

Índice alfabético da segunda edição média da CDU em língua portuguesa

Fátima Ganim

Bibliotecária – IBICT

Mauro Kenji Sujii

Analista de Sistema – IBICT

Odilon Gonçalves Martins

Técnico de Informática – IBICT

Odilon Pereira da Silva

Prof. Dep. de Biblioteconomia da UnB e Tec. de Inf. – IBICT

Zairton Bastos Pinheiro

Técnico de Informática – IBICT

SCN – Quadra 2 – Bloco K – 70000 Brasília, DF

Resumo – Descrição do Índice Alfabético da 2ª Edição Média da CDU em língua portuguesa, elaborado com emprego dos recursos oferecidos pela automação. São apresentadas algumas de suas características principais, bem como as decisões que presidiram à feitura do mesmo.

Antecedentes

A primeira edição média da Classificação Decimal Universal (CDU) em língua portuguesa, preparada pelo Instituto Brasileiro de Bibliografia e Documentação (IBBD) e publicada, em 1976, pelo órgão que lhe sucedeu, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), teve sua tiragem esgotada muito mais cedo do que o esperado. Em decorrência disso, na virada da década, já havia uma demanda e um clamor razoáveis por uma segunda edição.

Estudos sobre usuários e usos da CDU no Brasil, a que, por aquela época, já se vira forçado a proceder o Instituto, confirmaram a demanda, e levaram sua direção a realizar os primeiros estudos de viabilidade técnica e financeira dessa segunda edição. Os estudos amadurecidos, surgiu um acontecimento fortuito, que veio colaborar, imensamente, para o início dos traba-

ÍNDICE ALFABÉTICO DA SEGUNDA EDIÇÃO

lhos: o lançamento, em 1985, da versão inglesa da primeira edição média internacional trilingüe, projeto antigo da Federação Internacional de Informação e Documentação (FID).

Do projeto inicial, constava o lançamento simultâneo de uma versão alemã, uma inglesa e uma francesa, previstas para 1977, em comemoração ao primeiro centenário da primeira edição da Classificação Decimal de Melvil Dewey (CDD). Somente a versão alemã ficou pronta a tempo e foi publicada conforme o plano. A francesa, em virtude de dificuldades de natureza vária, até hoje não veio à luz, e a inglesa só ocorreu em 1985, mas com o texto guardando fidelidade ao acordo entre os editores dos três idiomas oficiais da FID, correspondendo ao estado de desenvolvimento do sistema em 1977.

Razões da escolha do texto

Além da circunstância óbvia de ser o inglês um dos três idiomas oficiais da Federação, e de sua maior acessibilidade por parte dos usuários de língua portuguesa (no Brasil), acresceu-se a de ter a versão inglesa sido produzida através do emprego de recursos automatizados e se encontrar disponível, no mercado, também em meio magnético.

Verificada a competência técnica e a conveniência de o Instituto ousar o emprego dessas novas tecnologias editoriais, foi adquirida, à British Standards Institution, a versão em fita magnética, a partir da qual se produziu uma linguagem de computador, base dos trabalhos de tradução para a língua portuguesa. E foi esse texto, traduzido, adaptado e atualizado (com as devidas licenças da FID) até à Extensão e Correção 12(2), correspondente ao ano de 1984, que se refere ao **Índice Alfabético**, sobre cuja metodologia de elaboração prestaremos alguns esclarecimentos técnicos a seguir.

Cenário

Lançado o primeiro volume (Tabelas) em 1987, havia um compromisso moral (e uma promessa pública e solene, feita durante o 14º Congresso Brasileiro de Biblioteconomia e Documentação, no Recife) do Instituto, de providenciar a publicação de seu índice alfabético com a maior brevidade possível. A meta estabelecida foi 1988. Diante da magnitude do empreendimento, da exigüidade do tempo e face aos constrangimentos de toda ordem, entre os quais se inscrevem os financeiros e os relacionados com recursos humanos, tivemos que optar por uma metodologia que, levando em conta essas circunstâncias adversas, pareceu a mais adequada à realidade do momento. Não

havia clima para aventuras maiores, para experimentações mais ousadas em matéria de tecnologias avançadas de indexação, sobretudo de indexação de algo tão complexo e tão peculiar quanto uma linguagem estruturada pré-coordenada, com um sistema de classificação decimal.

Acresça-se o fato de estarmos trabalhando com uma equipe técnica inferiorizada em relação à magnitude da tarefa, e, embora entusiasta das teorias mais modernas e dos avanços tecnológicos à sua disposição, na verdade ainda bastante inexperiente para ousar, sem razoável dose de experimentação, um empreendimento de tamanha responsabilidade. Daí termos optado, prudentemente, responsabilmente (pensávamos), por uma metodologia conciliatória, em que pudéssemos explorar ao máximo os recursos postos à nossa disposição por essas teorias e tecnologias de indexação, mas sem abdicarmos demais do controle, da intervenção humana direta, da manipulação intelectual (perdoem a antinomia), dos ajustes, das achegas, dos cortes e acréscimos, que nos parecem inevitáveis, se não desejáveis, quando é ainda bastante questionável a competência das maravilhosas máquinas “inteligentes”, para executar trabalho satisfatório, em áreas em que predominam as idéias e os conceitos (não apenas os termos); a sintaxe (não apenas a morfologia); a semântica (não apenas a terminologia e a taxonomia).

Preferimos reconhecer razão em quem advertiu: “contrariamente ao que se espera mesmo dos computadores mais sofisticados, do indexador (competente) espera-se que exercite sua faculdade de julgar, decidindo quanto ao que deve ser aceito no índice ou dele omitido”. E, sem dúvida alguma, não foi apenas com surpresa, mas com alguma simpatia que tomamos conhecimento da descoberta de indexadores experientes de que “um número incrivelmente significativo de usuários fica mais bem impressionado com o **layout** e as qualidades tipográficas de um índice, do que com suas qualidades intrínsecas e técnicas”.

†

Metodologia

O ponto de partida foi, como dissemos, o texto já atualizado, na versão em língua portuguesa. No formato em meio magnético, esse texto está dividido em cinco áreas diferentes (registros), correspondentes a:

Registro “A”: Número principal e texto correspondente

Registro “B”: Notas especiais

Registro “C”: Nota “subdivisão paralela”

Registro “D”: Exemplos

Registro “E”: Remissivas “ver” e “ver também”

ÍNDICE ALFABÉTICO DA SEGUNDA EDIÇÃO

Para efeito de análise e indexação, foram considerados, em princípio, pertinentes, os conceitos presentes nos registros "A". Eventual, mas muito raramente, os dos registros "D" (área de EXEMPLOS).

A equipe de técnicos de informação procedia à análise dos textos à procura de termos úteis nesses registros. Num primeiro momento, eram assinalados, para a equipe de digitação, os textos aprovados para indexação, ao mesmo tempo em que se indicavam as inversões mais óbvias e se criavam os qualificadores e, eventualmente, os modificadores aparentemente mais adequados para representar fielmente o(s) conceito(s) no contexto.

A decisão inicial de efetuar essas intervenções numa fase em que se tinha em mãos o contexto, confirmou-se acertada, como o demonstraram as raras oportunidades em que precisamos deixar a decisão final para etapa posterior: o retorno ao (con)texto tornava-se, então, imprescindível.

Após uma primeira digitação, de que resultou a primeira listagem em ordem alfabética, ainda sem um formato convencional qualquer de índice, foi possível tomar decisões quanto a este.

Céticos, desde o início, quanto à eficiência e à respeitabilidade recente entre os profissionais da informação, dos índices com formato KWIC, KWOC, KWAC e quejandos, solicitamos, da equipe de informática do Instituto, um índice com feições semelhantes aos de fabricação manual (chamados, às vezes pejorativamente, de tradicionais), com possibilidade de entradas para subcabeçalhos em primeiro e segundos níveis, o que nos parecia um layout suficientemente sóbrio, claro, prático, elegante, talvez, visualmente, até agradável. Parecia uma decisão acertada, no sentido de ser uma solução que evitava o inconveniente dos índices com uma série interminável de referências (números) para cada termo, e, ao mesmo tempo, o outro inconveniente, não menor nem menos grave, de subdivisões infundáveis, multiplicação de subcabeçalhos em primeiro, segundo, terceiro, quarto níveis, que tornam desnecessariamente confusos, de difícil consulta, e, por isso, detestáveis, os índices com essas características.

Num segundo momento, obtidas listagens já em formato que pretendíamos final, e onde era possível, em virtude da ordem alfabética rigorosa dos cabeçalhos e subcabeçalhos, tomar decisões quanto aos demais aspectos das sofisticadas técnicas pretendidas para o índice, fomos paulatina, e também repetida e alternadamente, introduzindo essas melhorias, até obtermos o produto final, de cujas características básicas demos conhecimento sucinto na APRESENTAÇÃO do texto impresso da Tabela (volume um), e que, aqui, pedimos vênua para transcrever.

Contrariamente à praxe adotada no feitiço de outros índices CDU, não nos pareceu oportuno indicar sempre, como qualificadores, as áreas de assunto de todos os termos homógrafos. São empregados qualificadores, apenas quando julgados necessários, ainda que, no índice, não co-ocorram em relação de vizinhança com seu(s) par(es). Igualmente, deixa de haver qualificador, quando os termos homógrafos são os de aceção e emprego mais usuais na língua.

Sintaticamente sóbrio, evita, deliberadamente, ser uma réplica da estrutura do sistema a que serve de apoio, ao mesmo tempo em que se despe das pretensões de emular os tesouros, que diferente é sua finalidade.

Dentro do mesmo critério de economia sintática, em vez de mil remissivas, referências cruzadas, modificadores em segundo, terceiro e até quarto níveis, qualificadores em profusão e abundantes notas "cf", preferimos remeter os consulentes para os números relativos aos diferentes contextos na tabela onde o termo ou o tema procurado ocorrem. Inclusive para desestimular a tentação de se classificar em base ao índice, sem consulta às tabelas sistemáticas. Daí não termos, também, feito controle de sinônimos à moda dos tesouros, com remissiva dos termos não-adotados, ou pouco conhecidos, para os de emprego mais freqüente ou os aprovados pelo sistema. Simplesmente, incluímos todos os sinônimos encontrados no texto, em sua respectiva ordem alfabética, seguidos do(s) número(s) CDU pertinente(s). Foram-lhe (ao índice) acrescentados, entretanto, inúmeros sinônimos ausentes da tabela sistemática, sobretudo nas classes 5 e 6 (Ciência e Tecnologia, respectivamente), especialmente termos relativos a nomes de plantas e animais da fauna e da flora brasileira, com destaque para os de origem indígena e os conhecidos, popularmente, por formas que não as registradas na tabela. Tal não impede, contudo, que determinada variante de um nome, ou mesmo determinado termo de emprego generalizado no Brasil, deixe de figurar no índice. Afinal, não se deve esquecer de que "*quandoque dormitât (et) Homerus*". Cochilos ocorrem.

Outros pontos – sobre os quais parecem oportunos, aqui, alguns esclarecimentos – são:

Nomenclatura científica – Nas áreas de Botânica e Zoologia, sobretudo, optamos por oferecer, a par da forma vernácula, a latina, tradicionalmente adotada na taxonomia daqueles dois campos do conhecimento, visando a uma mais rápida e inequívoca identificação do assunto quando aquela terminologia é a empregada (e com freqüência o é) na comunicação científica.

Ortografia – Acolhemos a do texto, exceto nos casos de evidente falha

ÍNDICE ALFABÉTICO DA SEGUNDA EDIÇÃO

de revisão tipográfica, oportunidade em que fizemos apelo aos instrumentos lexicográficos de maior autoridade e aceitação na língua, como o “Novo Dicionário da Língua Portuguesa”, de Aurélio Buarque de Holanda Ferreira, e o “Vocabulário Ortográfico”, da Academia Brasileira de Letras, entre outros.

Pontuação e sinais diacríticos – Sinais de pontuação não contam para fins de alfabetação. Semelhantemente, acentos e sinais diacríticos não são levados em consideração, contrariamente ao que estatui a P-NB-106. Assim, Ü = U; Ö = O, e não UE e OE, respectivamente. Julgamos, aqui, mais aconselhável aceitar as instruções previamente inseridas nos programas do sistema e na Tabela ASC II para alfabetação, do que conformá-las às normas estabelecidas pela ABNT.

Preposições – Pró, meso ou enclíticas, empregamo-las apenas quando pareceram absolutamente essenciais à clareza da relação sintática existente entre os termos.

Qualificadores – Fizemos uso deles com dupla função: a tradicional e mais freqüente, de indicar o contexto em que ocorre um determinado termo, sobretudo nos casos de homógrafos; e como meros sinônimos ou equivalentes dos termos de entrada. Em ambos os casos, com a finalidade de esclarecer o âmbito (contexto) ou a extensão semântica do termo em questão.

Referências/remissivas – Preferimos evitar o aparato solene e a complexidade de um índice com esse tipo de recurso sintático, em virtude, aliás, de suas próprias peculiaridades: refere-se a um documento sintática e hierarquicamente estruturado, cujas relações múltiplas entre os conceitos e os assuntos por eles representados constituem sua feição mais característica, tornando pleonástico o emprego desses mesmos recursos em seu índice alfabético relativo.

Singular x plural – A preferência recaiu sobre a primeira das duas formas sempre que o permitiram a clareza, a semântica e o uso corrente do vernáculo. Do contrário, empregamos alternativa ou conjuntamente a ambas. Acolhemos, sistematicamente, a forma do plural para: nomes designativos de gêneros e famílias na Botânica e na Zoologia; nomes de pessoas, sobretudo os associados a raças, povos, seitas e denominações religiosas; termos das matemáticas e termos que só possuem essa forma no vernáculo. Evitamos, contudo, por nos parecer deselegante, além de demasiado simplista, a solução (?) de fundir num radical único às formas singular e plural, visto tal procedimento afigurar-se-nos um desrespeito à semântica.

Topônimos – Indicamos, sistematicamente, o auxiliar comum geográfico e seus correspondentes números principais designativos da Geografia e da História dos países, tanto do mundo antigo quanto do moderno. O mesmo fizemos com os estados e municípios brasileiros. Omitimos, entretanto, por facilmente localizáveis nas tabelas, os nomes de cidades, mesmo capitais.

As pequenas discrepâncias ocasionais, que ocorrem entre as três categorias de números referentes a países, devem ser imputadas ao texto, que assim (discrepantes) os registra no próprio original em que se baseou a tradução portuguesa: a versão inglesa da Primeira Edição Média Internacional Trilíngüe.

Variantes morfológicas – Adotamos, por norma, as formas registradas no texto das tabelas, mas introduzimos algumas que nos pareceram igualmente relevantes como termos de busca, uma vez que são de emprego generalizado em alguma das regiões brasileiras. São elas, sobretudo, as referentes a nomes de plantas e de animais, especialmente os étimos de procedência indígena.

Desenvolvimento do sistema de automação – O processo de automação do índice da CDU foi, basicamente, calcado em experimentalismo. O tipo peculiar de dado a ser manipulado e a falta de literatura sobre a forma de se proceder com tal índice resultaram na proposta de desenvolvimento de um sistema automático, que não se enquadrava nos padrões normais existentes. A própria inadequabilidade dos equipamentos utilizados constituiu-se em obstáculo ameaçador ao sucesso do projeto. O computador utilizado não reconhecia a acentuação da língua portuguesa e obedecia o padrão ASC II (American Standard Code II). O próprio Biomumps, sistema operacional e linguagem de programação adotado, não satisfazia as necessidades de ordenação exigidas na CDU.

Durante o processo de desenvolvimento do sistema, o primeiro problema detectado foi a necessidade de se representar os caracteres acentuados. Para isso, foi criada uma convenção que estipulava uma determinada sequência de caracteres para cada letra acentuada ou para a cedilha.

Vencido este problema, surgiram os de ordenação dos números da CDU e dos termos de indexação. Para resolver o primeiro, foi criada uma tabela com os códigos que compõem um determinado número. Cada caracter componente do número da CDU é pesquisado na tabela e a posição de cada caracter indica a forma de classificação do número.

Já a ordenação dos termos de indexação foi efetuada com a retirada de

ÍNDICE ALFABÉTICO DA SEGUNDA EDIÇÃO

acentos e pontuação das palavras componentes de cada termo de indexação. Um outro artifício utilizado foi a criação de grupos. Um grupo representava um conjunto de entradas no índice, que deviam estar agrupadas para que a emissão apresentasse uma hierarquia ou subordinação. Nesta hierarquia, a primeira entrada apresentava uma expressão comum a todas as entradas, e as entradas seguintes apresentavam somente o acréscimo à primeira entrada. Devido a isso os grupos alteraram a forma de ordenação, de modo a permitir que as entradas de um mesmo grupo estivessem juntas.

O desenvolvimento do sistema de automação do índice da CDU foi dificultado por limitações existentes tanto no equipamento utilizado (minicomputador Sisco 10.300) quanto no sistema operacional e linguagem Biomumps. Estas dificuldades foram sanadas através de certos artifícios práticos. Estes artifícios não se constituem em propostas genéricas de implementação em um sistema automatizado para índices de CDU. Trata-se, somente, de um relato que apresenta a forma de implementação de um sistema em um ambiente particularizado.

Conclusão

Durante todo o tempo de elaboração do índice, procuramos o equilíbrio entre duas preocupações conflitantes que parecem presidir ao trabalho de todo indexador, expressas magistralmente por esses dois pensamentos de Norman Knight:

"muitas vezes o indexador é tentado a pensar que a finalidade do índice é evitar o trabalho de ler o livro. Um índice pode fazer isto, mas esta não é sua primeira finalidade"; e

"há tantas utilidades diferentes que os leitores dão aos livros, que o indexador nunca pode estar certo de que não está esquecendo uma necessidade peculiar". Teremos conseguido o equilíbrio desejado? A crítica dos usuários o dirá. E ajudará enormemente no preparo das próximas edições e de trabalhos semelhantes.

Abstract – A description of UDC 2nd Portuguese Medium Edition Alphabetical Index – produced with the aid of the computer – is presented. Some of its basic characteristics are described, together with the decisions taken in the process of making it.