



TOGETHER
for a sustainable future

OCCASION

This publication has been made available to the public on the occasion of the 50th anniversary of the United Nations Industrial Development Organisation.



TOGETHER
for a sustainable future

DISCLAIMER

This document has been produced without formal United Nations editing. The designations employed and the presentation of the material in this document do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations Industrial Development Organization (UNIDO) concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries, or its economic system or degree of development. Designations such as “developed”, “industrialized” and “developing” are intended for statistical convenience and do not necessarily express a judgment about the stage reached by a particular country or area in the development process. Mention of firm names or commercial products does not constitute an endorsement by UNIDO.

FAIR USE POLICY

Any part of this publication may be quoted and referenced for educational and research purposes without additional permission from UNIDO. However, those who make use of quoting and referencing this publication are requested to follow the Fair Use Policy of giving due credit to UNIDO.

CONTACT

Please contact publications@unido.org for further information concerning UNIDO publications.

For more information about UNIDO, please visit us at www.unido.org

20329

Projecto DP/MOZ/86/014
Documento de Campo No. 25

**LEGIS: BASE DE DADOS DE
LEGISLAÇÃO INDUSTRIAL**

ÂNGELO SI'TOLE

FERNANDO L. GARAGORRY

MINISTÉRIO DA INDÚSTRIA E ENERGIA

Maputo, Dezembro de 1992

LEGIS: DATA BASE OF INDUSTRIAL LEGISLATION

ABSTRACT

The paper presents the main aspects related to the development of a data base of legislation, at the Ministry of Industry and Energy, Mozambique, as part of the work conducted under UNDP/UNIDO Project DP/MOZ/86/014.

The data base, named LEGIS, concentrates on legislation related to industry and energy, but also contains references to legal documents of wider application.

All types of legal documents are acceptable (i.e., laws, decrees, decisions, etc), and there is no limitation as to the period covered, although preference has been given to legislation approved after Independence, in 1975.

With respect to the nucleus of the information contained in the records, it was found that it would only be feasible to include the summary of the legal texts, as published in the official bulletins, with the addition, eventually, of some explanatory notes. Beside this nucleus, there are other fields which are required to identify the text, and a field with additional descriptors.

As far as possible, descriptors have been taken from the U.N. Macrothesaurus, which proved to be quite satisfactory for the description of the industry-related matters. Carefully controlled descriptors had to be added to the basic vocabulary, mainly in connection with administrative and legal matters.

At the time of preparation of the present paper, the data base, implemented under CDS/ISIS, contained slightly over 900 records, which means the most substantial part of the legislation approved over the years 1975 to 1990, within the relevant subject matter.

1. INTRODUÇÃO

O presente trabalho visa apresentar os aspectos mais importantes relacionados com o desenvolvimento de uma base de dados de legislação, no Ministério da Indústria e Energia (MIE), como parte das actividades realizadas sob o Projecto DP/MOZ/86/014, executado pela ONUDI.

O principal objectivo, para se ter uma base de dados de legislação, está relacionado com o suporte que pode ser oferecido a outras actividades do Projecto ou, em forma mais geral, do MIE.

De acordo com o conhecimento disponível, o exercício revestiu-se de certa novidade no âmbito da ONUDI e, nesse sentido, contou com o apoio necessário por parte da Secção de Informação Industrial e Tecnológica da sede dessa organização. Em particular, foi facilitado o processo de obtenção do programa CDS/ISIS (usualmente designado apenas como ISIS), distribuído pela UNESCO [3], que foi utilizado para implementar a base de dados. Cabe assinalar que esse programa é usado para todas as bases de dados textuais existentes na sede da ONUDI, tais como a de referências bibliográficas e as diversas bases relacionadas com tecnologias.

Na época da elaboração do presente documento, a base de dados de legislação, denominada LEGIS, contava com algo mais de 900 registos.

2. DESCRIÇÃO DA BASE DE DADOS

2.1 Abrangência

Inicialmente, delimitou-se a abrangência aproximada da base de dados, no que diz respeito ao conteúdo temático. Entendeu-se que deveriam ser incluídos, prioritariamente, registos referentes a diplomas que estivessem claramente associados com indústria ou energia. Além dessas referências, poderiam ser incluídos registos referentes a textos de abrangência mais geral.

No que se refere ao tipo de textos (leis, decretos, despachos, etc), não houve limitação. Todos eles são considerados como fazendo parte da legislação. Em princípio, haveria uma limitação prática, no sentido de que a fonte principal para a selecção dos textos são os boletins oficiais.

No que se refere ao período coberto não há limitação, mas deu-se prioridade à legislação que surgiu após a Independência.

Com respeito ao núcleo do conteúdo dos registos, entendeu-se que só seria viável incluir os resumos dos textos legais, tal como aparecem publicados, adicionando-se, eventualmente, notas com os esclarecimentos considerados necessários. Além desse núcleo, foram incluídos outros campos requeridos para completar a identificação dos textos, e um campo de descritores adicionais, para facilitar a recuperação e a elaboração de relatórios indexados.

2.2 Campos

À continuação serão apresentados os campos que definem a base de dados, sem entrar em maiores detalhes técnicos (tamanho máximo, etc). Pode ver-se, mais adiante, a Ficha Geral de Entrada de Dados, que é utilizada para a codificação dos novos registos; nela aparecem todos os campos da base de dados.

Nas explicações que seguem, o nome de cada campo está precedido por uma "etiqueta" (número 10, 20, etc), que é a forma interna usada pelo ISIS para identificar o campo.

Inicialmente, há dois campos utilizados para identificar a fonte documental de onde foram obtidos os dados. Esses dois campos são obrigatórios em todos os registos.

10. FONTE. Em geral, trata-se do jornal oficial utilizado na época da publicação do diploma. Na actualidade, denomina-se Boletim da República; anteriormente, denominava-se Boletim Oficial. No entanto, não há limitação para se utilizar outras fontes documentais. No caso dos boletins mencionados anteriormente, usam-se as abreviaturas B.R. e B.O., respectivamente. Além da designação da publicação, indica-se o número, se for o caso, para facilitar a identificação. Por exemplo: B.R. 18.
20. ANO DE PUBLICAÇÃO. Trata-se do ano em que foi publicado o diploma, na publicação indicada no campo FONTE. O uso do ano de publicação como um campo separado da FONTE visa facilitar a recuperação, bem como algumas tarefas de controle (crítica de dados, emissão de relatórios, etc).

À continuação, há cinco campos que fornecem dados adicionais para identificar o diploma.

30. TIPO. Indica se o diploma é lei, resolução, decreto, despacho, etc. Este campo é obrigatório.
40. NÚMERO. Se for o caso (por exemplo, nas leis), indica o número do diploma. Por exemplo: 12/83. Cabe notar o seguinte: 1) alguns diplomas não têm número (isso acontece, por exemplo, com os despachos) e, nesses casos, o campo é deixado em branco; 2) quando há número, o ano aparece nos dois últimos dígitos, que indicam o ano de aprovação; 3) o diploma pode aparecer publicado num ano posterior ao da sua aprovação e, portanto, mesmo que tenha número, não há redundância semântica com o campo ANO DE PUBLICAÇÃO.
50. DATA. Quando for conhecida a data de aprovação, esta aparecerá na forma DD/MM (isto é, dia e mês).
60. ANO. Se o diploma não tiver número, este campo é utilizado para indicar o ano de aprovação, quando este for diferente do ANO DE PUBLICAÇÃO. Nos outros casos (isto é, se o diploma tiver número, ou se o ANO DE PUBLICAÇÃO for o mesmo que o da aprovação), este campo é deixado em branco.
70. ORIGEM. Este campo é utilizado para indicar o órgão ou órgãos de onde emanou o diploma. Por exemplo: Presidência da República, Ministério da Indústria e Energia, etc. Este

campo, que é obrigatório, é tratado como "repetitivo" (de acordo com a terminologia do ISIS), já que um diploma pode emanar da acção conjunta de mais de um órgão.

Finalmente, há três campos que se referem ao conteúdo do diploma.

80. RESUMO. Este campo obrigatório contém o texto completo do resumo do diploma. Em princípio, não são feitas correcções ao texto publicado, salvo em alguns casos em que se detecta um erro tipográfico, ou quando é necessário separar duas orações com um ponto (em lugar do travessão usado na publicação) para permitir alguns tipos de recuperação de dados (usando os recursos do ISIS para recuperação mediante proximidade de palavras contidas em uma oração).

O RESUMO é utilizado como base para a atribuição de descritores ao registo. Na medida do possível, usa-se um recurso do ISIS que permite criar descritores dentro de um texto. Por exemplo, se um texto menciona "...preço da gasolina...", basta digitar "...preço da <gasolina>..." para que o termo GASOLINA seja incorporado, automaticamente, ao índice geral de termos mantido pelo programa. Desse modo, um grande número de descritores é gerado a partir dos resumos, incluindo, por exemplo, nomes de empresas, siglas, nomes de proprietários, produtos, indústrias, acções administrativas (nomeação, demissão, intervenção, etc), nomes de províncias, etc. Cabe observar que esse recurso do ISIS é extremamente flexível, permitindo extrair descritores de parte de uma palavra (por ex., digitando "...empresas <sal>ineiras..." o programa extrairá o descritor SAL, que será incorporado ao índice geral).

90. DESCRITORES. Este campo é opcional. Usa-se para dar descritores adicionais aos que foram indicados no RESUMO. Na prática, este campo tem sido usado em quase todos os registos, particularmente para dar descritores ligados à actividade industrial (por ex., <INDÚSTRIA DO VIDRO>) ou relacionados com diversas acções administrativas (por ex., <NOMEAÇÃO>, <EXONERAÇÃO>, <INTERVENÇÃO DO ESTADO>, etc).

100. NOTA. Este campo é opcional. Usa-se para fornecer esclarecimentos essenciais sobre o diploma, já que, em alguns casos, o resumo não é suficiente para especificar o conteúdo do texto.

2.3 Captação e entrada de dados

Foram criadas as seguintes fichas para a captação de dados:

- 1) CAPA DE LOTE. Pode ser utilizada para encaminhar à digitação lotes de 50 fichas de entrada de dados.
- 2) FICHA GERAL DE ENTRADA DE DADOS. É utilizada para fornecer os dados correspondentes a um novo registo, ou a um registo que deve ser substancialmente alterado; pode também ser usada para indicar um registo que deve ser eliminado. Essa ficha aparece na página seguinte.
- 3) FICHA AUXILIAR DE ENTRADA DE DADOS. Pode ser utilizada para indicar actualizações de um único campo.

LEGIS : BASE DE DADOS DE LEGISLAÇÃO

FICHA GERAL DE ENTRADA DE DADOS

FICHA: ___ / ___ OPERADOR: _____ DATA: ___ / ___ / ___

OPERAÇÃO: ADICIONAR ___ ACTUALIZAR ___ ELIMINAR ___ MFN _____

FONTE DOCUMENTAL

10. FONTE: _____

20. ANO DE PUBLICAÇÃO: 19___

DADOS GERAIS SOBRE O DIPLOMA

30. TIPO: _____

40. NÚMERO: _____ 50. DATA: ___ / ___ 60. ANO: 19___

70. ORIGEM: _____

DETALHES

80. RESUMO: _____

90. DESCRITORES: _____

100. NOTA: _____

Na prática, até o momento, foi utilizada somente a FICHA GERAL DE ENTRADA DE DADOS.

Foi elaborada uma tela de entrada de dados. Devido à simplicidade dos conceitos envolvidos, não foi necessário criar instruções especiais de ajuda, se bem que o ISIS oferece facilidades nessa matéria.

2.4 Relatórios

Foram criados quatro formatos de relatório. Dois deles correspondem ao "modo de prova", onde aparecem todos os sinais auxiliares utilizados na entrada de dados (isso inclui, por exemplo, os símbolos < e > utilizados para delimitar os termos de indexação); esses formatos são utilizados, principalmente, para detectar erros durante a etapa de crítica visual dos dados. Os outros dois formatos correspondem ao "modo de dados", onde os símbolos especiais são eliminados ou substituídos por caracteres de pontuação, de modo que os textos aparecem numa forma de escrita mais corrente.

2.5 Índice

O instrumento essencial para acelerar a recuperação dos dados é o índice criado pelo programa. O programa mantém um único índice, com todos os termos extraídos de cada registo, a partir de uma tabela de selecção. Esse índice é periodicamente impresso, para auxiliar no processo de crítica dos dados. Cada termo do índice pode ter até 30 caracteres. Portanto, nos casos em que o nome de uma pessoa ou instituição tem mais de 30 caracteres, é necessário criar um descritor apropriado, usando alguma forma de abreviatura.

3. ASPECTOS OPERACIONAIS

3.1 Preparação de dados

Os textos são escolhidos por um advogado, basicamente a partir do exame de diários oficiais. Para cada texto é preparada uma ficha de entrada de dados, com o preenchimento manual dos campos apropriados, salvo o de resumo, que tem recebido um tratamento especial.

Devido ao tempo requerido e à possibilidade de introdução de erros, que estariam envolvidos na transcrição manual do resumo para a ficha de entrada de dados, tem sido usado, com bons resultados, o seguinte procedimento:

- 1) o texto seleccionado é fotocopiado com ampliação, para facilitar a leitura por parte do operador encarregado da digitação;
- 2) a fotocópia é recortada e colada na ficha, dentro do espaço reservado para o campo RESU 10.

Um aspecto delicado está relacionado com a atribuição de descritores. Inicialmente, o advogado que selecciona os textos

atribui descritores segundo critérios gerais, que foram estabelecidos no começo da actividade. Portanto, dentro desses critérios, há uma primeira etapa em que os descritores são atribuídos com bastante liberdade. Os dados assim preparados são entrados imediatamente para a base de dados. Numa segunda etapa, a partir de um relatório de prova contendo os novos registos, são feitas as seguintes operações:

- 1) correcção de erros;
- 2) atribuição de descritores com base na utilização de um tesouro (vocabulário controlado).

Esse procedimento foi utilizado para colocar os registos à disposição dos usuários no menor espaço de tempo, já que a escolha de descritores com base num tesouro é um processo relativamente demorado.

Decidiu-se iniciar a escolha de textos pelo ano de 1987, trabalhando depois com os anos mais recentes e com os anteriores a 1987. A escolha desse ano tem a ver com o início do Programa de Reabilitação Económica, o que torna o ano de 1987 num marco referencial histórico de inegável importância legislativa, com abundância de diplomas de natureza económica, de muita utilização pelos principais usuários da LEGIS.

3.2 Uso de tesouro

Um tesouro para processamento de informação é, essencialmente, um vocabulário controlado e hierárquico. Em princípio, só deveriam utilizar-se descritores que aparecessem no tesouro, salvo no que se refere a nomes próprios (pessoas, instituições, províncias, etc). Na prática, é comum que se faça necessário acrescentar alguns termos ("extensões"), para cobrir situações não previstas no vocabulário original.

A utilização de um vocabulário controlado torna-se necessária para evitar a explosão do índice (o que acontece quando os descritores são atribuídos em forma totalmente livre) e para facilitar a procura. Por exemplo, no caso de textos que façam referência a "tecidos", "empresas têxteis", "indústrias têxteis", etc, é atribuído o descritor "INDÚSTRIA TÊXTIL".

Existiu certa dificuldade para se escolher um tesouro que resultasse adequado para as finalidades do trabalho, que requer certo detalhamento em matéria de termos legais e industriais. Finalmente, decidiu-se pela adopção do Macrothesaurus [1], fornecido pela sede da ONUDI, como referência primária. Esse tesouro é utilizado em muitos sectores das Nações Unidas, existindo em espanhol, francês e inglês (uma versão em português estaria sendo preparada). Foram fornecidas as primeiras duas versões, de modo que optou-se por utilizar como base a versão em espanhol.

Corresponde assinalar que o Macrothesaurus foi desenvolvido para atender necessidades gerais de indexação, como as que podem surgir no tratamento de documentos administrativos ou em bibliotecas não especializadas. Como aparece na Introdução [1], a finalidade foi "criar uma linguagem documental que permita processar com a maior cobertura possível a informação relativa ao

desenvolvimento económico e social e seja um ponto de convergência dos tesouros sectoriais consagrados a aspectos mais específicos, como a agricultura, a indústria, o trabalho, a educação, a população, a ciência, a tecnologia, a cultura, a comunicação, a saúde ou o meio ambiente".

Encontrou-se que o Macrothesaurus é satisfatório para as necessidades da LEGIS, no que diz respeito aos termos relacionados com a indústria. É, no entanto, um pouco limitado no que diz respeito aos termos jurídicos; portanto, aceitou-se que, nessa área, seria necessário utilizar extensões, cuidadosamente controladas, do vocabulário fornecido pelo tesouro.

Cabe acrescentar que a experiência acumulada na utilização do Macrothesaurus indica que ele pode ser uma boa opção para indexar documentos processados dentro do MIE.

Como referência secundária tem sido consultado o Tesouro SPINES [2], em português, elaborado através de um projecto UNESCO/IBICT (Brasil)/JNICT (Portugal). Em geral, é muito similar ao Macrothesaurus, mas mostra-se menos satisfatório para a indexação de indústrias.

3.3 Recuperação de dados

O ISIS oferece diversos recursos para a recuperação de dados, que não apresentam maiores dificuldades de aprendizagem. Foi feita a adaptação requerida para suportar todos os caracteres do português, mediante a modificação de certas tabelas usadas pelo programa. Se bem que o ISIS utiliza uma forma padronizada para incluir termos no seu índice (por exemplo, o termo "nomeação" é gravado como "NOMEACAO"), as conversões necessárias são feitas automaticamente pelo programa, através das referidas tabelas.

Uma recuperação típica, baseada no uso de descritores existentes no índice, não demora mais que uns poucos segundos. É possível, no entanto, fazer consultas com termos de livre escolha; obviamente, essas consultas devem revisar cada registo e requerem mais tempo.

4. CONCLUSÃO

Na sua forma actual, LEGIS contém a maior parte de legislação relacionada com indústria e energia, aprovada a partir de 1974. Como em toda base de dados, surgem dois problemas essenciais:

- 1) a manutenção da base de dados (que envolve, principalmente, a captação de novos textos e a preparação, entrada e crítica dos dados);
- 2) a facilitação do acesso aos usuários principais (nomeadamente, funcionários do Departamento Jurídico e do GREI, no MIE).

Com respeito ao primeiro problema, entende-se que ele pode ser devidamente contornado mediante o envolvimento de um advogado por curtos períodos, dedicados à selecção de novos diplomas e preenchimento das fichas de entrada de dados. O restante das

tarefas relacionadas com a manutenção da base de dados poderá ser atendido pelo pessoal lotado no Centro de Informação Industrial, a ser criado no MIE. Evidentemente, surgirão consultas eventuais ao assessor jurídico, para dirimir diversas dúvidas.

Com respeito ao segundo problema, sem chegar a considerar a utilização de uma rede local, é possível solucioná-lo com a instalação de microcomputadores, contendo cópias da LEGIS, junto aos principais usuários. Essa solução torna-se viável devido à diminuição do preço das máquinas. Além disso, cabe lembrar que a entrada de dados na LEGIS realiza-se, principalmente, por lotes. Portanto, uma vez que um novo lote de registos foi entrado e criticado, no Centro de Informação Industrial, é possível instalar rapidamente cópias actualizadas da LEGIS, nos computadores dos usuários.

Obviamente, o uso de cópias do ISIS, que é distribuído gratuitamente pela UNESCO, vai requerer a obtenção de licenças adicionais para a utilização do programa, o que não oferece dificuldade.

Um benefício adicional ao desenvolvimento da LEGIS, consiste na experiência acumulada nos diversos aspectos relacionados com a criação e utilização de uma base de dados textuais. Outras bases de dados que podem ser obtidas da ONUDI utilizam as mesmas técnicas experimentadas no caso da LEGIS.

5. REFERÊNCIAS

- [1] NAÇÕES UNIDAS. MACROTHESAURUS para el procesamiento de la información relativa al desarrollo económico y social. 3ª ed. Publicação No. ST/ESA/160. Nova Iorque: Nações Unidas, 1985.
- [2] UNESCO. Tesouro SPINES. Versão preliminar. Brasília: UNESCO, 1988.
- [3] UNESCO. Mini-micro CDS/ISIS. Reference Manual (Version 2.3). Paris: UNESCO, 1989.