Leonardo Meo-Evoli, Giliola Negrini, Tamara Farnesi
CNR - Istituto di studi sulla ricerca e documentazione scientifica, Rome, Italy

ICC and ICS: Comparison and Relations between two Systems Based on Different Principles

Abstract: CIRT, the Italian Reference Centre for Technical and Scientific Terminology, is a network of sites designed to manage and retrieve terminological information. In view of the multiplicity of knowledge base structures which go together to form CIRT, as well as the need to preserve them, an interface system was developed for the integrated management of the different order systems. Work began with an analysis and comparison of the ICC and ICS classifications. Automatic procedures were defined to manage these classifications in an integrated way and to compare and correlate their classes. This paper describes the different structural characteristics of the two systems in question with a formal presentation of the typology of relations between classes, the problems that emerged during the research and the results achieved.

1. Introduction
Vocabularies, dictionaries, glossaries, thesauri and, in general, all terminological resources with a specific information content about a given sector are invaluable sources of knowledge and language. The structure of a data bank, or a terminological information network, has to enable the user to source the domain he is interested in. A subject is generally the principal key for accessing terminological resources in which other items - language/s, structure, the presence or otherwise of definitions and sources, format, authors, etc. - are also important. In order to be able to search for the domain using different interfaces, it is necessary to organise and manage all subjects. This is the aim of the comparative study that the ISRDS (Istituto di Studi sulla Ricerca e Documentazione Scientifica) responsible for CIRT, the Italian Reference Centre for Technical and Scientific Terminology, is carrying out on two major conceptual systems. CIRT, which is promoted by the Italian Association for Terminology, is made up of a network of terminological nodes interconnected in a telematic network. This includes the ISRDS itself, a National Research Council institute which acts as a central reference node, as well as a number of private-sector institutions. The Centre’s aim is to gather information about terminological resources and to manage and disseminate terminological information. It is an on-line reference system connected through the Internet with institutions in other countries and international bodies and networks which share the same aims.

(\text{http://www.isrds.rm.cnr.it/HypwerDocs/cirt.icirt.html}).

The study aims to produce an interface system for the integrated management of the classifications used to order the terminological resources encompassed by CIRT.

2. The classifications analysed
When CIRT was still at the design phase, the problem emerged of the multiplicity of knowledge-based structures. The need to preserve such structures prompted a project comparing the various systems and managing them in an integrated way. With this objective in mind, the ISRDS decided to organize the CIRT’s central reference structure, which it first
defined, and now manages according to a metasystem allowing the integrated management of
other classifications of the terminological resources encompassed by the CIRT.

The complex central ISRDS data bank was indexed following the principles on which
the metasystem is based. To survey the specific nature of the subjects to which the
terminological resources refer, terminological archives were classified following the different
criteria by which subjects are dealt with, and which may be of interest to the user for
information retrieval.

The work described below does not address the conceptual ordering of the subject
‘Terminology’ nor, as we shall see, does it deal with the typology of terminological sources,
but speaks exclusively about the fields of knowledge in which terminological resources are
available.

2.1 ICC: the conceptual reference system

ICC (Information Coding Classification) was adopted as a conceptual reference model.
Its order system is the result of a rigorously logical deductive formulation, which takes
ontological principles into account (Dahlberg, 1995b). The metasystem, designed and built by
Ingetraut Dahlberg, is based on the ‘Theory of Integrated Levels’ whereby nine levels of being
or categories of being (Form & Structure Area; Energy & Matter Area; Cosmo & Geo Area
etc) combine with as many aspects or categories of form, which represent the common order
principles. The concepts of each category of being may be clustered according to these
common principles. In turn, each concept cluster, which represents a knowledge field, may be
analysed according to the same principles, and may generate clusters of more specific
concepts. The process may thus be continued ad infinitum, organising knowledge fields at any
level according to a single model (Dahlberg, 1974, 1982).

The ICC was chosen on account of the following important features.

1) It enables us to identify the ‘subject field’ - i.e., the subject of the terminological data
   bank - not from a list of disciplines, but from primitive category levels. It is thus possible to
   review a subject from different points of view and represent the latter with a variety of
   notations. Albeit highly specific, the subject of a collection of terminology does not refer to a
   single discipline traditionally defined by consolidated divisions of learning. Indeed, precisely
   because it is so specific, it reveals an individuality of its own, expressed by concepts referring
to different disciplines, which may be clustered by various aspects and criteria but are
integrated in a unique way to the subject. E.g., Cereal cultivation, classified with the
notations: 465, Plant crops production; 43, Plant biology and cultivation; 412, Genetics; 484,
Food-based on plants.

2) It only contains classes which denote subject fields. These classes cover broader
   concepts relating to the broader concepts of other systems. This feature proves to be an
effective tool for ‘mapping’ the hierarchically higher classes of different classification
   systems.

3) The order principles on which the classification is based allow us to represent
   composite concepts according to a combinatorial procedure (e.g., Object +
   activity/application). This procedure is especially useful when new subjects - e.g.,
   Microbiology of Water - have to be introduced.

2.2 ICS: the International Classification for Standards

The CIRT data bank contains numerous glossaries classified according to ICS
(International Classification for Standards), produced by ISO. ICS is designed to classify all
special sectors of a technical nature disciplined by common standards. At this stage of the
project, class 01 Generalities. Terminology. Standardization. Documentation was reviewed.

Advances in Knowledge Organization, Vol.6(1998)
Both the classifications taken into account break down over three levels. The International Classification for Standards is based on inductive principles of an experimental nature, which offer a broad spectrum of very up-to-date fields of application. This 'open', highly specific structure allows us to 'arrange' new special subjects of a scientific-technical and experimental nature whose use and communication require common standards. The general composition of ICS would not appear to be based on clearcut, permanent structuring principles.

In principle, an inductive approach adopts an abstraction procedure to identify concepts which refer to more general common properties in concepts which represent concrete or abstract elements of reality. The concepts which represent these common properties show a higher conceptual level than the more specific ones. Hence, in ICS, concepts which identify properties common to more specific concepts flow from the third, most specific level and converge into a higher class, hence generating a hierarchy. It is then possible to identify conceptual elements common to these second-level classes to create a new hierarchical level, the first class.

In some cases, the inductive procedure is not altogether transparent. It is not always easy at the higher level to find sufficient concepts common to all underlying classes to create a single higher class. This means that several non-correlated concepts, which represent common elements of different groups of subclasses, end up in the same class coded with a single notation. Hence, for example, class 03 Sociology, Services, Company Organization and Management, Administration, Transport.

In our comparison and correlation of ICS with ICC, this particular feature created serious problems of univocality.

Other difficulties of comparison emerged in cases in which concepts identical at the lower levels were present in the higher class. E.g.,

39 Precision Mechanics, Jewellery
39.020 Precision Mechanics
39.040 Horology
...
39.060 Jewellery

It is worth remembering that, as things stand, ICS cannot be regarded as a universal classification, since it fails to take into account more strictly cultural and common aspects for which no standardization has yet been studied.

3. The problem of coding the formal aspect of the information source

If the information system manages miscellaneous sources (standards, books, periodicals, encyclopedias, glossaries, letters, etc.), it is sometimes vital to record the formal aspect of an information source. If the system manages only terminological resources, as in the case in point, the recording of form may instead be considered as accessory. Here it is unquestionably useful to distinguish lists, dictionaries, thesauri and so on, but such information is recorded inside the data bank structure and not included in the notation of the subject of the terminological resource.

The problem of recording the information source can be solved in two ways. A) The formal aspect is not highlighted if it refers to a system of homogeneous information (e.g., for terminological resources). In this case, the notation reveals only knowledge fields relative to these resources. B) The formal aspect is recorded in the notation anyway, if the classification system envisages identification of form and, subsequently, the disciplinary sector. In this case
the system envisages a faceted organization taking into account the formal aspect.

The classification systems considered respect these two different recording criteria. In order to compare such systems, they had to be made homogeneous.

Under *Vocabularies* (code 01.040), the ICS envisages a specific subdivision of the disciplinary sectors to which terminology refers and indicates form and content under a single code. This content corresponds to the first level of the whole classification.

To represent the formal aspect of the information source, the International Coding Classification uses a specific subdivision, *Form Divisions* (codes 01-09). The form 'terminology' is identified by notation 03 *Dictionaries, Terminologies*.

The solution adopted to make the structure of the two systems homogenous, hence comparable in accordance with the principles of the metasystem, was that of *Structural Composition* described below.

4. The requirements of the Formal Model

In order to compare classes and establish relations, it was necessary to find a system to manage ICC and ICS and define their relations using automatic procedures. The approach adopted was to define a semantic model which could subsequently be implemented with physical data structures. The method followed is summed up below. It was necessary:

- to establish the requirements for the representation of the two classifications and relative comparative relations;
- to formally define a semantic model (CoReC: Comparison and Relation of Classifications);
- to implement the semantic model with data structures using inexpensive software;
- to begin the process of analysing the concepts of the two classifications to define relations.

Let us now define requirements by typology.

4.1. Basic notions

**Classification**

A structured list of concepts which represents the semantic concept of a field of knowledge or universe of knowledge; concepts of a classification are connected hierarchically.

Let A be a classification: below we use \( \{a_1, a_2, \ldots, a_n\} \) to denote the set of concepts of classification A.

**Concept**

Represented by a description and identified by a code. The code facilitates the consultation of a classification and the identification of a concept.

**Specifications**

The set of concepts which descend hierarchically from a father concept.

Let \( a_1 \) be a father concept; the set of specifications of \( a_1 \) is \( s(a_1) = \{a_{11}, \ldots, a_{1m}\} \).

**Role of specifications**

In the ICC classification, every specification has a role which is different from that of its brothers (see par. 2.1). The ordering of specifications on the basis of the code represents this difference between the roles of specifications.

**Univocality of the parent**

A child concept can only have one father.

**Root**

The fatherless concept from which the whole hierarchical structure descends is the root of the whole concept hierarchy.
Plurality of children
A father concept may have more than one child concept.

4.2 Composition
The ICC's characteristic to identify new composite concepts (point 3 of par. 2.1) solves many of the problems of our work. In fact, in relating distinct classification concepts, it may happen that no concept in classification A corresponds to concept b in classification B. The need thus arises to build a new concept (i.e., not originally present) in A, assembling concepts a1 and a2 of A. The CoReC model thus has to support the Simple Composition operator. Furthermore, to make the structure of the two systems homogeneous (see paragraph 3), it was necessary to adopt Structural Composition.

Simple Composition
\{a_1,a_2,\ldots,a_n\} being the set of concepts of classification A, the composite concept is defined as \( a = K(a_j,a_k) \).

Semantic interpretation of a Simple Composition
The expression \( a = K(a_j,a_k) \) should be interpreted as: 'a IS_THE FIELD \( a_j \) APPLIED_TO \( a_k \).
From the previous expression it is possible to obtain the role of operands \( a_j \) and \( a_k \). The first operand identifies the object; hence, '\( a_j \) IS_AN OBJECT_OF a'.
The second operand identifies the application; hence, '\( a_k \) IS_ONE APPLICATION_OF \( a_j \).
For example, the concepts of ICC:
'General & theor. foundations of Hydrosphere' (ICC 351) and 'Microbiology and cultivat.' (ICC 428) may be composed to form the concept: 'Microbiology of water' (ICC 351:428). This concept is expressed formally as 351:428=K(351,428). The colon notation is the standard for ICC classification.
This latter expression should be interpreted as: 'Microbiology of water IS_THE FIELD General & theor. foundations of Hydrosphere APPLIED_TO Microbiology and cultivat.'
Hence: 'General & theor. foundation of Hydrosphere IS_AN OBJECT_OF Microbiology of water', or 'Microbiology and cultivat. IS_ONE APPLICATION_OF General & theor. foundation of Hydrosphere' (Fig. 2).

Structural Composition
Structural Composition creates a new concept which denotes form and content as introduced in par. 3). Structural Composition is an operator \( H \) which operates on two operands: \( a = H(a_j,a_k) \); \( a_j \) denotes the form of the information source and \( a_k \) denotes the subject/content. The result a of Structural Composition inherits the complete structure descending from the second operand.

Semantic interpretation of a Structural Composition
Let F3 be the ICC code for 'Dictionary, Terminology' concept, let S be the ICC code for the 'Subject Fields' concept\(^2\) and 'Terminology by ICC Subjects' be the concept name assigned to structural composition F3/S=H(F3,S). This expression should be interpreted as: 'Terminology by ICC Subjects IS_THE STRUCTURE Subject Fields OF Dictionary, Terminology'.
Note that the entire conceptual compatibility table between ICC and ICS was developed using Structural Composition concept F3/S=H(F3,S) (Figs 1 and 3).

Concept Type
The concept types that have to be distinguished are:
- Basic concepts: defined by the classification;
- Composite concepts: result of a Simple Composition;
- Form/Subject concepts: result of a Structural Composition. Composite concepts and Form/Subject concepts are part of the classification, but have to be identified as such. They may also be used to compose further concepts. Note that the ‘Simple Composition’ operator generates a concept devoid of specifications.

Adoption relationship

Composite concepts identify a knowledge field and as such may identify a set of descendant concepts. The possibility of identifying the descendants of a Composite concept which determines the transformation of the concept network from ‘tree’ to ‘graph’. In the CoReC model it is necessary to distinguish between the relationship between Basic concepts and relative descendants (a ‘natural’ relationship) and the relationship between Composite concept and its descendant (an ‘adoption’). The adoption relationship introduced here was not used in the research. It is formally defined in reference (Meo-Evoli L, Negrini G, 1998).

4.3 Relation between concepts of the two classifications

To compare the systems, their English-language published versions were used. Since no explicit definition exists of the concepts included in the two classifications3, we considered their hierarchical positions and, even more so, their specifications. The latter allowed us to identify the different semantic contents of the concepts under review, hence to compare them and establish relations between them.

Starting from the most general level of the two classifications, comparative relations emerged which were different from those usually attributed to concepts belonging to a single system.

Between a concept a of classification A and a concept b of classification B, the following types of relation may exist.

corresponds_to
a and b are said to be equivalent (i.e., a corresponds_to b), if they identify the same field of knowledge.
E.g., 01.040.79 Wood Technology (Vocabularies) corresponds_to F3/S477 Wood Technology.
“corresponds_to” is a semantic relation and does not fix any constraints on structures. This implies that if a corresponds_to b, this does not necessarily mean that there is an equivalence between their specifications; hence, in general, s(a)?s(b).
“corresponds_to” relation satisfies the reflexive property:
if a corresponds_to b, then b corresponds_to a.

includes
a includes b, if some of the knowledge present in a is lacking in the field of knowledge represented by b.
E.g., 01.040.93 Civil Engineering (Vocabularies) includes F3/S753 Construction Technology. In this case too, the relationship is semantic and it is impossible to formalise anything about their specifications.

is_included_in
is the opposite of the includes relation; i.e. if a includes b, then b is_included_in a.
The comparison phase revealed a high frequency of cases of ‘inclusion’ of specific ICS concepts in more general ICC fields.

is_about
is the relation which exists between two similar concepts; in general, if a is_about b, some sectors of a are not present in b and some sectors of b are not present in a.
E.g., 01.040.75 Petroleum and related technologies is_about F3/S376 Oil & gas extraction technology.

Advances in Knowledge Organization, Vol.6(1998)
"is_about" relation satisfies the reflexive property:
if a is about b, then b is about a.

union
The union of concepts relates a concept of classification B with the union of two or more concepts of classification A. By union of concepts, we intend the union definition of the set theory.

E.g., 01.040.83 Rubber and plastics industries (Vocabularies) union
(F3/S386 Organic materials, sci. and techn., F3/S387 Science & technology of plastics)
The union relation between concept b of B and the concept pair <a1,a2> of A is formally represented by b union (a1,a2).
For the union relation, the following property applies:
if b union (a1,a2), then a1 is_included_in b and a2 is_included_in b.

4.4 The CoReE model
The CoReE model, drawn up to represent classifications and comparison relations, is formally an oriented graph introduced in reference (Meo-Evoli L., Negrini G. 1998). This model was implemented with FileMaker Pro on a Macintosh. It was thus necessary to define a FileMaker Pro archive for every node typology envisaged by the model:
- ICC classification concepts;
- ICS classification concepts;
- Simple and Structural Compositions;
- adoptions;
- relations between concepts.
The edges of the CoReE graph were implemented with 'FileMaker Relation': this construct allows us to connect information from different physical archives.
We intend to develop CGI (Common Gateway Interface) programmes to make the results achieved available on the Internet. This is necessary to allow a user of CIRT to browse information in the different databases.

5. Results
A comparison table sums up the relations between the ICC and ICS concepts and their relative codes. The table can be drawn up in either systematic or alphabetic order.
It was noted that equivalent concepts are often represented by different denominations. In the alphabetic list, these denominations are addressed as synonyms: e.g., 'Road vehicle engineering', as a synonym of 'Road vehicle technology'.

<table>
<thead>
<tr>
<th>ICS code</th>
<th>ICS description [Vocabularies]</th>
<th>relations</th>
<th>ICS code</th>
<th>ICS description [Vocabularies]</th>
</tr>
</thead>
<tbody>
<tr>
<td>01.040.37</td>
<td>Image technology</td>
<td>union</td>
<td>F3/S572</td>
<td>Plastics including micrographic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>F3/S573</td>
<td>Printing</td>
</tr>
<tr>
<td>01.040.59</td>
<td>Precision mechanics</td>
<td>corresponds</td>
<td>F3/S746</td>
<td>Precision instrumentation</td>
</tr>
<tr>
<td></td>
<td>Jewellery</td>
<td></td>
<td>F3/S767</td>
<td>Engineering for work and leisure</td>
</tr>
<tr>
<td>01.046.43</td>
<td>Road vehicle engineering</td>
<td>corresponds</td>
<td>F3/S774</td>
<td>Road vehicle technology</td>
</tr>
<tr>
<td>01.046.35</td>
<td>Railway engineering</td>
<td>corresponds</td>
<td>F3/S775</td>
<td>Rail vehicle technol. (rolling stock)</td>
</tr>
<tr>
<td>01.046.37</td>
<td>Ship building and marine</td>
<td>corresponds</td>
<td>F3/S776</td>
<td>Ship and boat technology</td>
</tr>
<tr>
<td></td>
<td>structures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01.046.49</td>
<td>Aircraft and space vehicle</td>
<td>corresponds</td>
<td>F3/S777</td>
<td>Aircraft engineering</td>
</tr>
<tr>
<td></td>
<td>engineering</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1 ICS-ICC relations table

In Fig. 1 it is possible to note that 01.040.37 includes F3/S746:F3/S236. The concept

Advances in Knowledge Organization, Vol.6(1998)
F3/S746:F3/S236 is an example of the Simple Composition shown in Fig. 2.

![Fig. 2: example of Simple Composition](image1.png)

The semantic interpretation of Structural Composition is shown at the bottom of Fig. 3.

![Fig. 3: example of Structural Composition](image2.png)

The next step in the project will be to extend the comparative analysis of ICC and ICS to classes other than 01 Generalities. Terminology. Standardization. Documentation again using the five relation typologies identified in this paper. It is important to point out that these relation typologies may also be applied in comparisons between ICC and other classification systems. The definition of comparative relations should not present the difficulties which emerged in this work, if classification systems fail to highlight the formal aspect of the source of information.

It is also worth noting that the work carried out enabled us not only to achieve the scheduled result of producing a comparison table, but also to develop a special model which may be used to correlate different systems in the future.

Notes
1 Curly brackets identify a set of elements. \( \{a_1, a_2, \ldots, a_n\} \) should be read as 'the set of elements \( a_1, a_2, \ldots, a_n \). Note that it cannot be said that \( \{a_1, a_2, \ldots, a_n\} = A \) since the set \( \{a_1, a_2, \ldots, a_n\} \) represents not the hierarchical structure, but only the list of concepts.
2 A German-language definition (yet to be published) exists for all ICC concepts.
3 We introduce node S 'Subject Fields' to identify the complete structure of ICC.

References
Dahlberg, I. (1982). Information Coding Classification - principles, structure and application possibilities. Int. Classif. 9/2, p.87-93
1995, p.32-45


