Explotaciones 460*	a
Directory 565		Export 565	
Direito 14,88		Exportação 353	
Discriminada 893*	c	Exposição 1383	
Distrito Federal 511		Exterior 1381	
DNER 577*,609*	a	Familiares 511*	a
Economic 118		Federativa 149*	c
Econômica 1139		FGTS 112*	a
Econômico 208,224,1005,1008		FINEP 1443*	a
Econômicos 500*	a	Fiscal 149*	c

Economics 61		Física 31* b 427	
Educação 979		Folha 1167*	a
Educacional 1429		Interpretação 194*	c
Education 97,118		Interpretada 1302*	c
Educational 8*	a	Issues 8*	c
Força 1076*	c	Lagarto 1240	
Fotocorrente 1143*	a	Latinoamericano 1457	
Fundações 204*	a	Legal 14*	a
Fundos 234*	a	Legislação 579	
Futures 8*	b	Leis 577,609,766,768,782	
Gaps 4*	b	Limitações 1339*	b
Ge-si 1143*	a	Local 393*,833*	b
Geologia 986*,1167*	a	Manga 556*	a
Government 626		Manual 224,569,1000	
Grandes 1443*	b	Mão-de-obra 194	
Grão Mongol 833		Mar 1023	
Grau 979*	b	Mascote 1167*	a
Guia 655,1302,1901		Mata 192*	a
Heterojunções 1143*	a	Matéria 609*,577*	c
História 1139		Mecânicas 1143*	a
Indicadores 100*	b	Member 118*	c
Industria 1339		Mercado 1457,2011*	a
Industrial 579,699,879*a 2011*a		Metas 230*	c
Influencia 1143*	b	Methods 8*	c
Infra-estrutura 1443*	a	Metropolitano 393*	a
Infraestructura 826*	a	Microlinha 1140*	a
Innovation 118*a,214,626		Minas Gerais 192	
Institucional 211*	a	Modelo 211*,1095*	c
Integra 2*	a	Mogiquiçaba 986*	a
Integração 1228*	a	Município 80,833,850,916,925,945,	
Integrado 235*,833*,850*,916*,		1053,1064,1214,1231,1236,1239,	
925*,945*,1053*,1064*,1214*,		1240,1258,1269,1278	
1231*,1236*,1239*,1240*,1258*,		Municípios 893*	a
1269*,1278*	a	Nacional 80*,149*	b
Interamericano 569*	b	Não-tradicionais 353*	b
Interesse 577*,609*	c	Nazionale 506	

Interinvest 1901*	c	Nordeste 986*a,1228,1443	
Internacionais 879		Noroeste 1167*	a
Internacional 1901*	b	Oceanografia 31	
International 118*	b	Princípios 14*	b
OEA 149*	a	Problema 194*	b
Operações 569*	b	Problemas 653*	b
Orçamentos 511*	a	Processo 379*	c
Orgânico 609*	c	Produção 893	
Organização 80*,208*	a	Produtos 353*,893	a
Organizacion 460*	a	Programa 127,149*,1429*,1443*	b
Oscilador 1140		Programming 132	
Países 2011*	c	Projeto 1076*	a
Paper 132*	c	Projetos 224,485	
Parintins 925		Promulgados 577*,609*	b
Participação 234*	b	Publica 14	
Patrimônio 1000		Publicas 204	
Permanentes 893*	b	Publico 88,856	
Perspectivas 1339*	c	Quadriculas 986*	a
Pesquisas 379		Quimica 1339	
Piano 506		Radicular 67*	a
Planejamento 393,1173*	b	Recursos 569*	b
Planificação 1130		Reforma 595*	b
Planificación 826		Reformas 127*	b
Plano 234,1008,1064*	b	Região 67*,192*	b
Policy 8		Regional 1228,2011	
Política 1005		Relativa 879*	c
Político 1095		Relatório 833*,850*,916*,925*,1053*,	
Porto Seguro 986*	a	1214*,1231*,1236*,1239*,1240*	,
Posição 879*	c	1258*,1269*,1278*,	b
Pratico 655*,1302*	b	Relevance 8*	b
Preços 653		Renda 1939	
Prejulgados 2*	a	Report 118*	b
Preliminar 833*,850*,916*,925*,		República 149*	c
945*,1053*,1076*,1214*,1231*,		Research 118	
1236*,1239*,1240*,1258*,1269*,		Rio-Doce 235*	a
1278*,	b	Rondonia 1076*	a
Previdência 1302		Roraima 1076*	a

Primários 353*	b	São Paulo 393*	a
Serviço 856		São Raimundo Nonato 1278	
Servidor 88		TST 2*	a
Sistema 67*	b	Tunel 1140*	b
Sociais 857*	b	Unidades 511*	b
Social 1008*b,1302		United Kingdom 97	
Sociedade 584		United States 8	
Solos 67		Urbanas 1076*	a
Soure 916		Urbano 393	
STF 2*	a	Urbanos 1443*	a
Success 214*	c	Versando 577*,609*	c
SUDENE 900		Vigia 1214	
Sul 67*	b	Zona 192*	b
Súmulas 2*	b	Zoo'ogia 399	
Sviluppo 506		Zootecnica 1173	
Tarefa 1076*	c	* valor 0% recebido na comparação da palavra com o cabeçalho de nível 1 da cadeia. Total 199 - palavras (62,97% do total).	
Taubaté 1064			
Technical 97,626			
Technological 61,118,214			
Technologique 101			
Technology 4*	a	a	palavras que receberam o valor
Técnica 149,460*	a	0%, mas que eram realmente significativas no caso dado. Total 85 palavras (26,89% do total)	
Tecnologia 699,979			
Tensões 1143*	a		
Teoria 653*,1005*	b		
Territorial 1023*	a	b	palavras que receberam o valor
Territórios 1076*	a	0%, mas que seriam frases-chave quando coordenadas com outras - palavras do mesmo título. Total 66 palavras (20,88% do total)	
Tiangua 1269			
Tipos 67*	b		
Trabalhista 655			
Trabalhistas 194*	a		
Transferência 699			
Transmissão 1140*	a	c	palavras supérfluas, sem significação. Total 48 palavras -
Treinamento 569*	a	(24,12% do total).	

#### APÊNDICE 4

Relação das palavras e frases-chave dos títulos resultantes do índice KWOC reformado.

Total de frases: 99 e palavras: 163 (Total: 262)

##### Frases-chave

##### Uso generalizado

Absorção de mão de obra	80	Indicadores econômicos	0a
Analytical report	0a	Indústria química	80
Áreas urbanas	70	Mar territorial	80
Avaliação de projetos	80	Mercado Comum	0a
Bancos de desenvolvimento	100	Mercado Comum	80
Capital internacional	0a	Microlinha de transmissão	0a
Centros urbanos	0a	Oceanografia física	80
Classificação de cargos	80	Participação dos estados	0a
Comércio exterior	80	Planejamento metropolitano	80
Cooperação técnica	80	Planejamento urbano	80
Culturas permanentes	0a	Previdência social	80
Desenvolvimento econômico	80,100	Produção agrícola	80
Desenvolvimento industrial	80	Produtos cultivados	80
Desenvolvimento integrado	0a	Produtos primários	0a
Desenvolvimento local	0a	Reformas de base	0a
Diodo tunel	80	Research & Development	80
Direito administrativo	80	Serviço público	80
Direito constitucional	80	Servidor público	80
Economic exchanges	80	Sistema radicular	0a
Electronic exchanges	80	Sociedade afluyente	80
Empresa pública	80	Technical education	80
Empresas internacionais	80	Technical innovation	80
Empresas públicas	80	Technological change	80
Encargos trabalhistas	0a	Technological innovation	80
Ensino industrial	80	Tensões mecânicas	0a
Estatística básica	80	Teoria dos preços	80
Estatutos sociais	80	Tipos de solos	80
Explotaciones agrícolas	80	Transferencia de tecnologia	80
Guia prático	80	Zona da mata	0a
História econômica	80		

Frases-chave  
Formadas para a coleção

Background paper	0a	Prejulgados na íntegra	0a
Banco interamericano	80	Produção dos municípios	80
Catálogo da exposição	80	Programa de ação	70
Centro integrado	0c	Programa de apêlo	0a
Conceitos e definições	0c	Programa educacional	0c
Conditions for success	0c	Programa nacional	0a
Decretos promulgados	0a	Projeto especial	0a
Diagnóstico preliminar	0c	Relatório preliminar	0c
Educational futures	80	República Federativa	0c
Export directory	80	Versando sobre matéria	0c
Força tarefa	0c		
Guia Interinvest	80		
Integração regional	80		
Interesse específico	0c		
Interesse orgânico	0c		
Interpretação do problema	0c		
Limitações e perspectivas	0c		
Manual de treinamento	80		
Mascote noroeste	0a		
Member countries	0c		
Metas e bases	0c		
Modelo institucional	0c		
Organização nacional	0a		
Pesquisas em processo	80		
Piano nazionale	80		
Planejamento das atividades	0a		
Plano de aplicação	80		
Policy relevance	80		
Porto Seguro Nordeste	0a		
Posição relativa	0c		

Palavras-chave

Ação	0c	Coari	80
<u>Acoplado</u>	0b	Codó	80
Acre	0a	Coleção	0c
Administración	80	Comparisons	0c
Administrativa	0,80a	Conjunturais	0b
Agrícola	80	Consolidação	0b
Agricultura	80	Constitucionais	0a
Agroindustria	80	Contribuição	0c
Ajuda	0b	Country	0c
Alternative	0c	Curso	70
Amapá	0a	Debate	0c
Amazônia	80	Definição	0c
America Latina	0a	Desarrollo	80
Anais	0a	Desenvolvimento	80,0a
Bahia	80	Desigualdades	0b
Biaxiais	0b	Diagnóstico	0c
Bibliografia	80	Dirección	0c
BNB	0a	Direito	80
BNH	0a	Discriminada	0c
Brasil	80,0a	Distrito Federal	80
Brasileira	80,0a	DNER	0a
Brasileiro	80	Economics	80
Brazil	80	Educação	70
Bridges	100	Education	80
Brumado	80	Elaboração	80
Cacaueira	80	Elementos	0c
Cacaueiro	80	Energia	0a
Camocim	80	Ensaio	0c
CEF	0a	Especial	0b
Chapadinha	80	Esplanada	80
CIRD	0a	Estado	0a
Clientes	0a	Estatística	80

Palavras-chave

Estratégia	0b	Latinoamericano	80	Reforma	0b
Estrutura	0b	Legal	0a	Região	0b
Estudo	0c	Legislação	80	Regional	80
Estudos	0c	Leis	80	Renda	80
Europe	80	Local	0b	Rio Doce	0a
Exercício	0c	Manga	0a	Rondônia	0a
Exercícios	0c	Manual	80	Roraima	0a
Exportação	80	Methods	0c	São Paulo	0a
Familiares	0a	Minas Gerais	80	Social	0b
FGTS	0a	Modelo	0c	Soure	80
FINEP	0a	Mogiçuçaba	0a	STF	0a
Fiscal	0c	Município	80	SUDENE	100
Física	80	Não-tradicionais	0b	Sul	0b
Folha	0a	Nordeste	80	Súmulas	0b
Foto-corrente	0a	OEA	0a	Sviluppo	80
Fundações	0a	Operações	0b	Taubaté	80
Fundos	0a	Orçamentos	0a	Technologique	80
Gaps	0b	Organização	0a	Technology	0a
Geologia	0a	Organización	0a	Técnica	0a
GeSi	0a	Oscilador	80	Tecnologia	80
Government	80	Países	0c	Teoria	0b
Grandes	0b	Parintins	80	Territórios	0a
Grão Mongol	80	Patrimônio	80	Tianguá	80
Grau	0b	Planificação	80	Trabalhista	80
Heterojunções	0a	Planificación	80	TST	0a
Industrial	0a	Plano	80,0b	Unidades	0b
Influencia	0b	Política	80	United Kingdom	80
Infra-estrutura	0a	Político	80	United States	80
Infraestructura	0a	Principios	0b	Zoologia	80
Integrado	0a	Problemas	0b	Zootecnia	80
International	0b	Programming	70		
Interpretada	0c	Projetos	80		
Issues	0c	Quadrículas	0a		
Lagarto	80	Recursos	0b		

\* os números (0,80,100,70) ao lado das palavras e conceitos representam os valores de similaridade obtidos no nível 1 da cadeia. (Total de palavras e frases de valor 0% = 142, isto é, 54,19% do total)

a palavras e frases que receberam o valor 0% mas que eram realmente significativos no caso dado. Total: 74 (28,24% do total).

b palavras que receberam o valor 0% mas que seriam frases-chave quando coordenadas com outras palavras do mesmo título. Total: 28 (10,68%)

c frases e palavras sem significação. Total 40 (15,26%).

01	Historiografia, Antropologia		02	Geografia, Climatologia
02	Geografia, Climatologia		03	Geografia, Climatologia
03	Geografia, Climatologia		04	Geografia, Climatologia
04	Geografia, Climatologia		05	Geografia, Climatologia
05	Geografia, Climatologia		06	Geografia, Climatologia
06	Geografia, Climatologia		07	Geografia, Climatologia
07	Geografia, Climatologia		08	Geografia, Climatologia
08	Geografia, Climatologia		09	Geografia, Climatologia
09	Geografia, Climatologia		10	Geografia, Climatologia
10	Geografia, Climatologia		11	Geografia, Climatologia
11	Geografia, Climatologia		12	Geografia, Climatologia
12	Geografia, Climatologia		13	Geografia, Climatologia
13	Geografia, Climatologia		14	Geografia, Climatologia
14	Geografia, Climatologia		15	Geografia, Climatologia
15	Geografia, Climatologia		16	Geografia, Climatologia
16	Geografia, Climatologia		17	Geografia, Climatologia
17	Geografia, Climatologia		18	Geografia, Climatologia
18	Geografia, Climatologia		19	Geografia, Climatologia
19	Geografia, Climatologia		20	Geografia, Climatologia
20	Geografia, Climatologia		21	Geografia, Climatologia
21	Geografia, Climatologia		22	Geografia, Climatologia
22	Geografia, Climatologia		23	Geografia, Climatologia
23	Geografia, Climatologia		24	Geografia, Climatologia
24	Geografia, Climatologia		25	Geografia, Climatologia
25	Geografia, Climatologia		26	Geografia, Climatologia
26	Geografia, Climatologia		27	Geografia, Climatologia
27	Geografia, Climatologia		28	Geografia, Climatologia
28	Geografia, Climatologia		29	Geografia, Climatologia
29	Geografia, Climatologia		30	Geografia, Climatologia
30	Geografia, Climatologia		31	Geografia, Climatologia
31	Geografia, Climatologia		32	Geografia, Climatologia
32	Geografia, Climatologia		33	Geografia, Climatologia
33	Geografia, Climatologia		34	Geografia, Climatologia
34	Geografia, Climatologia		35	Geografia, Climatologia
35	Geografia, Climatologia		36	Geografia, Climatologia
36	Geografia, Climatologia		37	Geografia, Climatologia
37	Geografia, Climatologia		38	Geografia, Climatologia
38	Geografia, Climatologia		39	Geografia, Climatologia
39	Geografia, Climatologia		40	Geografia, Climatologia

APÊNDICE 5

Quadro das Divisões Principais

C D U

0	GENERALIDADES	3	CIÊNCIAS SOCIAIS
00	Prolegômenos. Fundamentos gerais do conhecimento e da cultura.	30	Sociologia. Sociografia.
01	Bibliografia. Catálogos.	31	Estatística.
02	Bibliotecas. Biblioteconomia.	32	Ciência política. Política.
03	Enciclopédias gerais. Dicionários. Livros de referência.	33	Economia política. Economia.
04	Coleções de ensaios, separatas, folhetos, opúsculos.	34	Direito. Legislação. Jurisprudência.
05	Publicações periódicas, revistas gerais.	35	Administração pública. Direito administrativo. Ciência militar. Defesa.
06	Instituições. Associações. Congressos. Museus.	36	Assistência e socorro social. Seguros.
07	Jornais. Jornalismo.	37	Educação.
08	Poligrafias. Trabalhos correctivos. Vária	38	Comércio. Comunicações.
09	Manuscritos. Obras raras e notáveis.	39	Etnografia. Costumes e tradições. Folclore. Antropologia social e cultural.
1	FILOSOFIA. METAFÍSICA. PSICOLOGIA. LÓGICA. ÉTICA.	4	FILOLOGIA. LINGUÍSTICA.
11	Metafísica.	41	Filologia e linguística gerais.
13	Filosofia do espírito. Metafísica da vida espiritual.	42	Línguas ocidentais em geral.
14	Sistemas filosóficos.	420	Inglês.
159.9	Psicologia.	43	Línguas germânicas.
16	Lógica. Teoria do conhecimento. Metodologia.	44/46	Línguas românicas ou neolatinas.
17	Ética. Moralidade. Convenção.	440	Francês.
18.01	Estética.	450	Italiano.
19	História da Filosofia.	460	Espanhol.
2	RELIGIÃO. TEOLOGIA.	469.0	Português.
21	Teologia natural.	47	Línguas clássicas. Latim e grego.
22	Sagrada Escritura. A Bíblia.	48	Línguas eslavas e bálticas.
23	Teologia dogmática.	49	Línguas orientais, africanas e outras.
24	Teologia moral. Prática religiosa.	5	CIÊNCIAS PURAS
25	Teologia pastoral.	51	Matemática.
26	A Igreja Cristã em geral.	52	Astronomia. Levantamentos topográficos. Geodesia.
27	História geral da Igreja Cristã.	53	Física.
28	Igrejas e seitas cristãs.	54	Química. Cristalografia. Mineralogia.
29	Religiões e cultos não cristãos.	55	Geologia e ciências afins. Meteorologia.
		56	Paleontologia.
		57	Antropologia. Ciências biológicas.
		58	Botânica.
		59	Zoologia.

6	CIÊNCIAS APLICADAS. MEDICINA. TECNOLOGIA.	8	LITERATURA.
61	Ciências médicas. Saúde. Segurança.	8.0	Ciência e técnica literárias.
62	Engenharia e Tecnologia em geral.	820	Literatura inglesa.
63	Agricultura. Silvicultura. Zootecnia. Produtos de origem animal. Caça e pesca.	830	Literatura alemã.
64	Ciências domésticas. Economia doméstica.	840	Literatura francesa.
65	Administração e organização da indústria, comércio e transportes.	850	Literatura italiana.
66	Indústrias químicas. Tecnologia química.	860	Literatura espanhola.
67	Indústrias e profissões diversas.	869.0	Literatura portuguesa.
68	Ofícios, artes e indústrias especializados.	869.0(81)	Literatura brasileira.
69	Indústria de construção, materiais, profissões, construção.	87	Literatura clássica (latina e grega).
7	BELAS-ARTES. DIVERTIMENTOS. DESPORTOS.	88	Literatura eslava. Literatura báltica.
71	Urbanização. Planejamento. Arquitectura paisagística, etc.	89	Literatura oriental, africana e outras.
72	Arquitectura.	9	GEOGRAFIA. BIOGRAFIA. HISTÓRIA.
73	Escultura e artes afins.	91	Geografia, explorações, viagens.
74	Desenho. Artes menores.	92	Biografia.
75	Pintura.	93	História em geral. Fontes. História antiga.
76	Arte da gravura. Gravuras.	94	História medieval e moderna.
77	Fotografia e cinematografia.	940	História da Europa.
78	Música.	946.9	História de Portugal.
79	Divertimentos. Passatempos. Jogos. Desportos.	950	História da Ásia.
		960	História da África.
		970	História da América do Norte.
		980	História da América do Sul.
		981	História do Brasil.
		990	História da Oceania e das regiões.

## A B S T R A C T

SOUSA, G.H.B.P.de - Comparison between a KWOC Index (Keyword-out-of-Context) and a Chain Index derived from the UDC (Universal Decimal Classification), Rio de Janeiro, 1975.  
p. Thesis.

The two indexes systems applied in the Studies & Projects Financing Agency (FINEP) were compared: a KWOC Index and a Chain Index derived from the UDC. A second KWOC Index (built with the combination of each two consecutive keywords of the titles) was also compared to the Chain Index.

Two distinct methods were utilized in this comparison: 1) similarity measure of the indexes, where the level 1 heading of the Chain Index was considered the pattern of excellence and the titles compared to it, having each keyword or key-phrase of the title received a value of the scale 100%, 90%, 80%, 70% e 0% according to the higher or lower level of similarity; 2) specificity measure of the indexes where the number of specifying terms of each of them respective to the analyzed work was compared.

Relating to the similarity measure, the overall average (titles compared to the 1st five heading of the Chain Index) obtained was 19,03% for the keywords of the titles and 22,61% for their key-phrases. The overall average obtained with the comparison of the titles to the level 1 headings of the Chain, (the pattern headings), was 38,85 % (keywords) and 47,41% (key-phrases). All these averages confirmed a non-coincidence of vocabulary of the indexes, and an analysis of the lowest level key-terms (value 0%) revealed that in a great number they were significant terms as indexers, which accounted for the titles as indexers.

In the second method it was verified that the analyzed Chain Index was more specific than the key-terms of the titles.

It was concluded that it was not possible to generalize the results, but the methods, and that any indexing system can be handled (to a certain extent), to better fit its users needs.