Introduction

THES-BID is a structured *thesaurus* of descriptors (keywords and phrases) used in the «International Bibliography on Computers and Law» (BID) and its computerized data base for the indexing and retrieving of bibliographical material relating to computer science, its application to the law and to computer law problems. It is the first attempt at giving a general structural systemization to recurrent concepts within legal informatics and computer law. As such it is brought to the attention not only of documentation experts, the users of the Bibliography and the BID data base but also of all those who are in any way interested in this new discipline.

THES-BID was compiled with the aid of a computer by means of a specially prepared automated checking and printing procedure designed to facilitate the authors' work.

The description of the procedures and of the programmes – which were designed, using a new and original methodology, at the Istituto per la documentazione giuridica – will be the subject of a complete and articulated study to be published in a forthcoming issue of the journal «Informatica e diritto». The authors, in fact, believe that the research experience gained in the building of THES-BID could be used profitably in the building of other *thesauri* in various different disciplines.

In the first phase of the work, about 2.000 descriptors were chosen from the technical literature on the basis of the frequency in their use. Once normalized, they were numbered progressively and then classified with a code number taken from a special scheme, the *Classification Table*, also used for the classification of the bibliographical units in BID.

This data was collected together in an electronic file called the MASTER file, whilst non-preferred terms, as they were singled out, along with particular annotations (SCOPE NOTES) relating to the use of the descriptors, were registered in a separate file called the NOTE file.

In the second phase of the work, the most important relations of a vertical kind (conceptual hierarchy) and of a horizontal kind (synonymy, quasi-synonymy and other associative-type relations), existing between the various descriptors selected, were individuated. Moreover, the instructions to go from non-preferred terms to preferred terms within the *Thesaurus* and the connection between some particular annotations in the NOTE List and the descriptors to which they referred were decided upon. The relations be-

tween the descriptors (including the instructions and connectors), after being codified, were registered in an electronic file called the RELAT file. The computer, employing a set of specially designed programmes, carried out the first series of checks on these three files (MASTER, NOTE, RELAT), verifying the formal exactness of the classification and codification and eliminating eventual duplications in the strings.

Then, on the basis of the properties of the relational operators, defined in advance, the computer expanded the network of relations registered in the RELAT file. The expansion of the reciprocity, symmetry and transitivity of some operators, on the one hand and of the checking by the computer of irreflexivity, on the other hand, have permitted the number of originally defined relations, to be enlarged considerably and meanwhile the *Thesaurus* has been purified of inevitable errors. This automated expansion of relations has been particularly useful. The nearly 4.500 relations initially registered in the RELAT file have almost trippled with the help of the computer and the data obtained in this way has been organized in a new file, called the RELAT-ESP file.

The vertical hierarchic relations between the terms (identified by the operator BT, *Broader Term* and its reciprocal NT, *Narrower Term*) have been formulated as algebraic tree-structures, whose formal properties have been expanded and checked automatically. A tree-structure with only one root and numerous intermediate and terminal nodes has been constructed for every set of hierarchic chains having the same conceptual origin. Then, with the aid of the computer, each tree-structure was checked to make sure there were no «jumps» in the hierarchy during the passage from one node to another (in the BT/NT chain) and that each node has only one hierarchic superior (monohierarchic tree).

The horizontal relations of synonymy, quasi-synonymy and association between the terms (identified by the operator RT, *Related Term*, and expanded automatically according to the property of symmetry and directed transitivity) have, instead, been designed as lattice non-cyclic algebraic structures. They were also checked automatically in order to avoid logical errors and redundancy and, yet again with the aid of the computer, the observance of the rule of incompatibility between the relations of BT and that of RT for the same couple of terms has been verified.

$(x \text{ BT } y \wedge x \text{ RT } y \Rightarrow$ «error of direct incompatibility»;
$x \text{ BT } y \wedge y \text{ BT } z \wedge x \text{ RT } z \Rightarrow$ «error of indirect incompatibility»).

Finally, the computer was also used to organize graphically the printing of the numerous indices and lists which make up this edition of the *Thesaurus*.

KWOC Index of Permitted Terms

The *KWOC Index of Permitted Terms* allows us to go from the single words making up part of the phrases found in the *Thesaurus* to the com-

plete phrases. It is an alphabetical index built up according to the KWOC index technique: in the column on the left, the single terms are listed in alphabetical order as an index; in the column on the right the phrases from which the terms have been taken appear.

Consultation of the *KWOC Index* is the key to access to the *Structured Alphabetic List* for anyone who does not already know the order of the single words in the complex phrases found in the *Thesaurus*.

THE STRUCTURED ALPHABETIC LIST

The *Structured Alphabetic List* is the main index of the *Thesaurus* because it contains, in alphabetical order, all the keywords and phrases examined (including those non-preferred) with an indication of the different types of relations between them. The descriptors are followed, in brackets, by their respective codes of classification (on the basis of which they are set out in the *Class List*) and the relations between them are indicated using relational operators codified, according to common usage, as follows:

$$US = Use$$

A US B: the term A is a non-preferred term in the *Thesaurus*. The term B is suggested in its place. The US relation – defined (and checked by the computer) as irreflexive, asymmetrical and intransitive – has as its inverse (or reciprocal) relation UF, which has the same properties as US.

$$UF = Used for$$

B UF A: the term B has been used instead of the term A, which is a non-preferred term. The UF relation is expanded automatically by its reciprocity to US $(x\ US\ y \Rightarrow y\ UF\ x)$

$$TT = Top\ Term$$

A TT: the term A is defined as a *Top Term* whenever it is at the vertex of a conceptual pyramid or, in other words, whenever it is at the root of an algebraic tree-structure. The computer has permitted us to verify that only descriptors situated at the root of the trees are defined TT and, equally, that none of these descriptors are left without their appropriate definition as TT's. The TT relation is monadic having only one term on its left.

$$HT = Heading\ Term$$

B HT: the term B is defined as a *Nodal Heading* on the condition that it is a conceptually important node to which a substantial group of other terms (not less than 5) are joined. Even the HT relation, as that of TT, is monadic.

$$SC = Scope\ Note$$

A SC n: A is followed by a note (numbered progressively in the NOTE file and referred to by the number «n» made up of 5 figures on the right of SC) which clarifies its meaning and use. The SC relation is biadic, irreflexive, asymmetrical and intransitive.

<div align="center">BT = Broader Term</div>

A BT B: the term A has as its hierarchically superior concept the term B. The properties of the BT relation, which is also checked automatically, are irreflexivity, asymmetry and intransitivity. The inverse (or reciprocal) relation is the NT relation which has the same properties as the BT.

<div align="center">NT = Narrower Term</div>

B NT A: the term B has as its hierarchic inferior the term A. The NT relation is expanded automatically by reciprocity of the BT relation (x BT y $\Rightarrow$ y NT x)

<div align="center">RT = Related Term</div>

A RT B: the term B is associated with the term B (and symmetrically A is associated with B) because the two terms are in a relation of synonomy, quasi-synonomy or of generic association. The properties of symmetry and directed transitivity are applied automatically to the RT relation which is also irreflexive.

(symmetry)

$$x\ RT\ y\ \Rightarrow\ x\ RT\ y,\ \ y\ RT\ x$$

(direct. transitivity)

$$x\ RT\ y\ \wedge\ y\ RT\ z\ \Rightarrow\ x\ RT\ y,\ \ y\ RT\ x,\ \ y\ RT\ z,\ \ z\ RT\ y$$
$$x\ RT\ z,\ \ z\ RT\ x$$

(but)

$$x\ RT\ y\ \wedge\ z\ RT\ y\ \Rightarrow\ x\ RT\ y,\ \ y\ RT\ x,\ \ z\ RT\ y,\ \ y\ RT\ z$$

Finally, for the editing of the list under examination, a further operator called DL (*Delete*) was used. This allowed us to eliminate those relations of the RT type which, although produced by automatic expansion, were nevertheless judged as conceptually unacceptable. The DL relation whose properties are irreflexivity, symmetry and intransitivity, has always been registered in the RELAT file, but, being useful only in the compilation phase of the list, has, for obvious reasons, not been printed in the *Thesaurus*.

INDEX TO THE HIERARCHY

In the *Index to the Hierarchy,* the single descriptors (keywords and phrases) of the *Thesaurus* are set out in alphabetical order. They are completed with

their appropriate codes of classification. The root of the conceptual tree (or trees) where it can be found is indicated for each descriptor.

Therefore consultation of the *Index to the Hierarchy* permits the user to move up from a single descriptor to the *Top Term* in the hierarchic chain where the descriptor belongs.

LIST OF TOP TERMS (TT)

The *List of Top Terms* contains the list of descriptors which are at the root of the 52 conceptual trees making up the structure of the *Thesaurus*.

LIST OF HEADING TERMS (HT)

The *List of Heading Terms (HT)* is a list of the 156 descriptors which, in the structuring of the trees represent the conceptual nodes important for the number of relations dependent on them (not less than 5).

The list is made up of two parts, the first listing the descriptors in alphabetical order and the second dividing the descriptors into groups on the basis of the number of relations connected with them. Within each group, they are in alphabetical order.

HIERARCHY LIST

The *Hierarchy List* contains the description of the general hierarchic structure of this edition of the *Thesaurus*. In this list all the lexical material is grouped together and organized in 52 conceptual trees having as roots the same number of descriptors, considered of general semantic importance (see *List of Top Terms*).

The hierarchic level is graphically represented by a different number of indicator-dots preceeding the word or the phrase inside each tree. Therefore, in the search for the hierarchic superior or inferior of a particular concept it will be necessary to look up or down the list until reaching the row singled out by a number of indicator-dots – either one more or one less – than the concept started from. On the basis of the same criterion, the descriptors preceeded by the same number of indicator-dots are to be considered on the same level in the conceptual hierarchy.

By consulting the *Hierarchy List* the documentalist, who must index a given bibliographical unit, and the user of the *Bibliography* and the « BID » data base are given a sort of map of the field of knowledge where the subject matter, the object of his examination or research, has been placed. Therefore it is possible, according to one's needs using the hierarchy of concepts indicated in the list, to broaden or to narrow one's analysis in a systematic manner, testing efficaciously the choice of terms for indexing or for retrieving information.

CLASSIFICATION TABLE

The *Classification Table* is a decimal classification scheme which represents a systematic frame of reference for subject matter of importance to legal informatics and computer law. Each of the nine general classes (numbered from 0 to 8) making it up is internally structured in subclasses, in turn hierarchically organized according to their gradually increasing level of specificity. A code number of one more figures has been given to each hierarchically organized according to their gradually increasing level of specificity. A code number of one or more figures has been given to each through a process of systematic interpretation joining the heading taken into consideration with those which are gradually more general than it.
The constitutive elements of the *Classification Table* (the codes of classification and their corresponding headings) when related to a specific documentary unit permit the documentalist to arrange its contents within an organized system of knowledge.
Equally, in as far as this system of classification has also been linked to the single descriptors of the *Thesaurus* by the affixing of the relative codes to them, biunivocal cross-reference between keywords and the codes of classification and an analytical breakdown of information in both the indexing phase and in that of retrieval has been made possible.

CLASS LIST

In the *Class List* the headings of the *Classification Table* are to be found. The terms of the *Thesaurus* relating to each heading are listed under it, followed in brackets, if appropriate, by an indication of the other codes of classification where they are grouped together in the same manner. The terms are listed in two distinct groups, each list being in alphabetical order. The first group of terms includes descriptors taken from the *Classification Table* whilst the second group contains all the words and phrases registered in the *Thesaurus* which do not come directly from the *Classification Table* but directly from the technical literature examined.

GEOGRAPHIC LIST

The *Geographic List*, an alphabetic listing of the names of countries (or politically important groups of countries) used in the examined documentation to identify geographic areas where research or computer applications relevant to the Bibliography are being carried out. These names have been presented in a normalized from which is therefore binding on the indexer and on the researcher.
When the name of a country is followed by that of a state, province or region belonging to it, the two names are separated by a dash whilst the

names of the cities are indicated in brackets after the name of the country or the state, province or region to which they refer.

Acronim List (1 and 2)

The *Acronym List* is made up of abbreviations which frequently appear within the area of legal informatics or computer law and of the expressions from which these abbreviations are taken. Of the two lists in the *Acronym List* the first is based on the acronyms in alphabetical order and the second on the expressions relative to them.

The abbreviations contained in brackets when following expressions corresponding to institutes, public bodies, associations etc. indicate the country (according to international standards) in which their head office is situated or, where appropriate, their international character. (INT.). When, instead, we are dealing with the titles of journals, guides, inventories, and the like, they are indicated as pubblications (PUB.).

THE AUTHORS

Florence, June 1982