



## **Structuring and Organizing Nuclear Information Subject Analysis**

**Alexander Nevyjel**

Database Production and Development Group  
**INIS, IAEA**



## **INIS - International Nuclear Information System**



- ◆ **Information Reference Service**
- ◆ **Scientific and technical Literature**
- ◆ **Operated by IAEA**
- ◆ **Based on International Cooperation**
  
- ◆ **> 2.5 Mio. References**
- ◆ **since 1970**
- ◆ **annual growth 80.000 references**



## Subject Analysis

- ◆ Introduction
- ◆ Preparatory Analysis
- ◆ Subject Classification (Taxonomies)
- ◆ Abstracting
- ◆ The Thesaurus and its Structure
- ◆ Subject Indexing



## Introduction

- ◆ Each record of a bibliographic database consists of
  - ◆ **bibliographic description**  
descriptive cataloguing used to identify and locate the piece of literature
  - ◆ **subject description**  
means of identifying a document's information content so that database users can identify relevant records in the database which cite literature containing information in a particular subject area
    - Subject Classification
    - Subject Indexing
    - Abstracting

Document 3 of 4

Title Cesium-137 in soil texture fractions and its impact on Cesium-137 soil-to-plant transfer  
Author(s) Gerzabek, M.H.; Mohamad, S.A.; Mueck, K  
CA Oesterreichisches Forschungszentrum Seibersdorf GmbH (Austria)  
Source TYPE Report  
RP OEFZS--4632 {OEFZS4632}  
Jun 1992. 12 p  
Reprint from Commun. Soil Sci. Plant Anal., v. 23(3-4) p. 321-330 (1992)  
MF available from INIS under the Report Number

AV  
Abstract Field studies at two sites contaminated by the Chernobyl fallout showed <sup>137</sup>Cesium (Cs) soil-to-plant transfer factors in wheat, rye and potato. Transfer values ranged from 0.0017 (potato tuber) to 0.07 (wheat straw). Generally transfer coefficients in cereal grains and potato tubers were significantly below the values of the shoots. A comparison of the two sites led to the conclusion that for all plants investigated <sup>137</sup>Cs transfer factors were higher in Lower **Austria** (Calcic Chernozem) than in Upper **Austria** (Eutric Cambisol). The specific activities of the texture fractions of the two soil types increased from sand to silt and clay. In the Calcic Chernozem the ratio of the <sup>137</sup>Cs activity in the silt fraction to the total activity in the soil was considerably higher than in the Eutric Cambisol. At the same time extractability of <sup>137</sup>Cs from the silt fraction of the latter soil was clearly lower. Both results mainly were attributed to the differences between the soils according to the organic matter content of the silt fractions, the Calcic Chernozem being seven times higher. Therefore, the differences in the <sup>137</sup>Cs-soil-to-plant transfer can be attributed partly to these soil characteristics. (authors)

Descriptor(s)

DEI cesium 137; comparative evaluations; geographical variations; potatoes; radionuclide migration; rye; soils; wheat

DEC beta decay radioisotopes; beta-minus decay radioisotopes; cereals; cesium isotopes; environmental transport; evaluation; **food**; gramineae; intermediate mass nuclei; isotopes; liliopsida; magnoliophyta; mass transfer; nuclei; odd-even nuclei; plants; radioisotopes; tubers; variations; vegetables; years living radioisotopes

## Information Retrieval

- ◆ **Online Searching**
  - ◆ controlled terms --> subject indexing
  - ◆ free text words --> abstracting
- ◆ **Printed Product**
  - ◆ table of content --> subject categories
  - ◆ subject index --> subject headings
- ◆ **Select relevant references**
  - ◆ abstract, title, descriptors
- ◆ **Locating selected relevant pieces of literature**
  - ◆ descriptive cataloguing
- ◆ **Access to the full text**

## Preparatory Analysis

Input preparation require close collaboration between descriptive cataloguer and subject specialist in:

- ◆ selecting literature **relevant to the database scope** and checking whether the information content makes its entry into the database worthwhile
- ◆ deciding whether **subdivisions** should be made at the **bibliographic level** and each part treated as a separate piece of literature

## Selection of relevant Literature

- ◆ All the literature in the scope of the database should be included
- ◆ Within the scope the database include all relevant published material

*Rule: Select only those pieces of literature to which you can assign at least one appropriate subject category*

- ◆ In addition, it should be determined whether the information value of the piece of literature makes its entry into the database worthwhile. Non-scientific-technical publications, such as newspapers, newsletters, etc. should be excluded

## **Bibliographic Subdivision**

**Publications may consist of subunits such as chapters of a book or an annual report, or individual conference papers in conference proceedings.**

**For these publications the subject specialist must decide whether bibliographic subdivisions should be made and whether all or only some of the items should be analysed separately**

**A separate record for the entire publication must be supplied as a „lead“ record in addition to the „analytic“ records**

## **Subject Classification (Taxonomies)**

- ◆ **Checking of the relevance to the database**
  - **If a specific and appropriate subject category can be assigned to a piece of literature that contains worthwhile information, this piece of literature should be selected and reported to the database**
- ◆ **Retrieval tool**
  - **tool for increasing the relevance of retrieval by restricting the search strategy on specific concepts or by excluding other areas**
- ◆ **A scope description is given for each category**
- ◆ **The scope of INIS is defined by the scope of all categories**

## Assignment of Subject Categories (1)

- ◆ **Subject classifiers are expected to be subject specialists who identify the significant topics of each piece of literature and report the item to the database if it contains information that falls within its subject scope**

*Rule: Each piece of literature processed for the database must be assigned at least one subject category*

*Rule: Documents should be categorized under the most specific category appropriate to the topic treated by the piece of literature*

## Assignment of Subject Categories (2)

- ◆ **The main topic of the document determines the **primary subject category****
- ◆ **If there are other significant topics, one or more **secondary subject categories** can be assigned in addition**

*Rule: Categories with subcategories of the same hierarchy should not be assigned simultaneously to a piece of literature. Multiple subcategories of the same hierarchical level are permitted.*

## Abstracting

- ◆ **An abstract is a clear and concise statement of the content of a piece of literature, including the author's objective, methodology, results and conclusions**
- ◆ **Basis for free text retrieval**
- ◆ **Indication of the relevance of a retrieved record to the query**
- ◆ **Each input item should contain an English abstract (exception: short communications)**
- ◆ **Abstracts in other languages are optional**

## Thesaurus

**„A thesaurus is a terminological control device used in translating from the natural language of documents, indexers or users into a more constrained system language. It is a controlled and dynamic vocabulary of semantically and generically related terms which covers a specific domain of knowledge“**

**This definition has been adopted by UNESCO  
„Guidelines for the establishment and development of monolingual thesauri“, UNESCO, SC/W/255, Paris, September 1973**

## The Thesaurus and its Structure (1)

- ◆ A thesaurus is a tool in subject indexing for information retrieval
- ◆ Descriptors („controlled terms“) chosen from the thesaurus and assigned to a document should clearly indicate the information content of the piece of literature
- ◆ The meaning of descriptors must be well defined and unambiguous
- ◆ In case of ambiguity, controlled terms are defined implicitly by the structure of the terminology

## The Thesaurus and its Structure (2)

- ◆ For each alphabetical entry in the Thesaurus a „wordblock“ is given containing all terms associated with that particular entry
- ◆ Synonyms of controlled terms may be included in the thesaurus as „forbidden terms“ with a cross-reference
- ◆ Homonyms: Terms in natural language sometimes have several meanings. Descriptors are well defined in the following ways:
  - ◆ The descriptor is placed in its correct semantic context by the association with its forbidden, broader, narrower and related terms
  - ◆ A scope note provides the exact meaning which is intended

## Thesaurus Sample Entry

### **SOLUTIONS**

(For chemical solutions only. For mathematics see the word block of **MATHEMATICAL SOLUTIONS**.)

**BT1** homogenous mixtures

**BT2** mixtures

**BT3** dispersions

**NT1** aqueous solutions

**NT1** hypertonic solutions

**NT1** isotonic solutions

## The Thesaurus and its Structure (3)

<b>Relationship</b>	<b>Symb</b>	<b>Cross reference</b>
hierarchical	<b>BT</b>	broader term (level 1, 2, etc.)
hierarchical	<b>NT</b>	narrower term (level 1, 2, etc.)
affinitive	<b>RT</b>	related term
preferential	<b>UF</b>	used for (reciprocally <b>USE ...</b> )
preferential	<b>SF</b>	seen for (reciprocally <b>SEE ... OR ...</b> )

## Subject Indexing

- ◆ Subject indexing is the concurrent assignment of a set of descriptors to indicate the subject content of a piece of documents
- ◆ Subject indexing means analysing the information content of a piece of literature and expressing the meaningful information content in the language of the database using the controlled vocabulary of the Thesaurus
- ◆ **The purpose of subject indexing is to enable useful retrieval**

## Subject Indexing

- ◆ **Document Interpretation Rule**
  - ◆ *Choose such information items for indexing as you would yourself expect to find in the document if you were the user searching for that information*
- ◆ **Specificity Rule**
  - ◆ *Always use the most specific appropriate descriptor*
  - ◆ *Do not assign a descriptor and one of its broader terms to the same item*

## Procedures for Indexing

- ◆ Read title and abstract, scan full text
- ◆ Identify the concepts about which the piece of literature contains useful information
- ◆ Translate the concepts into descriptors
- ◆ Check each descriptor to make sure that
  - ◆ the descriptor represents precisely the concept
  - ◆ the definition matches the use
  - ◆ the selected descriptor is the most specific appropriate choice
- ◆ If no suitable descriptor exists for a useful concept, propose a new one
- ◆ Avoid overindexing

## Specific Indexing Rules (1)

- ◆ Elements and Isotopes
  - ◆ element name EUROPIUM
  - ◆ with mass number EUROPIUM 144
  - ◆ with term ISOTOPES EUROPIUM ISOTOPES
- ◆ Inorganic compounds and complexes
  - ◆ specific CARBON SULFIDES
  - ◆ broader terms CARBON COMPOUND SULFIDES
  - ◆ e.g. sodium oxalate should be indexed by SODIUM COMPOUNDS and OXALATES
  - ◆ e.g. sodium platincyanate should be indexed by SODIUM COMPOUNDS, PLATINUM COMPOUNDS and CYANATES

## Specific Indexing Rules (2)

- ◆ **Nuclear Reactions** (e.g.  $^{15}\text{N}(\text{p},\alpha\gamma)^{12}\text{C}$ )
  - ◆ target: NITROGEN 15 TARGET
  - ◆ specific reaction: PROTON REACTIONS
  - ◆ reaction products: ALPHA PARTICLES  
GAMMA RADIATION  
CARBON 12
  - ◆ Energy range: MEV RANGE 01-10
- ◆ **Reactor Names**
- ◆ **Geographical Designations**
  - ◆ Bombay is indexed as INDIA
  - ◆ Chicago is indexed as ILLINOIS
  - ◆ Gulf of Genoa is indexed as MEDITERRANEAN SEA

## Quick reference: CATEGORIZATION

- ◆ Each piece of literature must be assigned at least one subject category
- ◆ Documents should be categorized under the most specific category treated by the piece of literature
- ◆ Categories and subcategories belonging to the same hierarchy are usually not assigned simultaneously to a piece of literature

## **Quick reference: ABSTRACT PREPARATION**

- ◆ **Submit an informative abstract whenever possible**
- ◆ **Emphasise what is novel about the information in the document**
- ◆ **Do not repeat the title of the document in the body of the abstract**
- ◆ **Do not exceed 2000 characters (300-400 words) in length of the abstract**
- ◆ **At the end of the abstract state the number of references, figures, tables and the indication who wrote the abstract**

## **Quick reference: SUBJECT INDEXING**

- ◆ **Choose such information items for indexing as you would yourself expect to find in the piece of literature if you were the user searching for that information**
- ◆ **Use the most specific appropriate descriptor**
- ◆ **Do not assign a descriptor and one of its broader terms to the same item**
- ◆ **If no suitable descriptor exists, propose a new one**