Ingrid Meyer, Lynne Bowker and Karen Eck,
University of Ottawa, Canada

COGNITERM: An Experiment in Building a Terminological Knowledge Base

ABSTRACT: This paper discusses two aspects of an experimental terminological repository, called COGNITERM, which is a hybrid between a conventional term bank and a knowledge base. The first aspect is our methodology, essentially a blend between conventional terminology practise and knowledge engineering techniques and technology. The second aspect is the advantages that a terminological knowledge base offers over a conventional term bank, in terms of the data itself, as well as support for acquiring and retrieving the data. Our paper is illustrated with examples from COGNITERM and TERMIUM III.

0. Introduction

This paper is based on the assumption that term banks would be more useful, and useful to a wider variety of people, eventually even machines, if they contained a richer and more structured conceptual (i.e. knowledge) component than they do at present. Drawn to its logical conclusion, this assumption implies that term banks should evolve into knowledge bases, or at least contain a knowledge base component in addition to a conventional linguistic component. This vision of a hybrid between a term bank and a knowledge base, or *terminological knowledge base* (*TKB*), has recently been paralleled in computational lexicology by the concept of a lexical knowledge base (LKB), as discussed for example in Atkins 1991, Boguraev and Levin 1990, and Pustejovsky and Bergler 1991.

At the Artificial Intelligence Lab of the University of Ottawa, we have been interested in the concept of a TKB for the past three years[1], and have recently begun constructing a prototype TKB called COGNITERM as a testbed for our initial ideas. The basic assumption underlying our work is that a TKB must *represent what a native speaker who is also a subject-field expert knows about both concepts and their corresponding terms*. We feel that this assumption is in principle fully compatible with traditional terminological practise, not only as regards the "terms" component (obviously, term banks contain terms and other strictly linguistic information), but also as regards the "concepts" component. Although terminologists have not traditionally constructed *knowledge bases*, the acquisition of conceptual knowledge (from documents and/or experts) has always been the starting point for any terminology project, and an activity that goes hand-in-hand with all stages of the project (cf. section 1). Where our work differs from traditional terminology, however, is in the *degree of explicitness and structure* of the representation we are aiming at, namely a degree similar to that found in knowledge bases as they are known in Artificial Intelligence (AI). The difficulties that AI researchers have encountered in acquiring the knowledge contained in knowledge-based systems have given rise to a new area of AI research called *knowledge engineering*, which, as we have argued in more detail elsewhere[2], shares many of the conceptual analysis problems of terminology, and which – fortunately for terminology – has begun to develop tools to facilitate solving these problems.

Consequently, we are using one such tool, called CODE (Conceptually Oriented Description Environment)[3], to construct COGNITERM. CODE is essentially designed to provide various types of support for any person faced with the tasks of acquiring, structuring, debugging, revising – in other words, "managing" – information about the concepts of a specialized subject field. In a nutshell, CODE could be a described as a "concept manager" which is at the terminologist's elbow throughout a terminology project, from initial introductory reading on the field, to the selection of information sources (documents and subject-field experts), to in-depth analysis of concepts, to revision by other terminologists or subject-field experts. Currently, term and knowledge acquisition for COGNITERM are focussing on the subject field of optical storage technologies (e.g. optical media, devices, processes, standards), and at the time of writing, we have described several hundred concepts.

This paper is organized as follows. Part 1 provides a general overview of the methodology (still very much in progress) that we have developed to date. Part 2 outlines some of the advantages that a TKB offers over a conventional term bank, in terms of the data itself, as well as support for acquiring and retrieving this data. These advantages are illustrated by comparing examples from COGNITERM and TERMIUM III, the world's largest term bank maintained by the Canadian government.

1. Methodology

Although we are building something new (a hybrid between a term bank and a knowledge base), we have found it useful to adopt the following basic principle to guide the development of our methodology: *as far as possible, the methodology will be consistent with that traditionally used by terminologists*[4]. Our justification for adopting this principle is grounded in the assumption introduced above, namely that while terminologists do not build *knowledge-based* term banks, conceptual analysis has always been a crucial part of their work nonetheless: terminologists are keenly aware of the importance of a certain depth of understanding of the subject-field concepts, which implies an understanding of the interrelations between these concepts, for orienting the search for documentation and experts, judging the quality of documentation, communicating effectively with experts and other resource persons, identifying synonymy, constructing definitions, establishing interlinguistic equivalence, handling neology, revising and updating term records, etc.

On the other hand, what differentiates our methodology from the traditional one is that it aims at representing conceptual structures with a very high degree of explicitness and structure. In conventional term banks, of course, conceptual information is only implicitly available in the form of definitions, contexts, subject-field labels, etc. The degree of explicitness and structure we are aiming for has only recently become practically feasible with the availability of knowledge engineering tools. In a nutshell, then, our methodology might be described as being grounded in traditional terminological methodology, but *enhanced* whenever possible through knowledge engineering technology, in this case the CODE system. Since CODE is by design a very *generic* knowledge engineering tool, we hope that our methodology will be transportable to other knowledge engineering environments as well.

The basic goal of our methodology is to establish an explicit representation of the *conceptual structure* of a field. This conceptual structure is seen as a dynamic entity, into which new concepts are constantly integrated, and in which descriptions of existing concepts are constantly modified as the terminologist's understanding of the subject field deepens. In effect, the *structure* may be more or less *structured* at any given point in a project. Because of this understanding of the conceptual structure as highly dynamic, our methodology does not treat it as something that we merely *aim for*, but also something that we *work with* throughout a project. This implies that every new term is integrated into the conceptual structure *as soon as it is encountered*, and not just when we feel we really understand it "well enough". As a result, the conceptual descriptions in the TKB may exist in different degrees of completeness, since a terminologist's under-

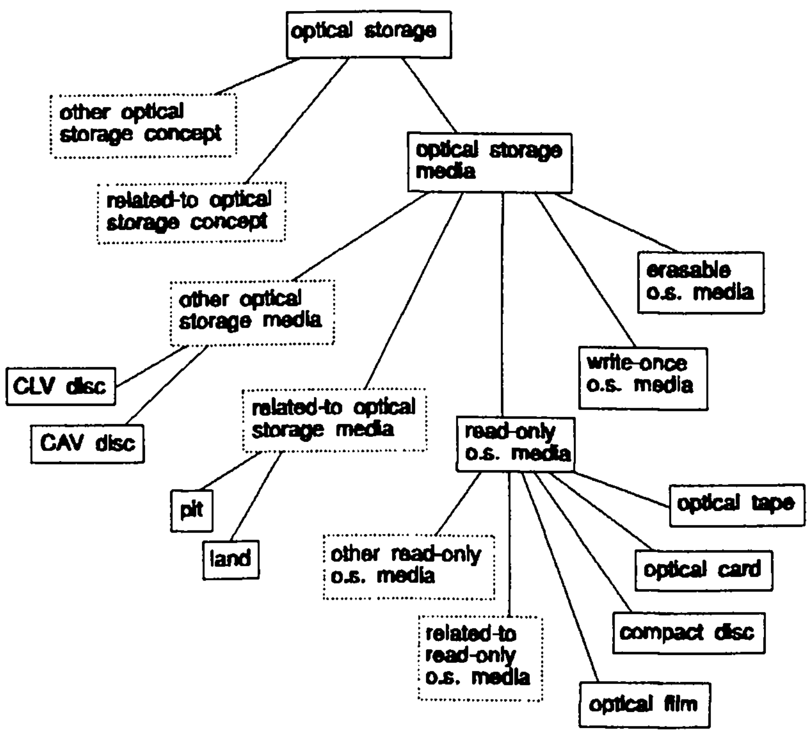

Figure 1. Dotted lines indicate "dummy concepts" that regroup unclassified concepts. Two categories of unclassified concepts exist to date: *other* unclassifieds indicate a generic-specific relation (e.g. a CLV disc is a kind of optical storage media, but that's all we know about it); *related-to* unclassifieds regroup all other relations (e.g. a pit is related somehow to an optical storage media, but we're not sure how).

standing of any concept will evolve gradually throughout the knowledge acquisition process. To indicate "degree of completeness" in our TKB, we have introduced the concept of "classification status", as illustrated in figure 1. Associated with this is the concept of "dummy concepts", which are created to regroup concepts whose classification status is unsure.

Our current methodology consists of the following general steps, which are described only briefly because of space constraints[5]:

1) Establishment of a rough conceptual map. We begin by doing general introductory reading on the subject field (which we roughly describe as optical storage), using the CODE system to draw an informal sketch of conceptual associations, a technique that has been advocated by numerous educational psychologists, as described in Sowa (in press). The goals of this step are 1) to identify a number of subfields, which show up as "concept islands" on the CODE graph, and 2) to establish the limits of the field (i.e. adjacent and related subfields that we decide not to handle). In our case, we identified the primary subfields of *media, devices, standards, software,* and *production techniques* at the outset of the project, and decided to exclude some related fields such as *magnetic storage.*

2) Establishment of a skeletal conceptual framework for the selected subfield. Once the conceptual map is sketched out, the subfield that appears the most fundamental to understanding all the others is selected for in-depth analysis. In our case, we selected *optical media,* which seemed to be not only the most fundamental to understanding the other subfields, but fortunately the most tractable as well[6]. Using a number of documents providing a general overview of this subfield, we identify its most generic concepts at this point, and design a template of conceptual characteristics common to these concepts. This template is a dynamic entity in the sense that it inherits to more specialized concepts, where it can acquire additional characteristics as well. The inherited characteristics are used as a guide to seeking out conceptual information in step 3.

3) Scanning of documentation and conceptual analysis. More documentation on the subfield is selected to supplement the very general documents used in step 2, and all the documentation is then scanned (i.e. terms recorded and contexts analyzed). In CODE, a conceptual descriptor (CD) is created for every concept. As figure 2 shows, a CD has a zone for Conceptual Information, which is the knowledge base component of COGNITERM. Here, information (always referenced as to its source) is entered in the following format: the *name* of the characteristic (e.g. *content*), followed by a *value* for this characteristic (e.g. *video*). Particular attention is paid to the template of characteristics established in step 2, though other characteristics are added as necessary. Every CD also has a zone for Linguistic Information, which is the term bank component of COGNITERM. Note that not all the conceptual information must be entered for every concept, as CODE features inheritance mechanisms[7] that allow more specific concepts to inherit information from more general ones. When a CD is created, the user is asked whether he/she knows what the more generic concept (i.e. superconcept) is. In this way, most concepts eventually become part of a hierarchical knowledge structure, which can be viewed graphically, as illustrated in figure 3. While the subfield illustrated in this figure (media) is dominated by hierarchical relations of the generic-specific type, another current subfield of interest (devices) appears to be more dominated by part-whole relations, and yet another (production processes) by stage-substage relations.

cdName: videodisc
super: optical disc, read-only optical disc
hasPropsOf:
kinds:
subConcepts: optical videodisc, misc videodisc
inheritPropsTo:
instanceOf:
instances:

SPECIAL
classification status: classified
creation date: 19 July 1991
D-disc by physical form:
D-disc by writability:
D-media by physical form:
D-media by writability:
done by: bowker

DEFINITION
intensional definition:

CONCEPTUAL INFORMATION
available recording surfaces: generally two, but sometimes only one.
content: video ; the range of information is taken from the section on videodiscs in CHEN89b p.15 since videodiscs are capable of holding the widest variety of information
degree of writability: one of: read-only, write-once, erasable
dimensions/diameter: 3.5, 5.25, 8, 12, or 14 inches
encoding method: digital or analog or both ; ELSHAMI90 pp.6&7 provides the property description
error correction: no ; RANADE90 p.40 states that error correction can be addressed by the disk drive itself and/or the file system software
laser type: gas laser or semiconducter laser
method of reading: the laser is either reflected off the surface of the disc, or passes through the disc
physical form: disc
recording technology: optical
rotation technique: one of: CAV, CLV ; more information available in "rotation technique", "CAV

LINGUISTIC INFORMATION
English synonym/syn1: LaserVision ; MATTHEWSON83 p.13 uses a lowercase "v" (i.e. Laservision)
English synonym/syn2: laser videodisc ; ELSHAMI90 p.8 notes that this term is used loosely as a synonym of videodisc
English synonym/syn3: LV videodisc
English synonym/syn4: videodisk
English synonym/syn5: VD
English synonym/syn6: Laserdisc
English synonym/syn7: LV-ROM
English term: videodisc
English textual support/def1: A 12-inch diameter disc which is used for recording video.
French synonym/syn1: disque vidéo [acceptability rating: unconfirmed / gender: masc] ; A file
French synonym/syn2: vidéodisque à laser [gender: masc]
French synonym/syn3: vidéodisque optique [gender: masc]
French synonym/syn4: vidéodisque LaserVision [gender: masc]

Figure 2. A CD (Concept Descriptor) for the concept VIDEODISC. The upper left-hand corner gives the name of the concept (cdName), its generic concept (super), and its subconcepts. The Conceptual Information zone is the knowledge base component of COGNITERM, while the Linguistic Information zone is the term bank component. There is no space limitation for any of the zones (they are scrollable).

4) *Revision and validation of information.* When the description of the subfield seems as complete[8] as practical constraints allow, the terminologist revises[9] the information using CODE's Browser, shown in figure 4. The Browser is a hypertext-like interface that allows the user to quickly view all the conceptual and linguistic information for a concept, and also to compare this information for different concepts. Both the Browser and the Graph (which can be open at the same time, as can any number of CDs) can be used to validate information with subject-field experts and other terminologists.

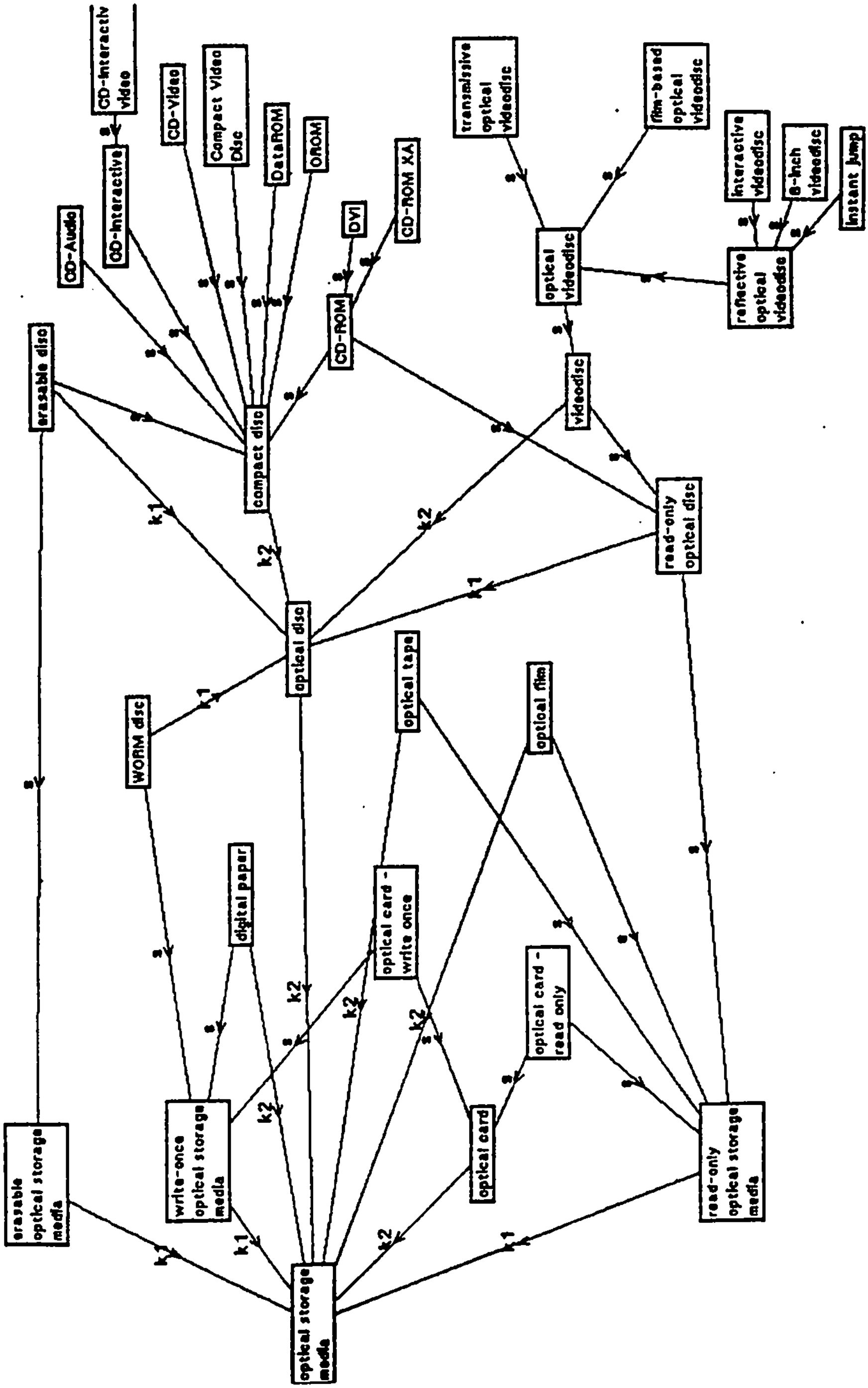

Figure 3. A Graphical Representation of the media subfield of the knowledge base. Arrows point towards more generic concepts. Links labelled "s" indicate a normal subconcept relationship. Links labelled "k" indicate a subconcept relationship affected by the presence of different dimensions (all concepts labelled k1 are in one dimension, k2 in another, etc.).

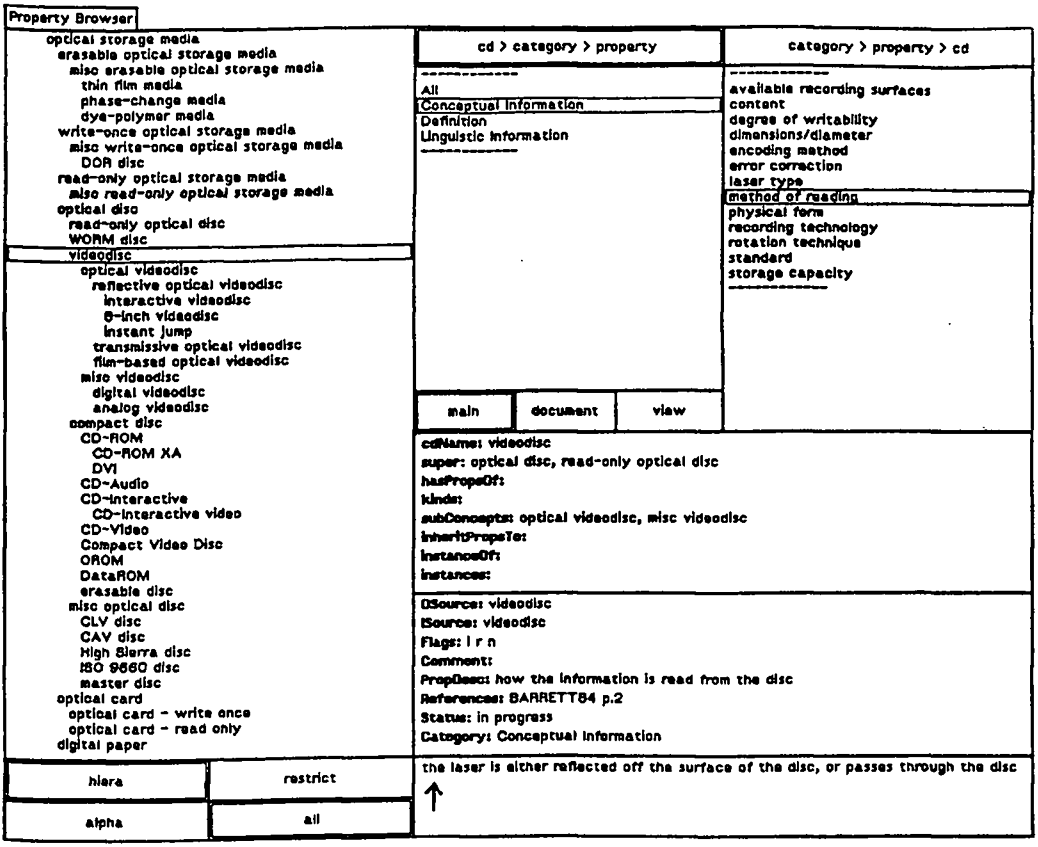

Figure 4. A Characteristic Browser. The left column contains all concepts in COGNITERM, with indentations indicating hierarchical organization. The concept VIDEODISC has been selected. The second column indicates the categories of information available for VIDEODISC. The Conceptual Information category is selected. The third column indicates all the Conceptual Information characteristics that exist for the concept VIDEODISC. The characteristic *method of reading* is selected. The value of this characteristic is indicated in the bottom-right zone of the browser (where the arrow is).

5) Computer-assisted definition construction. When the information for a subfield is felt to be reasonably complete and has been validated by at least one expert, the terminologist can set about constructing a definition for each concept. To date, we have worked only on intensional (i.e. logical, generic-specific) definitions, a task that is facilitated by the fact that the genus term is always explicitly recorded for every fully classified concept, and that the conceptual characteristics are systematically noted (i.e. with a view to consistency between co-ordinate concepts[10]). To assist the terminologist in determining the characteristics that differentiate the concept-to-be-defined from its co-ordinate concepts, CODE offers a feature called a Characteristic Comparison Matrix, illustrated in figure 5, that shows all co-ordinate concepts (listed at the tops of the columns) and the union of all their characteristics (at the left of the rows).

Once a given subfield has been completed to the terminologist's satisfaction, including definition construction and establishment of equivalent terms in the target language, the cycle is begun again with another subfield.

Property Comparison Matrix

Supers: READ-ONLY OPTICAL DISC	VIDEODISC	CD-ROM
available recording surfaces	generally two, but sometimes only one.	n/a
content	video	text, digital audio, video, graphics
degree of writability	one of: read-only, write-once, erasable	read-only
dimensions/diameter	3.5, 5.25, 8, 12, or 14 inches	4.72 inches (12 cm)
encoding method	digital or analog or both	digital
error correction	no	CIRC, Layered ECC
laser type	gas laser or semiconducter laser	n/a
method of reading	the laser is either reflected off the surface of the disc,	n/a
rotation technique	one of: CAV, CLV	CLV method
standard	one of: National Television System Committee (NTSC),	Yellow Book, High Sierra Group Standard
storage capacity	800 - 1000 MB	total: 600 MB, user data: 550 MB
introduction date	n/a	1983
observation	n/a	CD-ROM has to be used as an adjunct to mini- and

Figure 5. A Characteristic Comparison Matrix. The columns are headed by the names of all co-ordinate concepts of VIDEODISC, which is the concept-to-be-defined. The rows correspond to the union of the characteristics of all the co-ordinate concepts. "n/a" in a cell indicates that the characteristic does not exist for this concept.

2. A Comparison: Knowledge-based *vs.* Conventional Term Banks

As indicated above, COGNITERM is designed to provide all the strictly linguistic information found in a conventional term bank such as TERMIUM III. The reader can be quickly assured that this is true by looking at TERMIUM's bilingual record for the concept VIDEODISC, illustrated in figure 6, and comparing this with the COGNITERM CD for the same concept, in figure 2. In this section, we would like to address the question, *what advantages does COGNITERM offer that conventional term banks do not?* To illustrate some of these advantages, we will compare COGNITERM with TERMIUM III from three points of view: 1) the type of data that one can find in each, 2) the ways in which each provides support for acquiring and systematizing the data, and 3) the facilities that each provides for retrieving the data once it has been recorded.

DATA COLLECTION Terminology	FONDS Terminologie
FILE Single-concept File	FICHIER Fichier uninotionnel
SUBJECT FIELD(S) Storage Media (Data Processing)	DOMAINE(S) Supports d'information (Informatique)
EN	FR
videodisk *b;c *CORRECT,CORRECT videodisk *a *CORRECT,CORRECT	vidéodisque *g;d;e *CORRECT,CORRECT; MASC,MASC disque vidéo *f *A VÉRIFIER; UNCONFIRMED;MASC,MASC
DEF*A device, similar in appearance to a phonograph record, that contains audio and video material recorded on spiral or circular tracks.*b	DEF*Appareil de stockage sur disque d'images lues par laser. *d CONT*Le vidéodisque peut se comparer à un disque phonographique mais les signaux vidéo, en l'occurrence les images en mouvement, aussi bien que les signaux audio, sont encodés sur le vidéodisque. Les technologies du disque optique (semblables aux vidéodisques de type laser qu'on trouve sur le marché) peuvent constituer une alternative au stockage sur disques magnétiques. Les vidéodisques peuvent contenir beaucoup plus de données numérisées au pouce carré que les disques magnétiques.

Figure 6. Excerpt from the TERMIUM III record for VIDEODISC. "Single-concept File" means that this record is considered high quality (i.e. has undergone revision and is considered complete).

2.1. The Data

Quantity and consistency of conceptual information. By comparing the COGNITERM and TERMIUM data for the concept VIDEODISC, it is immediately obvious that COGNITERM provides much more conceptual information. This difference is, arguably, somewhat trivial, since in principle TERMIUM could have much longer definitions and/or contexts. The significant difference is that TERMIUM's conceptual information is embedded within free natural language text, while COGNITERM's information is presented in a highly structured characteristic-value format. In particular, COGNITERM makes very explicit the conceptual relations that apply to the concept. For example, as indicated in the top left zone of the CD, VIDEODISC has two generic concepts (= superconcepts, in this case optical disc, read-only optical disc). One might be able to extract this type of information from some TERMIUM definitions, but ultimately, acquiring a complete picture of the network of relations into which a concept enters would be like trying to put together the pieces of a very difficult puzzle, to use an analogy proposed by Kukulska-Hulme and Knowles (1989). One important by-product of the structure imposed by COGNITERM is *consistency*: for example, since generic concepts are explicitly indicated, definitions of all co-ordinate concepts must have the same genus term; since characteristics are automatically inherited to subconcepts, they will correspond from one co-ordinate concept to another; and so on.

Graphical as well as textual representation of conceptual information. As illustrated in figure 3, CODE provides a graphical representation of the subject field, normally in the form of a generic-specific, part-whole, or other type of hierarchy (though non-hierarchical relations can be graphed as well). This feature makes the data very attractive for learning purposes (for example, the terminologist who has just inherited a subject field from a predecessor, or the translator who needs to understand a field better before translating texts in it), as it provides a conceptual "map" of the field. The importance of conceptual relations to the understanding of concepts has been very aptly summarized by Sowa (in press):

> "None of these words [*sin, carburetor, tax shelter*] can be understood in isolation...The entry for sin, for example, might define it as a transgression against God. But that introduces the concepts of transgression and God. A transgression is a violation of a law, but that raises questions about how God gives laws and how they differ from human laws or laws of physics. A few more steps lead to the concepts of heaven and hell and eventually all of theology...In every field of human endeavour, from cooking and fashion to topology and quantum mechanics, the basic concepts can only be understood in relation to other concepts in tightly organized structures of thought. Knowledge acquisition may begin with words, but it must also find the connections that link those words in larger structures."

Possibility of multidimensional representation of reality. One of the fundamental problems with trying to describe conceptual structures is that a given field or subfield will often be divided up in different ways, depending on the point of view of the expert. For example, vehicles could be classified according to the characteristics *medium of transportation* (e.g. land, air, water), *type of propulsion* (e.g. motorized, non-motorized), and *principal type of load* (e.g. passenger, cargo). Some objects in a field such as this could thus be members of several dimensions simultaneously (e.g. a car could be classified as a kind of land vehicle, motorized vehicle, or passenger vehicle). The more levels one adds to a hierarchical classification, the greater the number of possible occurrences of multidimensionality, with the result that the conceptual structures can appear very complicated and messy indeed. COGNITERM offers a variety of mechanisms for managing the information overload that multidimensionality can present. For example, a masking capability will hide any selected "dimension(s)" in order to let the terminologist or end-user focus on just one at a time. Also, the characteristic underlying any given dimension is explicitly indicated and easily accessible. Dimensions can even be "ranked" from most to least important or frequent. Graphically, CODE indicates the presence of a multidimensional partition by a special link, labelled "k". For example, in figure 3, we see that the "world" of optical storage media can be partitioned in at least two different ways: first, according to the degree of writability of the media, indicated by the k1 links (*erasable* vs. *write-once* vs. *read-only* storage media), and second, according to the physical form of the media, indicated by the k2 links (*paper* vs. *card* vs. *disc* vs. *tape* vs. *film*).

2.2. Support for acquiring and systematizing the data

The COGNITERM environment (i.e. the CODE system) provides much more than a medium for *storing* data: it also offers a variety of mechanisms to help terminologists *acquire* and *systematize* data – in particular, conceptual data – in the first place. As we mentioned earlier, knowledge acquisition is by no means new to terminology: in our practical experience with working terminologists, we have seen terminologists make meticulous notes about individual concepts (very much like the characteristic-value notation we use), and sketch diagrams of conceptual relations (very much like our graph). COGNITERM offers the terminologist a variety of mechanisms to support the knowledge acquisition and systematization they have always done.

Inheritance mechanisms. On the simplest level, CODE's inheritance mechanisms free the terminologist from having to repeat information from one hierarchical level to another. On a more interesting level, inheritance offers the possibility of indicating to the terminologist the presence of inconsistencies: when a change is made at one hierarchical level, it will normally percolate throughout the knowledge structures, and CODE offers mechanisms for signalling undesirable repercussions (i.e. inconsistencies) to the terminologist. Another possibility offered by inheritance is that of doing "what-if" experiments (which will be familiar to anyone who has used a spreadsheet, for example). If the terminologist is not sure where a concept belongs, he/she can temporarily "attach" it to a part of the knowledge structure, see what is inherited, and better judge the appropriateness of this "place" in the knowledge structure.

Support for definition construction. The construction of sound terminological definitions is such a difficult task that in many working environments (e.g. the Secretary of State), terminologists are actually encouraged *not* to attempt their own definitions, but rather to extract or adapt them from documents. This can result in inconsistency, since definitions of related concepts may be taken from different experts (in person or from documents they have authored), and in poor quality, since experts, while they may know their field well, usually are not trained in linguistics, and certainly not in definition writing. The explicit rendering of conceptual information in COGNITERM provides the terminologist with a much better starting point, since the concept-to-be-defined will have an explicitly stated superconcept (hence the genus term for the definition), and the co-ordinate concepts will have corresponding characteristics, from which the differentia can be selected. Selection of differentia is facilitated by the Characteristic Comparison Matrix (cf. figure 5), which provides a clear overview of all co-ordinates and their characteristics. One can easily imagine (though it has not yet been implemented in CODE) a "dynamic" definition in the sense that when a differentiating characteristic is changed in the knowledge base, it is also changed in the definition (or at least, the terminologist is prompted to change it).

2.3. Retrieval of data

Conceptual vs. linguistic entry-points to the data. A severe limitation of conventional term banks is that the entry-points to the data they contain are essentially term-oriented: if one knows the term, one can expect the term bank to tell you what its French translation is, what grammatical peculiarities it exhibits, what its definition is, etc. However, one cannot easily find answers for concept-to-term oriented questions such as "What do you call the machine with function W?", "What do you call the material that has physical properties X, Y, and Z?", and so on, since the characteristics (i.e. W, X, Y, Z) are not explicitly recorded.

Hypertext-like browsing through the data. "Browsing" through conventional term banks is a rather cumbersome affair, and such browsing as is possible is, again, fundamentally terminological. For example, referring again to the record in figure 6, the user might not know what *tracks* means in the English definition, in which case he/she would have to go back to a central menu in TERMIUM III to look up this term. In the current version of CODE, the user could select this concept on the Browser (a much easier process than returning to TERMIUM's central menu), and in the forthcoming version of CODE, the user will have access to true hypertext features (e.g. all terms in a given CD that in turn have CDs of their own could be hypertext links). Any kind of extended browsing (particularly concept-oriented, rather than term-oriented) would be out of the question in TERMIUM. The COGNITERM Browser, on the other hand, assuming it were in the state shown in figure 4, would let the user make inquires such as "I wonder what the *method of reading* is for videodiscs in general?", simply by highlighting *videodisc* in the first column, and then "I wonder what a videodisc really is?", simply by selecting all of the conceptual characteristics for videodisc in turn; and so on. This kind of quick navigation, as we have mentioned above, makes the term bank a useful learning tool, and also a useful communication tool for interactions between experts and terminologists, or between various terminologists doing team work.

3. Summary

We have described the COGNITERM research project, whose basic purpose is to help us become clearer about (or "project", to use the terminology of Boguraev 1991:166) the concept of a TKB. Our approach to "getting clearer about the concept" of a TKB is to actually try to build one, and in this paper we have described a general methodology that combines aspects of traditional Terminology work with techniques and technology from Knowledge Engineering. While we feel that our concept of a TKB is still some distance from being fully "clear", we have also described some of the important advantages that such an artifact, even in its current primitive state of implementation, already offers over conventional term banks.

Acknowledgements

Various members of the COGNITERM research team have made valuable contributions to the work described in this paper, in particular Ken Iisaka, Timothy Lethbridge, and Douglas Skuce. The COGNITERM Project is supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) and Research Services of the University of Ottawa. Development of the CODE system is being supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the University Research Incentives Fund of the Government of Ontario, Bell Northern Research and Research Services of the University of Ottawa.

Endnotes

1 Cf. Meyer in press, Meyer and Skuce in press, Meyer and Paradis 1991, Skuce and Meyer 1990a/b. This past work essentially consisted in analyzing the conceptual analysis component of terminology work, developing CODE for terminological applications, and testing CODE and our preliminary methodology in a real terminology environment at the Dept. of the Secretary of State of Canada.

2 Cf. Meyer in press, Meyer and Paradis 1991.

3 Technical descriptions of CODE can be found in Skuce *et al.* 1989 and Skuce in press.

4 Since no internationally accepted generic methodology exists, we follow the one in use at the Terminology and Linguistic Services Directorate of the Dept. of the Secretary of State of Canada, as described in Cole 1987.

5 A detailed description of the COGNITERM methodology, and an analysis of outstanding methodological issues in TKB design, can be found in Meyer *et al.* 1992.

6 A matter which merits further research into the literature of educational psychology is what to do if the subfield that is the most fundamental to the subject field is not the most tractable. Might it be, for example, that for the acquisition of expert knowledge, doing the simplest thing is the most efficient approach under *any* circumstance?

7 Inheritance is a powerful AI technique that applies when concepts are arranged in a generic-specific hierarchy. Inheritance allows any characteristic of a given concept to be implicitly true for all specializations of this concept, for all specializations of these specializations, and so on.

8 The question of completeness is very complex in the context of a TKB, since both the type and amount of information required will vary according to the users one has in mind.

9 The revision of hypertext "documents" is a very new research problem to which we have no ready answers at all! To date, our revision has been done only intuitively.

10 Following the ISO International Standard on the vocabulary of terminology (ISO 1087), a coordinate concept is a concept "in a hierarchical system which ranks at the same level as one or more other concepts".

Bibliography

ATKINS, B.T.S. (1991): "Building a Lexicon: The Contribution of Lexicography". In: International Journal of Lexicography, Vol. 4, No. 3.

BOGURAEV, Branimir K. (1991): "Building a Lexicon: An Introduction". In: International Journal of Lexicography, Vol. 4, No. 3.

BOGURAEV, Branimir K. and LEVIN, Beth. (1990): "Models for Lexical Knowledge Bases". In: Proceedings of the Sixth Annual UW Centre for the New Oxford English Dictionary and Text Research. Waterloo, Canada: University of Waterloo.

COLE, Wayne. (1987): "Terminology: Principles and Methods". In: Computers and Translation, Vol. 2.

KUKULSKA-HULME, Agnes and KNOWLES, Frank. (1989): "L'organisation conceptuelle des dictionnaires automatiques pour textes techniques". In: META, Vol. 34, No. 3.

MEYER, Ingrid. (in press): "Concept Analysis, Terminology and the Document-Production Chain". In: Proceedings of the Symposium on Standardizing Terminology for Better Communication: Practice, Applied Theory and Results (Cleveland, Ohio, June 1991).

MEYER, Ingrid and PARADIS, Line. (1991): "Applying Knowledge Engineering Technology to Terminology: A Pilot Project". In: Terminology Update, Vol. 24, No. 2, pp. 3-8. (Publication of the Department of the Secretary of State of Canada)

MEYER, Ingrid and SKUCE, Douglas. (in press). "Computer-Assisted Concept Analysis for Terminology: A Framework for Technological and Methodological Research". In: Proceedings of the Euralex Fourth International Conference (Malaga, Spain, Aug. 28-Sept. 1, 1990).

MEYER, Ingrid, SKUCE, Douglas, BOWKER, Lynne and ECK, Karen. (1992): "COGNITERM: Structure and Methodology". Technical Report of the Artificial Intelligence Lab, Dept. of Computer Science, University of Ottawa, Canada.

PUSTEJOVSKY, James, and BERGLER, Sabine (Eds.). (1991): Proceedings of the ACL SIG Workshop on Lexical Semantics and Knowledge Representation. (To appear as a book published by Springer Verlag).

SKUCE, Douglas. (in press). "A Wide Spectrum Knowledge Management System". In: Knowledge Acquisition.

SKUCE, Douglas and MEYER, Ingrid. (1990a): "Computer-Assisted Concept Analysis: An Essential Component of a Terminologist's Workstation". In: Proceedings of the Second International Congress on Terminology and Knowledge Engineering Applications, Frankfurt: Indeks Verlag, pp. 187-199.

SKUCE, Douglas and MEYER, Ingrid. (1990b): "Concept Analysis and Terminology: A Knowledge-Based Approach to Documentation". In: Proceedings of the Thirteenth International Conference on Computational Linguistics (COLING 90), pp. 56-58.

SKUCE, Douglas, WANG, S., and BEAUVILLÉ Y. (1989): "A Generic Knowledge Acquisition Environment for Conceptual and Ontological Analysis". In: Proceedings of the Knowledge Acquisition for Knowledge-Based Systems Workshop (Banff, Canada, Oct. 1989).

SOWA, John. (in press): "Conceptual Analysis as a Basic for Knowledge Acquisition". In: The Cognition of Experts: Psychological Research and Empirical AI, R. R. Hoffmann (Ed.), Berlin: Springer Verlag.

KEYWORDS: terminology, term bank, Artificial Intelligence, knowledge engineering, knowledge base