

AUTOMAÇÃO DOS ÍNDICES DAS TABELAS DA
CLASSIFICAÇÃO DECIMAL UNIVERSAL

Elvia de Andrade Oliveira
Instituto Brasileiro de
Bibliografia e Documenta-
ção

Dissertação apresentada ao Instituto
Brasileiro de Bibliografia e Documen-
tação/Universidade Federal do Rio de
Janeiro para obtenção do grau de Ma
stre em Biblioteconomia e Documenta-
ção.

Orientador: Prof. Georges Schwachheim
do Centro Brasileiro de Pesquisas F
isicas

Rio de Janeiro
1973

SUMÁRIO

0 - INTRODUÇÃO	3
1 - HIPÓTESE	6
2 - MATERIAL	7
3 - MÉTODO	8
3.1 - ELVCDU1	8
3.2 - ELVCDU2	10
3.3 - ELVCDU3	11
4 - RESULTADOS	20
5 - CONCLUSÕES	21
6 - CITAÇÕES BIBLIOGRÁFICAS	22
7 - ANEXOS	23

0 - INTRODUÇÃO

A Classificação Decimal Universal (CDU), é uma tabela elaborada e editada pela Federação Internacional de Documentação (FID), com sede em Haia, Holanda, tendo por objetivo a sistematização do conhecimento humano expresso em números decimais.

Baseou-se na "Decimal Classification", de Melvil Dewey, agrupando em 10 classes subdivididas em divisões e seções, conforme as exigências de cada assunto.

Possui tabelas auxiliares para forma, local, língua, raça, subdivisão analítica e de ponto de vista, dando assim, maior flexibilidade ao sistema. Sua notação é mista, composta de números, letras e símbolos.

A CDU é utilizada em bibliotecas, centros de documentação etc., para classificar quaisquer tipos de documentos e materiais especiais.

É publicada em edições abreviadas e médias (abrangendo as 10 classes) e detalhada (contendo uma ou mais classes ou divisões).

Sua atualização e revisão são feitas, periodicamente, através das "Extensions and Corrections of the UDC".

As traduções, revisões e publicações em diversos idiomas são feitas por Comissões especiais, como a IBBO/COU, encarregada das tabelas em língua portuguesa, juntamente com a CP/CDU, Comissão Portuguesa da Classificação Decimal Universal.

Para a publicação das tabelas é importante a elaboração dos índices, até então feitos por processos manuais, onde os termos representativos dos assuntos são escolhidos e transcritos para fichas, juntamente com o número de classificação a eles correspondentes. Essas fichas são então ordenadas alfabeticamente para depois serem datilografadas em folhas, dependendo ainda de revisões para a uniformização dos termos. Os originais são encaminhados à gráfica para composição tipográfica e posterior revisão.

Esta forma morosa de elaboração dos índices, com a participação de várias pessoas, torna difícil a uniformização necessária em índices em linguagem artificial. A Federação Internacional de Documentação, recentemente, baixou instruções severas no sentido de não permitir a publicação de tabelas sem os respectivos índices.

Esses motivos nos lavaram ao interesse pelo assunto e o estudo da literatura nos mostrou que a maior parte das pesquisas sobre a automação da CDU, se refere a sistemas de recuperação da informação. No que se refere à automação de índices, estudos foram realizados por Malcolm Rigby (4), baseados no índice KWIC.

Foram então realizadas pesquisas que resultaram em três sistemas analisados, programados, testados, processados e apresentados por nós nesta dissertação para o processamento automático de índices das tabelas da CDU, a partir do texto, em linguagem natural.

Os dois primeiros sistemas (ELVCDU1 e ELVCDU2) apresentados neste trabalho forneceram experiência para a criação do terceiro sistema (ELVCDU3), considerado satisfatório, permitindo indexação inteiramente automática.

Com este trabalho procuramos dar nossa contribuição para a elaboração e atualização mais rápida de índices e tabelas da CDU através de processos automáticos e, com os resultados obtidos, fornecer subsídios para a preparação de novos textos traduzidos.

Para o desenvolvimento desta dissertação, contamos com:

i) experiências anteriores com as linguagens: de máquina, SPS-II, SNOBOL-III do computador IBM 1620 mod. II;

ii) aprendizado e treinamento com a linguagem PL/I para a análise, elaboração e processamento dos programas;

iii) aprendizado e treinamento em "Job Control Language" do Sistema Operacional OS/VS1 e em programa SORT/MERGE, do computador IBM/370 mod. 145;

iv) a eficiente orientação do Prof. Georges Schwachheim e a colaboração de toda sua equipe;

v) a compreensão do IBBO, na pessoa de sua Presidente, Profa. Heger Espanha Gomes, cedendo horas necessárias ao desenvolvimento dos estudos, sem nunca faltar incentivos aos nossos trabalhos;

vi) a colaboração do Centro Brasileiro de Pesquisas Físicas admitindo nosso treinamento em seu equipamento;

vii) a Comissão IBBO/CDU, pela aceitação destes estudos.

a sua anterior Presidente Irene de Menezes Dória pela autorização da utilização das tabelas da edição média e aos membros da Guanabara nas pessoas de seu atual Presidente Francisco Figueiredo L. de Albuquerque, Maria Emilia de Mello Cunha, Maria Helena G. de Paiva e Maria Thereza G. F. de Albuquerque, pela revisão da tabela 53 (FÍSICA) que foi utilizada para os testes;

viii) a Profa. Celia Ribeiro Zaher, por permitir, quando Presidente do IBBD, nossos estudos e treinamento em computador e também a iniciação deste trabalho;

ix) os Profs. Harold Borko e Jack Mills por sugestões aos dois primeiros estudos deste trabalho;

x) os colegas do Serviço de Bibliografia que muito me incentivaram e também os colegas do IBBD que, direta ou indiretamente, colaboraram para a realização deste trabalho;

xi) meu marido, Armando de Oliveira Filho, que com muita dedicação, acompanhou todas as fases desta dissertação.

1 - HIPÓTESE

Tendo em vista a dificuldade que apresenta a elaboração dos índices das tabelas de COU por processo tradicional, foi levantada a hipótese da viabilidade da elaboração automática das ses índices partindo das tabelas.

2 - MATERIAL

2.1 - Equipamento IBM/370 mod. 145

2.1.1 - Sistema operacional OS/VS1

2.1.2 - Unidade de disco 2314/2313

2.1.2.1 - Disco magnético 2313

2.1.3 - Leitora/Perfuradora de cartões, 2540

2.1.3.1 - Cartão

2.1.4 - Impressora 1403

2.2 - Equipamento periférico

2.2.1 - Perfuradora de Cartões IBM 023

2.3 - Publicações

2.3.1 - Classificação Decimal Universal

2.3.1.1 - Edição abreviada portuguesa

2.3.1.2 - Edição completa espanhola

2.3.1.3 - Edição média

2.3.1.4 - Edição trilingüe

3 - MÉTODO

O estudo dos três sistemas seguiu a um método geral:

a) a escolha da linguagem PL/I (Programming Language/One), deve-se ao fato de ser uma linguagem adequada para este trabalho, por permitir programação fácil com rápida análise e processamento da informação;

b) a escolha das tabelas da CDU para testes foi feita a partir de análise, procurando exemplos mais complexos para os testes a serem desenvolvidos. Em todas as etapas deste trabalho, considerou-se sempre a experiência obtida em etapas anteriores;

c) formato de entrada de dados em cartões:

- 1 - A classificação é citada a partir da coluna 1, ocupando até 25 (vinte e cinco) colunas;
- 2 - O elemento separador da classificação e do texto é 2 (dois) espaços;
- 3 - O texto ou assunto é perfurado em cartão a seguir dos 2 (dois) espaços após a classificação até a coluna 80 (oitenta) inclusive. Previu-se o uso de mais de um cartão, quando necessário, desde que se repita o número da classificação e se mantenham 2 (dois) espaços separadores;
- 4 - Todas as notas são colocadas após o texto, precedidas de 2 (dois) espaços;
- 5 - As remissivas são perfuradas com um asterisco (*) na coluna 1 (um).

d) A impressão da tabela que gerou o índice, por tornas o trabalho mais completo.

Cada um dos três estudos teve método próprio como passaremos a descrever:

3.1 - ELVCDU1 - 1º Estudo. Processamento de índices com marcação do texto da tabela.

3.1.1 - A parte escolhida para testes preliminares foi retirada da classe 3 (CIÊNCIAS SOCIAIS) que, após estudos, foi considerada como base para os testes.

3.1.2 - Descrição do programa. Leitura de cada cartão unindo os

textos com a mesma classificação. Captação de cada frase. Preparação da área de saída juntamente com a classificação correspondente. Gravação dessas áreas para posterior ordenação alfabética. Listagem dessas áreas formando assim o índice. Listagem da tabela.

O programa foi escrito e testado e aos primeiros testes surgiram os primeiros problemas: muitos termos vagos sem nenhuma subordinação aparente. Resolveu-se assinalar os termos gerais ao perfurar os cartões com o sinal maior que (>) e, quando algum termo vago necessitasse alguma explicação assinalava-se com o sinal menor que (<) conseguindo-se então, o seguinte resultado:

Cartões perfurados:

347,79 >DIREITO MARÍTIMO.
 347,794 <ALUGUEL. <FRETAMENTO.
 347.795.4 <MERCADORIAS.
 347.795.6 <REBOQUE.

no índice:

ALUGUEL (DIREITO MARÍTIMO) 347.794
 DIREITO MARÍTIMO 347.79
 FRETAMENTO (DIREITO MARÍTIMO) 347.794
 MERCADORIAS (DIREITO MARÍTIMO) 347.795.4
 REBOQUE (DIREITO MARÍTIMO) 347.795.6

Houve necessidade de outras associações mais internas que obedeceram às seguintes codificações:

347.791 INAVIOS. 2COMPRA. 2REGISTRO.

onde os termos assinalados com 2 (dois) seriam subordinados aos termos assinalados com 1 (um). Houve o problema do termo NAVIOS necessitar ser subordinado a DIREITO MARÍTIMO, resolveu-se então substituir o neste caso, por um asterisco garantindo coordenações duplas:

347.791 *NAVIOS. 2COMPRA. 2REGISTRO.

no índice:

ALUGUEL (DIREITO MARÍTIMO) 347.794
 COMPRA (NAVIOS) 347.791
 DIREITO MARÍTIMO 347.79
 FRETAMENTO (DIREITO MARÍTIMO) 347.794
 MERCADORIAS (DIREITO MARÍTIMO) 347.795.4
 NAVIOS (DIREITO MARÍTIMO) 347.791
 RESOQUE (DIREITO MARÍTIMO) 347.795.6
 REGISTRO (NAVIOS) 347.791

No entanto os sinais (>), (1) e (*) anulariam os sinais anteriores iguais. Qualquer assunto subordinado a DIREITO MARÍTIMO que fosse precedido dos sinais (<) ou (*), a ele seria acrescentado (DIREITO MARÍTIMO).

Previu-se também a exclusão de frases.

A frase iniciada com o sinal (%) não seria considerada para o índice a não ser que diante de um dos termos da frase, aparecesse um hífen (-), onde teríamos a inversão da mesma:

331.6 DESEMPREGO. %FALTA DE TRABALHO.
 341.2 %PESSOAS E COISAS EM -DIREITO INTERNACIONAL

No índice teríamos:

DESEMPREGO 331.6
 DIREITO INTERNACIONAL, PESSOAS E COISAS EM 341.2

3.1.3 - Discussão. Apesar das listagens do índice apresentarem resultados satisfatórios como produto final, este estudo (Anexo 7.1) foi considerado insatisfatório como sistema, pois exigiria a marcação de todo o texto da tabela com sinais pré-estabelecidos. Este sistema além de tornar o trabalho moroso, não aproveita o máximo da capacidade do computador, tornando o índice semi-automático e não automático como nos propusemos nesta dissertação. Este estudo foi então abandonado.

3.2 - ELVODU2 - 2º Estudo. Aplicação do índice KWIC (Key-words in context) na automação dos índices das tabelas da CDU.

3.2.1 - Escolha da tabela para teste. A parte escolhida foi 53 (AGRICULTURA), por conter texto considerado básico para os testes.

3.2.2 - Foi estabelecido um vocabulário de palavras não significativas baseado no texto, visando a exclusão dessas palavras na ordem alfabética dos termos importantes.

3.2.3 - O formato de saída foi de linhas com 112 posições para impressão.

3.2.4 - Descrição do programa. Leitura do vocabulário de palavras não significativas. Leitura de cartões garantindo o texto completo toda a vez que este fosse perfurado em mais de um cartão. Análise do texto com separação de cada palavra para pesquisa ao vocabulário. Preparação da área de saída citando o texto completo posicionando centralmente cada palavra significativa. Gravação de cada uma dessas áreas de saída para posterior ordenação alfabética. Listagem final do índice. Os cartões de entrada dos dados foram listados e medida que fossem lidos formando então a tabela. As setas (\Rightarrow) e (\Leftarrow) no índice, indicam início e fim de frase, respectivamente.

3.2.5 - Discussão. Este estudo (Anexo 7.2) mostrou a possibilidade de do emprego do índice KWIC na formação dos índices das tabelas da CDU, muito satisfatoriamente, pois o usuário teria cada termo significativo em ordem alfabética juntamente com todo o texto referente à classificação para um fácil reconhecimento da classificação desejada. O índice KWIC já havia sido empregado por Malcolm Rigby (4). Entretanto, consideramos que novos métodos deveriam ser tentados até que se obtivesse uma listagem formato tanto quanto possível semelhante aos índices relativos convencionais, pois este contaria, certamente, com melhor aceitação do usuário.

3.3 - ELVCDU3. O processamento dos índices das tabelas da CDU em linguagem natural, onde o trabalho manual se limita à perfuração do texto das tabelas em cartões, obtendo do computador, através de programas, um índice automático pronto para impressão, juntamente com o texto da tabela. Sua correção e atualização é feita através de troca ou inserção de novos cartões, levando ao computador para um novo processamento.

3.3.1 - A parte da tabela inicialmente escolhida foi 63 (AGRICULTURA), da edição abreviada portuguesa, de onde foram tirados exemplos mais complexos para a formação do índice. O programa foi escrito e testado. As tabelas 53 FÍSICA e 64 CIÊNCIAS DOMÉSTICAS da edição média, foram testadas tendo obtido resultados positivos. Parte da classe 3 CIÊNCIAS SOCIAIS das edições: abreviada portuguesa, completa, espanhola e edição trilingüe foi testada para as diversas línguas.

3.3.2 - Formação do índice. O texto referente à cada classificação é analisado. Seus componentes (palavras e símbolos) são separados e a cada um deles é dado um nível conforme sua função gramatical. Ao completar uma frase ou sentença, inicia-se a criação de novas frases ou sentenças. As frases são rotadas para colocar em evidência cada um de seus termos significativos.

3.3.2.1 - Como regra geral o ponto interrompe a frase. Os termos são distribuídos um após outro para a formação de novas frases.

63 Agricultura. Silvicultura. Zootecnia. Produtos de origem animal. Caça. Pesca.

Frases formadas para o índice

Agricultura 63

Silvicultura 63

Zootecnia 63

Produtos de origem animal 63

Origem animal, Produtos de 63

Animal, Produtos de origem 63

Caça 63

Pesca 63

3.3.2.2 - Diversas diretrizes foram estabelecidas para a formação de novas frases conforme a ordem dos elementos na frase original.

a) A vírgula (,) ou o E, mesmo quando seguidos de artigos e/ou preposições, levam a construção de mais de uma frase.

631.33 Máquinas de semear, plantar etc.

Frases formadas para o índice

Máquinas de semear 631.33
 Semear, Máquinas de 631.33
 Máquinas de plantar 631.33
 Plantar, Máquinas de 631.33

632.951 Produtos para a destruição dos artrópodos e dos moluscos

Frases formadas para o índice

Produtos para a destruição dos artrópodos 632.951
 Destruição dos artrópodos, Produtos para a 632.951
 Artrópodos, Produtos para a destruição dos 632.951
 Produtos para a destruição dos moluscos 632.951
 Destruição dos moluscos, Produtos para a 632.951
 Moluscos, Produtos para a destruição dos 632.951

b) A vírgula (,) ou o E juntamente com um possessivo formam uma única frase

631.531 Sementes e seu tratamento. Sementeira

Frases formadas para o índice

Sementes e seu tratamento 631.531
 Tratamento, Sementes e seu 631.531
 Sementeira 631.531

c) Os termos entre parênteses foram considerados como termos isolados e assim são recuperados

631.312 Charruas (Arados)

Frases formadas para o índice

Charruas 631.312
 Arados 631.312

d) Aos termos citados depois dos dois pontos, foi acrescentada entre parênteses a frase que o antecede, visando uma explicação para uma melhor recuperação

631.58 Métodos de cultura: pousio, intensiva, pastoril, eletro-cultura

Frases formadas para o índice

Métodos de cultura 631.58

Cultura, Métodos de 631.58

Pousio (Métodos de cultura) 631.58

Intensiva (Métodos de cultura) 631.58

Pastoril (Métodos de cultura) 631.58

Eletrocultura (Métodos de cultura) 631.58

e) O termo OU interrompe a frase

632.16 Doenças fisiológicas diversas ou devidas a causas complexas

Frases formadas para o índice

Doenças fisiológicas diversas 632.16

Fisiológicas diversas, Doenças 632.16

Doenças fisiológicas devidas a causas complexas 632.16

Fisiológicas devidas a causas complexas, Doenças 632.16

Causas complexas, Doenças fisiológicas devidas a 632.16

Complexas, Doenças fisiológicas devidas a causas 632.16

f) O termo OUTRO interrompe a frase e o que vem a seguir, formará nova frase:

633.17 Cera de abelhas e outras substâncias produzidas pelas abelhas

Frases formadas para o índice

Cera de abelhas 633.17

Abelhas, Cera de 633.17

Substâncias produzidas pelas abelhas 633.17

Produzidas pelas abelhas, Substâncias 633.17

Abelhas, Substâncias produzidas pelas 633.17

3.3.3 - Vocabulário de palavras não significativas. O texto é analisado para a verificação das palavras não significativas: artigos, preposições, advérbios, conjunções e também substantivos e adjetivos quando não devam ser colocados em evidência. Estas palavras são perfuradas em cartões para que sejam excluídas da ordem alfabética do índice (Anexo 7.3.4)

3.3.4 - Índice em outras línguas. Foram escolhidos textos em 5 (cinco) línguas: inglês, francês, espanhol, alemão e português para o processamento pelo mesmo programa. Acrescentou-se na análise do texto, os termos nestas línguas e ao vocabulário as palavras não significativas.

3.3.5 - Descrição dos programas

3.3.5.1 - Gravação dos cartões correspondentes à tabela COU em área de disco chamada ELVCOU.

3.3.5.2 - Produção do índice

Leitura do vocabulário. As palavras não significativas são perfuradas em cartões a partir da coluna 1 ocupando no máximo 20 colunas. Esses cartões são lidos e colocados em uma tabela chamada VOC.

Leitura de cada registro correspondente a um cartão. Cada registro é lido em área de trabalho até encontrar classificação diferente, garantindo um texto completo para cada classificação mesmo que este ocupe mais um registro correspondente a um cartão.

Análise do texto. Cada posição do registro é analisada e passada para uma área chamada INTER até que se encontre qualquer símbolo, palavra, sinal etc. que indique fim de frase. Cada palavra, símbolo, sinal etc. é agora considerado como um termo e é movido para uma tabela de nome TERMO. Cada um desses elementos captados é nivelado para um posterior reconhecimento. Após o nivelamento, se o termo em questão for ponto (.) ou dois pontos (:), o programa seguirá para a rotina de preparação da área de impressão. Caso contrário, o programa continuará na análise do texto.

Preparação da área de gravação. A frase captada é distribuída palavra por palavra em uma tabela de 2 dimensões. É então verificado o nível de cada termo para a construção das frases, seguindo as diretrizes determinadas no programa (3.3.2). Ca

da frase é rotada para que cada um de seus termos apareça em ordem alfabética. As palavras não significativas são omitidas na sua rotação através de testes com a tabela VOC. Cada frase é então gravada em área de disco chamada ELVCD1, para posterior ordenação alfabética, em área ELVCD2.

3.3.5.3 - Listagem do índice

Os registros correspondentes a cada frase, agora em ordem alfabética, são lidos da área de disco ELVCDU2 e distribuídos na área de saída para a formação do índice em 2 (duas) colunas.

Cada primeiro termo de uma frase é citado apenas uma vez omitindo-o nas outras linhas, caso seja idêntico ao anterior, mantendo o texto a seguir em margem interna.

Frases ou termos idênticos com mais de uma classificação só são citados uma vez com suas classificações separadas por vírgula.

Devido ao tamanho da frase criada, muitas vezes foi necessário a exclusão de parte da frase. Esta omissão é apontada por uma seta (<=).

A paginação do índice pode ser contínua à da tabela.

Guardando certas proporções o índice pode ter diversos gabaritos para impressão.

3.3.5.4 - Listagem da tabela

A tabela é listada dando destaque aos números de classificação até 3 (três) algarismos para facilitar a consulta.

Quando utiliza-se mais de um cartão para a representação de um número de classificação, este só será citado uma vez.

A paginação poderá ser iniciada em qualquer número.

Guardando certas proporções a tabela pode ter vários gabaritos para impressão.

3.3.5 - Discussão. O uso de linguagem natural no processamento de índices das tabelas da CDU.

As sentenças e frases que vão compor o índice (Anexo 7.3), são criadas a partir da análise do texto da tabela. As palavras são distribuídas umas após outras para a formação de novas frases. Isto implica em um texto bem estruturado, com fra-

ses com sentido completo para uma melhor recuperação.

3.3.8.1 - Problemas:

a) As frases com sentido ambíguo levam a falsas recuperações;

b) Os termos vagos quando citados isoladamente, levam à recuperação incompleta;

c) Os termos entre parênteses citados nas frases analisadas apesar de serem explicativos, foram apresentados de diferentes formas, como se vê a seguir, que dificultam o trabalho do computador.

1 - O(s) termo(s) podendo ser considerado(s) como um todo

631.312 Charruas (Arados)

631.537.2 Hidropônicos (Cultura sem solo)

64.0.3 Estudo científico do trabalho (Racionalização do trabalho)

(esta foi a forma adotada)

2 - O(s) termo(s) sem significação quando isolado(s), podendo ser considerado(s) como simples continuação do texto, sendo desnecessários os parênteses:

631.82 Adubos químicos (artificiais) em geral

631.92 Influência geofísica das plantações e culturas (sobre o clima, o solo, a água etc.)

3 - O(s) termo(s) sem significação quando isolado(s), sendo impossível sua recuperação lógica, devido à construção da frase:

634.0.4 Danos nas florestas e sua prevenção (proteção)

onde, quanto à recuperação do termo "Proteção", só seria lógica com a frase "Proteção das florestas" o que não poderia ser conseguido automaticamente.

d) o processamento do índice relativo devido à rotação de frases com inclusão de vírgulas, trouxe-nos problemas na ordenação alfabética.

Nas frases criadas:

- * Cozinhas de apartamentos 643.392
Apartamentos, Cozinhas de 643.392
- * Cozinhas de hotéis 643.36
Hotéis, Cozinhas de 643.36
Aparelhos para aquecimento das cozinhas 644.12
Aquecimento das cozinhas, Aparelhos para 644.12
- * Cozinhas, Aparelhos para aquecimento das 644.12

a ordem alfabética do termo "Cozinhas", seria:

Cozinhas de apartamentos 643.392
Cozinhas de hotéis 643.36
Cozinhas, Aparelhos para aquecimento das 644.12

a solução foi a criação normal das frases com sua rotação incluindo a vírgula e ao gravar para a ordenação alfabética, a vírgula é suprimida se aparece seguindo a 1.ª palavra da frase conseguindo a ordem alfabética ideal:

Cozinhas

aparelhos para aquecimento das 644.12
de apartamentos 643.392
de hotéis 643.36

Esta solução implica sempre na mudança de linha de impressão toda vez que a frase formada é composta de mais de um termo, o que nos leva a ter frases diretas ou rotadas, onde o texto a partir do 2º termo é levado para a próxima linha, mesmo que seja a única frase sobre o assunto.

Lenha

fogões à 643.332

Matérias

plásticas 645.11

e) Frases do texto das tabelas

- 1 - 634.0 Florestas, ciências florestais e utilização dos produtos florestais }

onde o ideal seria

Florestas. Ciências florestais. Utilização dos produtos florestais

pois o adjetivo *florestais* citado pela primeira vez pela rotina do programa levaria a ser tomado como adjetivo de ambos os substantivos que o antecede e assim seria recuperado:

Florestas florestais

Ciências florestais

3.3.6.2 - Vantagens

- a) As frases são desdobradas em tantas quantas o texto o exigir;
- b) A pesquisa é facilitada, pois são colocadas em evidência todas as palavras significativas, relacionando-as dentro das frases pela rotação;
- c) Possibilita a consulta direta por qualquer termo da nossa linguagem natural, evitando a antecipada busca a um termo que indicaria o assunto, como se faz em pesquisas em índices de linguagem artificial;
- d) A exclusão de qualquer palavra não significativa do índice;
- e) A apresentação em forma de índice relativo, facilita a consulta pela disposição das frases de forma semelhante à tradicional;
- f) O processamento das tabelas em outras línguas que não a portuguesa (Anexo 7.3.1).
- g) A possibilidade de processamento do índice de outros sistemas de classificação, cuja apresentação de texto se assemelhe ao da CDU, como a "Dewey Decimal Classification" etc.;
- h) A correção e/ou atualização de tabelas, bastando apenas a troca ou inserção de novos cartões;
- i) As listagens podendo ser encaminhadas diretamente à gráfica para impressão por processo "offset" evitando as tarefas de composição e revisão tipográficas;
- j) A rapidez do processamento por computador nos testes de Física (1338 classificações) e Ciências Domésticas (311 classificações) cujos índices foram gerados e listados em 1h e 14 min respectivamente.

4 - RESULTADOS

4.1 - Apuração

Nos testes realizados, o índice da tabela 54 (CIÊNCIAS DOMÉSTICAS) foi processado pelos sistemas ELVCOU2 e ELVCOU3 com os seguintes resultados:

	Classificações	Cartões	Reg. p/ índice	Tempo
ELVCOU2	311	348	1131	12 min
ELVCOU3	311	348	1237	14 min

Como se observa, o sistema ELVCOU3 requer maior tempo de processamento pela sofisticação dos programas e também pelo maior grau de recuperação.

Os testes realizados com a tabela 53 (FÍSICA) e parte da tabela 3 (CIÊNCIAS SOCIAIS) para as diversas línguas apresentaram os seguintes resultados no sistema ELVCOU3:

	Classificações	Cartões	Reg. p/ índice	Tempo
FÍSICA	1336	1542	5280	1 h
CI. SOCIAIS (poliglote)	48	52	226	3 min

5 - CONCLUSÕES

Os três estudos realizados com a finalidade de verificar a viabilidade da elaboração automática dos índices das tabelas da CDU, nos levaram às seguintes conclusões:

5.1 - ELVCDU1 - Sistema semi-automático por exigir marcação de toda a tabela a fim de se processar o índice. Apesar da morosidade da preparação dos dados, os resultados foram satisfatórios. Não é aconselhado como sistema.

5.2 - ELVCDU2 - Sistema automático com bom nível de recuperação. Pode ser utilizado, embora a pesquisa ao índice apresente dificuldade pelo tipo de visualização do referido índice.

5.3 - ELVCDU3 - Sistema automático com alto nível de recuperação. Como o índice é gerado a partir do texto, frases são construídas e uma única frase pode gerar tantas frases quantas sejam as necessidades de recuperação. A pesquisa ao índice é facilitada, pois as palavras significativas aparecem em destaque na frase. A forma de apresentação do índice foi escolhida por se assemelhar bastante aos índices elaborados de forma convencional de sorte a não exigir dos usuários qualquer mudança do comportamento do usuário na consulta.

6 - CITAÇÕES BIBLIOGRÁFICAS

- 1 - FEDERAÇÃO INTERNACIONAL DE DOCUMENTAÇÃO - Conversion des index alphabétiques des tables de la CDU en thesaurus. Rapport (FID/CCC/O - Document C 72-34).
- 2 - LLOYD, G. A. - Política da CDU, atividades correntes e novo programa de desenvolvimento para a década de 1970. In: FORUM aberto da CDU. Buenos Aires, 1970. p. 11-9.
- 3 - RIGBY, Malcolm - Computers and the UDC; a decade of progress 1960-1970. Rockville, Md., Scientific Information on Documentation Division, 1970.
- 4 - RIGBY, Malcolm - Experiments in mechanized control of meteorological and geostrophysical literature and UDC schedules in these fields. Revue Internationale de la Documentation, La Haye, 31(3):103-6, 1964.

7 - ANEXOS

7.1 - ELVCDU1

7.1.1 - 3 CIÊNCIAS SOCIAIS (parte)

7.1.1.1 - Listagem dos cartões

7.1.1.2 - Tabela

7.1.1.3 - Índice

7.2 - ELVCDU2

7.2.1 - 64 CIÊNCIAS DOMÉSTICAS

7.2.1.1 - Listagem dos cartões

7.2.1.2 - Índice

7.3 - ELVCDU3

7.3.1 - 3 CIÊNCIAS SOCIAIS (parte)

7.3.1.1 - Tabela

7.3.1.2 - Índice

7.3.2 - 53 FÍSICA

7.3.2.1 - Tabela

7.3.2.2 - Índice

7.3.3 - 64 CIÊNCIAS DOMÉSTICAS

7.3.3.1 - Tabela

7.3.3.2 - Índice

7.3.4 - Vocabulário de palavras não significativas