GUIDELINES FOR HANDLING MULTIDIMENSIONALITY IN A TERMINOLOGICAL KNOWLEDGE BASE

by

Lynne Bowker

School of Translators and Interpreters
University of Ottawa

Under the supervision of

Ingrid Meyer, Ph.D.
School of Translators and Interpreters
and
Douglas Skuce, Ph.D.
Department of Computer Science

Thesis submitted to
the School of Graduate Studies and Research
of the University of Ottawa
in partial fulfilment of the requirements
for the degree of M.A. (Translation)

ISBN 0-315-79998-6

Acknowledgements

First and foremost, I would like to express my sincere thanks to my thesis supervisor, Dr. Ingrid Meyer, for her constant encouragement, and her prompt and insightful feedback on all aspects of my work.

I would also like to thank my co-director, Dr. Douglas Skuce, for providing me with valuable comments, and for letting me use the CODE system, as well as the technical facilities in the Artificial Intelligence Laboratory of the Department of Computer Science at the University of Ottawa.

I am grateful to Dr. Geneviève Mareschal for reading this thesis and offering many constructive comments and suggestions. I would also like to thank Dr. Roda Roberts, who provided helpful comments on several papers that served as starting points for some of the chapters of this thesis.

I am indebted to the following subject-field experts, who patiently and willingly helped me to learn about their fields: Mr. Jean-Pierre Artigau, former terminologist with the Department of the Secretary of State of Canada (biological taxonomy); Dr. François Chapleau, Biology Department, University of Ottawa (biological taxonomy); Mrs. Cécile Prud'Homme, reference librarian, University of Ottawa (library science); Professor Nancy Williamson, Faculty of Library and Information Science, University of Toronto, and consulting editor to the journal *International Classification* (library science); Mr. Don Slaunwhite, software developer for the CD-ROM division of Corel Systems Corp. (optical storage technology); and Mr. Timothy Lethbridge, Ph.D. student, Department of Computer Science, University of Ottawa (hypertext).

I would like to thank the University of Ottawa for providing me with a scholarship that helped to make it possible for me to follow this degree. I would also like to thank the Social Sciences and Humanities Research Council (SSHRC) of Canada for funding the COGNITERM Project, as some of the work carried out within the framework of this project served as a basis for developing this thesis.

On a more personal level, I would like to thank my good friend and Masters cohort Karen (K.R.Y.) Eck. In addition to providing valuable comments on many aspects of this thesis, she was a constant source of moral support, and most importantly, she helped me hold on to my sense of humour. ☺

Others to be thanked for their encouragement and support include my parents, Joyce and Keith Bowker, and sister, Lisa Bowker — I'm sure Bell Canada appreciated their long distance support as much as I did! Honourable mention also goes out to my good friends Sanjay Rao and David Buchler, just for putting up with me.

Finally, I would like to thank the Ramones: they believe in miracles, and they gave me something to believe in.

Abstract

The goal of this thesis is to *develop* and *apply* a set of guidelines for handling multidimensionality in a *terminological knowledge base (TKB)*. By TKB we mean a hybrid between a conventional term bank and a knowledge base as this is known in artificial intelligence (AI).

Multidimensionality is a phenomenon of classification. Classification involves uniting like objects on the basis of a common characteristic. If more than one differentiating characteristic can be used to distinguish the objects, then the objects can be classified in more than one way. A *dimension* represents one way of classifying a group of objects; a classification with more than one dimension is said to be *multidimensional*.

The recognition and representation of multidimensionality, though highly important to terminology work, is a subject that has received very little attention in the terminology literature. One of the reasons for this neglect may be that multidimensionality is a potentially complex phenomenon, and until recently, terminologists did not have adequate tools for dealing with it. However, such tools are now being developed in the field of AI. One such tool is *CODE (Conceptually Oriented Description Environment)*, with which we have been working at the AI Lab of the University of Ottawa.

This thesis is divided into four main parts. In Part I, we discuss the general principles of classification, and explain multidimensionality.

In Part II, we develop an initial set of guidelines to help terminologists both recognize and represent multidimensionality in a TKB. The guidelines are based on 1) useful insights gleaned from an analysis of classification as practised in the disciplines of terminology, biology, and library science, and 2) practical experimentation with a TKB on the subject field of optical storage media.

In Part III, we develop a technical complement to the initial guidelines. We begin with a general description of the CODE system, and then we analyze those features that are particularly helpful for handling multidimensionality. The analysis is illustrated with a step-by-step example of how to use CODE to create a multidimensional representation.

Finally, in Part IV, we apply our guidelines by using the CODE system to construct a small TKB for concepts in a subfield of hypertext, namely hypertext links. The thesis concludes with discussions of some issues raised during the application of the proposed guidelines, recommendations for possible technical modifications to the CODE system that will allow it to better handle multidimensionality in future versions, and suggestions for possible areas for future research in multidimensionality.

THE FAR SIDE By GARY LARSON

Same planet, different worlds

Not *exactly* multidimensionality, but close...

Table of Contents

List of Figures

List of Tables

Notation Conventions

The following notation conventions will be used throughout this thesis:

1. New terms, many of which are also defined in the glossary, will be given in italics the first time they are used. Italics will also be used to show emphasis.

   e.g. A characteristic that is used to distinguish the subclasses of one class is called a *differentiating characteristic*.

   e.g. The initial guidelines are intended to help terminologists both *recognize* and *represent* multidimensionality.

2. Names of classes and subclasses will be given in capital letters.

   e.g. The class VEHICLE can be subclassified into three subclasses: LAND VEHICLE, AIR VEHICLE, and WATER VEHICLE.

3. Names of concepts being referred to only as concepts, and not as classes, will be given in bold 11 point Times Roman font.

   e.g. OPTICAL STORAGE MEDIA represents the generic class to which **optical disc** belongs.

4. Names of characteristics will be given in bold 11 point Helvetica font.

   e.g. OPTICAL DISC can be classified into READ-ONLY DISC, WRITE-ONCE DISC and ERASABLE DISC on the basis of the characteristic **degree of writability**.

5. Values of characteristics will be given in bold italicized 11 point Helvetica font.

   e.g. The value of the characteristic **medium of transportation** can be said to be ***water*** for the class WATER VEHICLE.

6. Explanations of meaning will be enclosed within single quotation marks.

   e.g. The term *simple* is used in the library science literature in the sense of 'single' and not the sense of 'easy to understand'.

<u>Glossary</u>

NOTE: words appearing in **bold** within a definition have their own entry in the glossary.

AI: see **artificial intelligence**.

analogous character: in biological classification, a **character** which has a common function with other characters, but which was not derived from a common ancestor. e.g. the wings of a bird and the wings of an insect; compare with **homologous character**.

artificial intelligence (AI): the subfield of computer science concerned with designing intelligent computer systems (i.e. systems that exhibit the characteristics we associate with intelligence in human behaviour, such as understanding language, learning, reasoning, solving problems, etc.).

character: in biology, the term used to designate **characteristic**; see **characteristic**.

characteristic: a quality that distinguishes or identifies an **object**. A characteristic has two components: a name and a value. e.g. the name of one of the characteristics of the object optical disc is **physical form**, and the value of the characteristic is ***disc***. (ISO uses the term *property* when referring to objects, and *characteristic* when referring to concepts. For the sake of simplicity, we have used characteristic when referring to both objects and concepts).

cladistics: in biology, a system of classification based solely on the evolutionary history of organisms; also known as *phylogenetic systematics*; compare with **evolutionary systematics** and **phenetics**.

cladogram: in biology, a branching diagram used to illustrate cladistic classifications; compare with **phenogram** and **phylogenetic tree**.

class: a group of **objects** sharing at least one common **characteristic**.

classification: a systematic arrangement of **objects** produced by grouping the objects on the basis of common **characteristics**; the process of producing such a systematic arrangement.

classify: to group like **objects** together on the basis of common **characteristics**.

complex subject: in library science, a **subject** that consists of more than one **concept** and could be classified in more than one place in the library. e.g. *Teaching Child Psychology*; compare with **simple subject**.

concept: a unit of thought constituted through abstraction on the basis of properties (see **characteristic**) common to a set of **objects** (ISO 1087: 1).

concept system: a structured set of **concepts** established according to the relations between them, each concept being determined by its position in this set (ISO 1087: 4).

co-ordinate concept: a concept in a hierarchical system which ranks at the same level as one or more other **concepts** (ISO 1087: 2). e.g. COMPACT DISC and VIDEODISC are co-ordinate concepts, since they are both a kind of OPTICAL DISC.

derived character: in **cladistics**, a **character** that is shared with a recent common ancestor. e.g. having hair; compare with **primitive character**.

differentia: the set of all **differentiating characteristics** that are presented in a definition.

differentiating characteristic: a **characteristic** that is used to distinguish the **subclasses** of one **class**. e.g. AIR VEHICLE, WATER VEHICLE and LAND VEHICLE are specializations of VEHICLE that have been separated from one another on the basis of the characteristic **medium of transportation**; therefore, we can identify **medium of transportation** as the differentiating characteristic at this level of the **classification**.

dimension: a classification that has been produced by classifying **objects** on the basis of one particular **differentiating characteristic**; Sager (1990: 31) refers to dimensions as *facets*.

enumerative classification: in library science, a hierarchical classification system which attempts to list all possible **subjects**, both simple and complex, within a defined subject field; compare with **faceted classification**.

evolutionary systematics: in biology, a system of classification that considers both evolutionary history and morphological similarities when classifying living organisms; compare with **phenetics** and **cladistics**.

facet: in library science, the term used to designate **dimension** in a **faceted classification**; see **dimension**.

faceted classification: in library science, a classification system whereby **subjects** are analyzed into their constituent concepts and then grouped into **facets**. The concepts can then be combined to form **complex subjects**; compare with **enumerative classification**.

false multidimensionality: a phenomenon that occurs when **objects** are classified on the basis of more than one **characteristic** at the same level and within the same **dimension**.

generic concept: a **concept** that has been classified into a number of **specific concepts**; see also, **superordinate concept**.

generic relation: a **hierarchical relation** which is based on the partial identity of the intensions of **generic, specific** and **co-ordinate concepts** (ISO 1087); sometimes called an *isa* or *type* relation because the specific concept *is a type* of the generic concept. e.g. a COMPACT DISC *is a type* of OPTICAL DISC.

hierarchical relation: a relation between **concepts** which is established by division of a **superordinate concept** into **subordinate concepts** forming one or more levels, or by the reverse process (ISO 1087: 3); a **generic relation** is one type of hierarchical relation.

hierarchy: a classification based on **hierarchical relations**.

homologous character: in biology, a **character** which derives from the same ancestor as another character, regardless of whether the two characters have a similar function. e.g. the forearm of a man and the forelimb of a dog, or the wing of a bird and the fin of a whale; compare with **analogous character**.

hypertext system: a computer system which mimics the brain's ability to access information quickly and intuitively by reference; users can connect screens of information using associative links.

inheritance: in **AI**, a powerful technique that applies when **concepts** are arranged in a generic hierarchy: it allows any **characteristic** of a given concept to be implicitly true for all **subordinate concepts** of this concept, for all the subordinate concepts of the subordinate concepts, etc. e.g. if the concept ALCOHOLIC BEVERAGE has a characteristic **alcohol content**, then its subordinate concepts (FERMENTED ALCOHOLIC BEVERAGE and DISTILLED ALCOHOLIC BEVERAGE) will inherit this characteristic, as will all their subordinate concepts (e.g. BEER, WINE, VODKA, RUM, WHISKEY).

intension: the set of characteristics which constitutes a **concept** (ISO 1087: 2).

KB: see **knowledge base**.

knowledge base (KB): in **AI**, a knowledge base is a specialized type of database that contains a highly structured set of knowledge particular to a subject field.

knowledge engineering: a subfield of **artificial intelligence (AI)** that is primarily concerned with acquiring, formalizing, and refining knowledge so that it can be used by machines.

morphology: in biology, the form and structure of a living organism.

multidimensional representation: a representation of a **concept system** which shows the classification of **concepts** according to all possible **characteristics** at the same level, but in different **dimensions**; Picht and Draskau (1985: 68–70) refer to multidimensional representations as *polydimensional* and *polyhierarchic systems*.

multidimensionality: a phenomenon that occurs when an **object** can be classified according to more than one **characteristic** at the same level of a classification. e.g. VEHICLE can be classified into LAND VEHICLE, AIR VEHICLE, and WATER VEHICLE (according to the characteristic **place of transportation**), or VEHICLE can be classified into MOTORIZED VEHICLE and NON-MOTORIZED VEHICLE (according to the characteristic **type of propulsion**).

object: any part of the perceivable or conceivable world. NOTE — Objects may also be material (e.g. engine) or immaterial (e.g. magnetism) (ISO 1087: 1).

outgroup: a genus or species only distantly related to another group which is used to determine **primitive characters** in **cladistic** classification.

phenetics: in biology, a classification system based on the total similarities and dissimilarities of living organisms; evolutionary history is not taken into consideration; compare with **evolutionary systematics** and **cladistics**.

phenogram: in biology, a branching diagram with a percentage scale used to illustrate **phenetic** classifications; compare with **cladogram** and **phylogenetic tree**.

phylogenetic systematics: see **cladistics**.

phylogenetic tree: in biology, a branching diagram with a geological time scale used to illustrate evolutionary relationships of species; compare with **cladogram** and **phenogram**.

phylogeny: the evolutionary history of a genetically related group of organisms.

primitive character: a **character** shared with a distant common ancestor. e.g. having DNA as a genetic material; compare with **derived character**.

property: see **characteristic**.

schedule: in library science, the hierarchical index of **subjects**.

simple subject: in library science, a **simple subject** is one that consists of only one **concept** and can be filed in only one place. e.g. *Psychology*; compare with **complex subject**.

specific concept: a **concept** that belongs to a larger **class**; this concept that has all the **characteristics** of its **generic concept** plus at least one additional or modified characteristic; see also **subordinate concept**.

subclass: a primary division of a **class**.

subordinate concept: a **concept** in a hierarchical system which can be grouped together with at least one more concept of the same level to form a higher ranking concept (ISO 1087: 2); e.g. AIRPLANE and HELICOPTER are subordinate concepts of AIR VEHICLE.

subject: in library science, the theme or topic of a document. e.g. *A Comparison of the Respiratory Systems of Land and Water Vertebrates*.

superordinate concept: a **concept** in a hierarchical system which can be subdivided into a number of lower-ranking concepts (ISO 1087: 1); e.g. OPTICAL DISC is the superordinate concept of COMPACT DISC and VIDEODISC.

taxonomic category: in biology, one of seven major categories to which groups of plants or animals are assigned for the purposes of **classification**; the seven categories, in descending order, are: kingdom, phylum, class, order, family, genus, species.

taxonomy: in biology, the science concerned with classifying and naming organisms.

terminological knowledge base (TKB): a hybrid between a conventional term bank and a **knowledge base** as this is known in **artificial intelligence**; note, this is the definition of terminological knowledge base that is used within the framework of the COGNITERM Project, as described in Meyer, Skuce, Bowker and Eck 1992 a and b, and Meyer, Bowker and Eck 1992 a and b.

TKB: see **terminological knowledge base.**

unidimensional representation: a representation of a **concept system** which shows the **classification** of **concepts** according to only one possible **characteristic** at each level; Picht and Draskau (1985: 68-70) refer to unidimensional representations as *monodimensional* and *monohierarchic systems.*

Introduction

Multidimensionality is a phenomenon of classification that arises when objects can be classified in more than one way. For example, VEHICLE can be classified into LAND VEHICLE, AIR VEHICLE, and WATER VEHICLE (according to the **place of transportation**), or VEHICLE can be classified into MOTORIZED VEHICLE and NON-MOTORIZED VEHICLE (according to the **type of propulsion**). Essentially, classification involves grouping similar objects into a class on the basis of a common characteristic. If the objects in a given class can be distinguished on the basis of more than one characteristic, then they can be classified in more than one way. A *dimension* represents one way of classifying a group of objects; a classification with more than one dimension is said to be *multidimensional*.[1]

Classification is an integral part of terminology work. Before terminologists can begin creating term records, they need to have a clear understanding of the subject field in question. As often as possible, terminologists consult existing classifications; however, if no classifications can be found, they must establish their own. When trying to establish a classification of a subject field, terminologists must carefully analyze the conceptual structures of this field. As a phenomenon of classification, multidimensionality is also important to terminology work: a multidimensional understanding of a subject field is more substantial than a unidimensional understanding, and if all dimensions are not considered, certain terms may be omitted from the term bank or terminological publication. Despite its importance, however, multidimensionality is a subject that has received very little attention in the terminology literature.

---

[1] The term *multidimensionality* was adapted from Sager (1990: 147, 160), who refers to "complex and multidimensional conceptual systems." Other terminological authors who explicitly discuss such systems include Picht and Draskau (1985: 68–70), who refer to "polydimensional" and "polyhierarchic systems."

One of the reasons for this neglect may be that terminologists have not had adequate tools for dealing with this potentially complex phenomenon. They have generally represented concept systems using pencil-and-paper techniques; however, a two-dimensional medium is not conducive to creating complex multidimensional representations. It is only recently that developments in computational technology, particularly in the field of artificial intelligence (AI), have provided tools that make it feasible for terminologists to handle multidimensionality. As pointed out by Sager (1990: 147):

> Full conceptual systems based on the sets of relationships which are found to be the most informative and effective for any one subject field have hitherto been developed only on paper. Complex and multidimensional conceptual systems can only be represented and managed by computer; it was therefore impossible until quite recently to have anything other than rather prescriptive and narrow conceptual systems.

Workers in AI, and specifically in a subfield of AI known as *knowledge engineering*, encounter many of the same problems as terminologists in trying to acquire, formalize, and refine information about concepts. Knowledge engineering has already made significant progress in developing computer aids to help with these tasks, and terminology researchers may find it profitable to investigate the usefulness of existing knowledge engineering tools for terminology purposes. In fact, recent international conferences focusing on terminology and knowledge engineering are an indication that a symbiotic relationship can exist between these two fields.

One knowledge engineering tool, called *CODE (Conceptually Oriented Description Environment)*, has been under development at the Artificial Intelligence Laboratory of the University of Ottawa for over five years. The CODE system has several features, including a graphical component, that can help terminologists create multidimensional representations. CODE has been used for terminological applications for three years, most recently within the framework of the COGNITERM Project, funded by the Social Sciences and Humanities

Research Council (SSHRC) of Canada. One of the objectives of the COGNITERM Project is to build a prototype term bank, called *COGNITERM*. COGNITERM features a rich and formally structured knowledge component and can be described as a *terminological knowledge base (TKB)*, which is a hybrid between a conventional term bank and a knowledge base as this is known in AI.[2] We have been involved with the COGNITERM Project over the past eighteen months, and in the course of our work, we encountered and were intrigued by problems related to multidimensionality. Hence, we decided to investigate some aspects of this problem in the context of this thesis.

Objectives and Methodology

The objectives of this thesis are to *develop* and *apply* a set of guidelines for handling multidimensionality in a TKB.[3] During the course of our research, we found that handling multidimensionality can be divided into two fundamental operations: *recognition* and *representation*. Although the guidelines are specifically designed for use with the CODE system, we feel that they could also be used for more general applications. Since the CODE system is generic by design, it is quite possible that the guidelines could be adapted to work with other knowledge engineering systems. Also, since the initial guidelines are non-technical, they may prove useful for terminologists working in a traditional fashion (i.e. without computer assistance).

The thesis is divided into four main parts. Part I consists of a single chapter and is intended to provide the background information necessary for understanding

[2] More details about the COGNITERM Project and TKB can be found in Meyer, Bowker and Eck 1992 a and b; and Meyer, Skuce, Bowker and Eck, 1992 a and b.

[3] In this thesis, we refer to a TKB in the same sense that it is used in the COGNITERM Project.

multidimensionality. The general principles of classification are discussed, the concept of multidimensionality is explained, and some causes of multidimensionality are examined.

The goal of Part II is to establish an initial set of guidelines, independent of technology, for both recognizing and representing multidimensionality. These initial guidelines have two sources: an investigation of the classification techniques used in several disciplines (Chapter 2), and practical experimentation (Chapter 3). Chapter 2 begins with an examination of the terminology literature to determine what insights for handling multidimensionality are already present in the field. Next, we explore classification as it is practised in biology and library science, two disciplines in which classification is of great importance, to see if terminologists can learn from these techniques. An extended example dealing with the classification of the subject field of optical storage media is presented in Chapter 3. The example demonstrates the type of classification problems that terminologists are likely to encounter in their source documentation; possible solutions are also proposed. In Chapter 4, the final chapter of Part II, an initial set of guidelines for recognizing and representing multidimensionality is presented.

Part III of the thesis focuses on the development of a technical complement to the initial set of guidelines developed in Part II. The technical complement is derived from two sources: an analysis of the CODE system (Chapter 5), and practical experimentation (Chapter 6). Chapter 5 introduces the CODE system, detailing those features which are especially useful for helping terminologists to handle multidimensionality. Chapter 6 continues the extended example that was started in Chapter 3, showing how CODE can be used to aid terminologists in the recognition and representation of multidimensionality when classifying the subject field of optical storage media. Part III concludes with Chapter 7, which presents a technical complement to the initial guidelines, summarizing the ways in which the CODE system can facilitate the recognition and representation of multidimensionality.

Finally, in Part IV, the guidelines are subjected to a modest test by applying them to the construction of a TKB on a subfield of hypertext, namely hypertext links (Chapter 8). Each of the guidelines are followed carefully (where applicable), and problems are noted. In Chapter 9, issues raised during the application of the guidelines are discussed, suggestions for technical modifications to the CODE system that will allow it to better handle multidimensionality in future versions are given, and possible areas for future research are considered.

An overview of the structure of this thesis is given in the following table:

Part/Chapter	Purpose
Part I	· *provides background information necessary for understanding multidimensionality.*
Chapter 1	· discusses general principles of classification; explains the concept of multidimensionality; examines some of the causes of multidimensionality.
Part II	· *establishes initial guidelines, independent of technology, for handling multidimensionality.*
Chapter 2	· examines terminology literature to determine the status quo for handling multidimensionality in the field; explores classification as practised in biology and library science to see if terminologists can learn from these techniques.
Chapter 3	· presents an extended example demonstrating the types of problems terminologists are likely to encounter in their source documentation; proposes possible solutions.
Chapter 4	· presents an initial set of guidelines for recognizing and representing multidimensionality based on research carried out for Chapters 2 and 3.
Part III	· *presents a technical complement to initial guidelines developed in Part II.*
Chapter 5	· introduces the CODE system; details those features which are useful for helping terminologists to handle multidimensionality.
Chapter 6	· continues the extended example started in Chapter 3, showing how CODE can be used to help terminologists handle multidimensionality.
Chapter 7	· presents a technical complement to the initial guidelines, summarizing the ways in which CODE can facilitate the handling of multidimensionality.
Part IV	· *tests the guidelines.*
Chapter 8	· applies the guidelines to the construction of a TKB, noting problems that arise.
Chapter 9	· discusses issues raised during the application, suggests technical modifications for CODE, and considers possibilities for future research.

Particular Problems Encountered

In the course of writing this thesis, we encountered four notable categories of problems: knowledge acquisition problems, terminological problems, complexity of the phenomenon and the technology, and potential vastness of the subject matter.

Knowledge acquisition problems. When beginning this thesis, we soon discovered, as previously mentioned, that the issue of multidimensionality has received very little attention in the terminology literature. Consequently, we decided to explore classification in other disciplines to see if we could gain some insights into the problem. It is always difficult to investigate a field about which one knows virtually nothing; therefore, when exploring the fields of biology and library science, we found it extremely useful to enlist the help of subject-field experts. When surveying the biology literature (section 2.2), we consulted Dr. François Chapleau, a biology professor at the University of Ottawa with a special interest in taxonomy, and Mr. Jean-Pierre Artigau, a former terminologist with the Department of the Secretary of State of Canada, who had specialized in the terminology of natural sciences. Similarly, when investigating the field of library science (section 2.3), we conferred with Mrs. Cécile Prud'Homme, a reference librarian at the University of Ottawa, and Professor Nancy Williamson of the Faculty of Library and Information Science of the University of Toronto, who is a consulting editor of the journal *International Classification*. The assistance provided by these subject-field experts was invaluable: they helped to direct our documentation search, clarified key concepts, and verified information. We also consulted subject-field experts as we conducted thematic research in the fields of optical storage media (Chapters 3 and 6) and hypertext (Chapter 8). Our optical storage expert was Mr. Don Slaunwhite, a software developer in the CD-ROM division of Corel Systems Corp., and our hypertext expert was Mr. Timothy Lethbridge, a Ph.D. student in Computer Science at the University of Ottawa.

Terminological problems. A second problem that we faced was related to the terminology encountered in the documentation. Not only was there new terminology to learn and absorb in the new fields (i.e. biology and library science), but the terminology used in the classification and terminology literature contained a considerable number of synonyms and inconsistencies. We therefore found it beneficial to create a glossary, containing definitions, of the terms we chose to use in this thesis. Whenever possible, for terms employed in the field of terminology, we adhered to the terms and definitions proposed by the International Organization for Standardization (ISO) in their Terminology Vocabulary (ISO 1087). In addition to creating a glossary, we italicized and defined new or important terms the first time they appeared in the text.

Complexity of the phenomenon and the technology. The third significant problem that we encountered was that both the phenomenon of multidimensionality, and the sophisticated AI tool used to handle it, are complex. Hence, we found it difficult both to understand and explain the phenomenon and the tool. With regard to multidimensionality, we attempted to illustrate our explanations with plenty of graphical representations and examples, starting with simple ones, and working up to more advanced ones. When describing the CODE system, we drew on our experience with the COGNITERM Project, and we included many screen dumps[4] to help the reader visualize what was happening on the screen at various stages of the project.

Potential vastness of the subject matter. Finally, we realized that the subject matter with which we were dealing was quite vast, and we thought it wise to restrict the scope of the thesis. We constrained our work in several ways. Firstly, although we realize that

[4] A screen dump is a printout of whatever is on the computer screen at any given time.

classification plays an important role in many disciplines, such as philosophy and cognitive psychology, we decided to focus on the classification techniques used in biology and library science. We chose biology because living organisms were among the first objects that people ever tried to classify; therefore, the techniques used in biological classification have been meticulously developed over a long period of time. Library science was our second choice because the primary concern of this field is classifying knowledge as it is found in documents, a concern that is shared by terminologists.

Secondly, we discovered that multidimensionality could occur not only in generic-specific relations between concepts, but in part-whole relations as well. For example, as pointed out by Cruse (1986: 169), HUMAN BODY can be classified into *segmental parts* (e.g. TRUNK, HEAD, LIMBS), or it can be classified into *systemic parts* (e.g. SKELETON, MUSCLES, NERVES, BLOOD VESSELS). For the purpose of this thesis, we decided to consider multidimensionality only as it occurs in generic-specific relations, and to exclude the consideration of part-whole or any other types of relations.

Finally, we decided to restrict our investigation of multidimensionality to the *recognition* and *representation* of the phenomenon. Multidimensionality may have implications in other areas of terminology, for example, in definition construction. These implications would make interesting topics for further research, but they are beyond the scope of the present thesis.

PART I

CHAPTER ONE

AN INTRODUCTION TO CLASSIFICATION AND MULTIDIMENSIONALITY

1.0 WHAT IS MULTIDIMENSIONALITY?

Multidimensionality is a phenomenon that occurs when one tries to classify *objects*.[5] In order to understand multidimensionality, it is necessary to first understand the general principles of classification.

1.1 CLASSIFICATION

It is generally held that classification is the act of grouping like objects together into classes on the basis of characteristics that are common to each class member. This automatically implies a separation of the unlike objects. In other words, all members of a class share at least one characteristic that members of other classes do not possess.

For example, consider the following list:

- car
- refrigerator
- boat
- dishwasher
- airplane
- oven

Three of the objects in this list have the common characteristic of being vehicles, so we can place them in one class which we will call VEHICLE. The remaining three have the common characteristic of being appliances, so we will place them in another class called APPLIANCE. By making the distinction between VEHICLE and APPLIANCE, we have

[5] In the Terminology Vocabulary produced by the International Organization for Standardization (ISO 1087: 1), *object* is defined as "Any part of the perceivable or conceivable world. NOTE — Objects may also be material (e.g. engine) or immaterial (e.g. magnetism)."

classified the objects in the list. That is, we have united the like objects and separated the unlike objects on the basis of common characteristics. Figure 1-1 illustrates this classification.

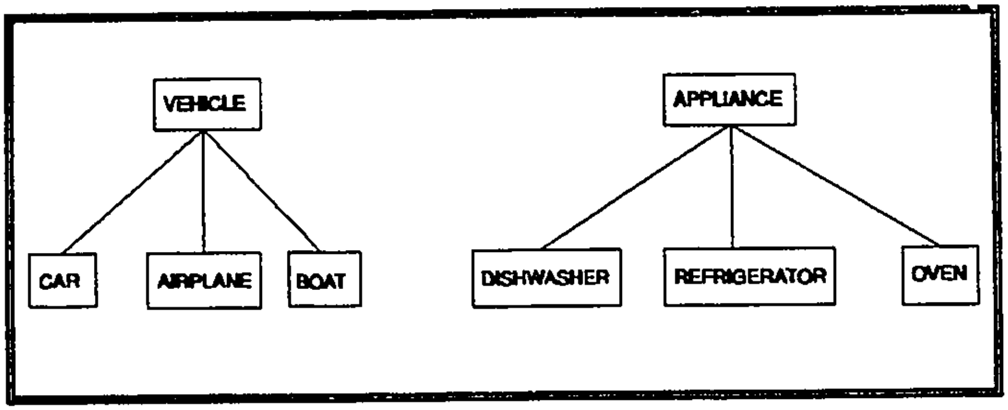


Figure 1-1. A classification showing the grouping of like objects.

Objects belonging to a class are said to be members of that class; they are related because they share at least one characteristic that is not shared by members of other classes. Any class may be a subclass of a larger class (e.g. in figure 1-1, AIRPLANE is a subclass of VEHICLE), and a subclass may itself be divided into subclasses. This process of subclassification can continue until all the members of a class share exactly the same characteristics, and hence cannot be classified any further. For example, figure 1-2 shows how the class VEHICLE has been classified into three subclasses: LAND VEHICLE, AIR VEHICLE and WATER VEHICLE. The class AIR VEHICLE has then been classified into the subclasses NON-MOTORIZED AIR VEHICLE and MOTORIZED AIR VEHICLE, and MOTORIZED AIR VEHICLE has been classified into HELICOPTER and AIRPLANE.

1.1.1 Characteristics

Characteristics are crucial to the process of classification; objects are grouped into classes on the basis of shared characteristics. A characteristic has two components: a name

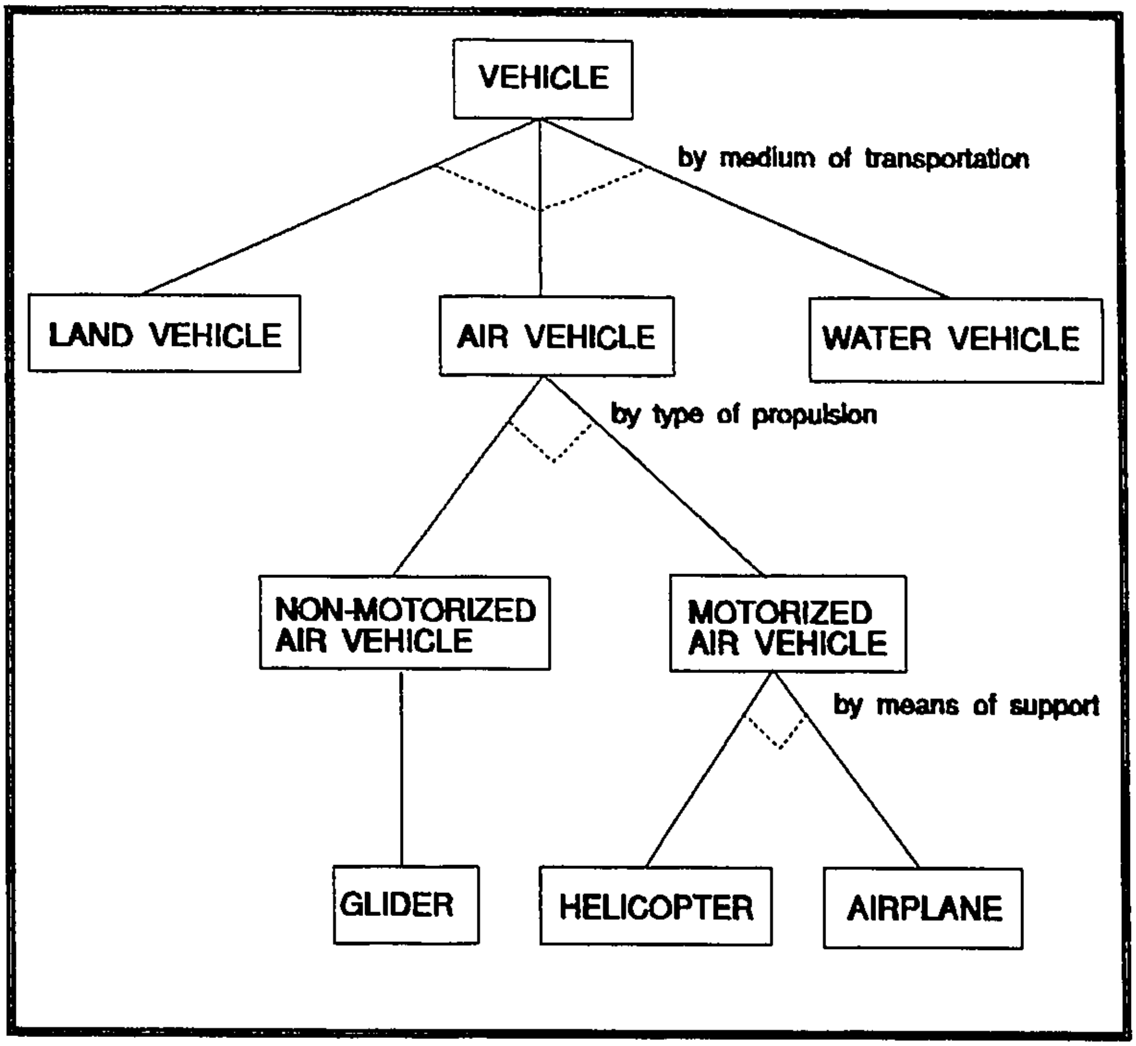


Figure 1-2. Classification and subclassification.

and a value.[6] For example, if we look back to figure 1-2, we can see that objects in the class VEHICLE were further classified into one of three subclasses: AIR VEHICLE, WATER VEHICLE or LAND VEHICLE. This classification was made on the basis of the characteristic **medium of transportation**. The classes AIR VEHICLE, WATER VEHICLE and LAND VEHICLE can all be described as having a characteristic with the name **medium of transportation**, but the value of this characteristic is different for each of them. In this

[6] Many people use the term *characteristic* to mean the combination of the name and the value; however, for our purpose, it is necessary to distinguish between the two parts.

example, the value of this characteristic can be said to be ***air*** for AIR VEHICLE, ***water*** for WATER VEHICLE, and ***land*** for LAND VEHICLE. A characteristic that is used to distinguish the subclasses of one class is called a *differentiating characteristic*. Because AIR VEHICLE, WATER VEHICLE and LAND VEHICLE are specializations of VEHICLE that have been separated from one another on the basis of the characteristic **medium of transportation**, we can identify **medium of transportation** as the differentiating characteristic at this level of the classification.

1.1.2 Relations

Relations between objects are based on characteristics. The type of relation that we have seen in the examples so far is a *generic relation*. There can be other types of relations such as partitive relations (e.g. part–whole) or sequential relations (e.g. cause–effect), but they are beyond the scope of this thesis.

The ISO terminology vocabulary defines a generic relation as "a hierarchical relation which is based on the partial identity of the intensions of generic, specific and co–ordinate concepts" (ISO 1087: 3). A hierarchical relation is one that is "established by division of a superordinate concept into subordinate concepts forming one or more levels" (ISO 1087: 3). In other words, the original or *generic* concept is divided and subdivided into a number of more *specific* concepts. A specific concept has all the characteristics of its generic concept, plus an additional or modified characteristic. According to ISO, a *co–ordinate* concept is "a concept in a hierarchical system which ranks at the same level as one or more other concepts" (ISO 1087: 2). In other words, two specific concepts that come from the same generic concept (i.e. two subclasses of the same class) are *co–ordinate* concepts. They both have all the characteristics of the generic concept plus at least one additional or modified characteristic called the differentiating characteristic. Although both co–ordinate concepts possess this differentiating characteristic, they each have a different value for it, and thus can be distinguished from one another (cf. section 1.1.1).

Generic relations are sometimes referred to as *isa* or *type* relations because the specific concept *is a type* of the generic concept. For example, a GLIDER *is a type* of NON-MOTORIZED AIR VEHICLE, and a NON-MOTORIZED AIR VEHICLE *is a type* of AIR VEHICLE, etc. It should also be noted that the identification of a concept as being *generic* or *specific* is relative. In other words, a specific concept at one level of the classification can be the generic concept at another. For example, as shown in figure 1-3, AIR VEHICLE is one of the specific concepts of VEHICLE, but it is the generic concept of NON-MOTORIZED AIR VEHICLE and MOTORIZED AIR VEHICLE.

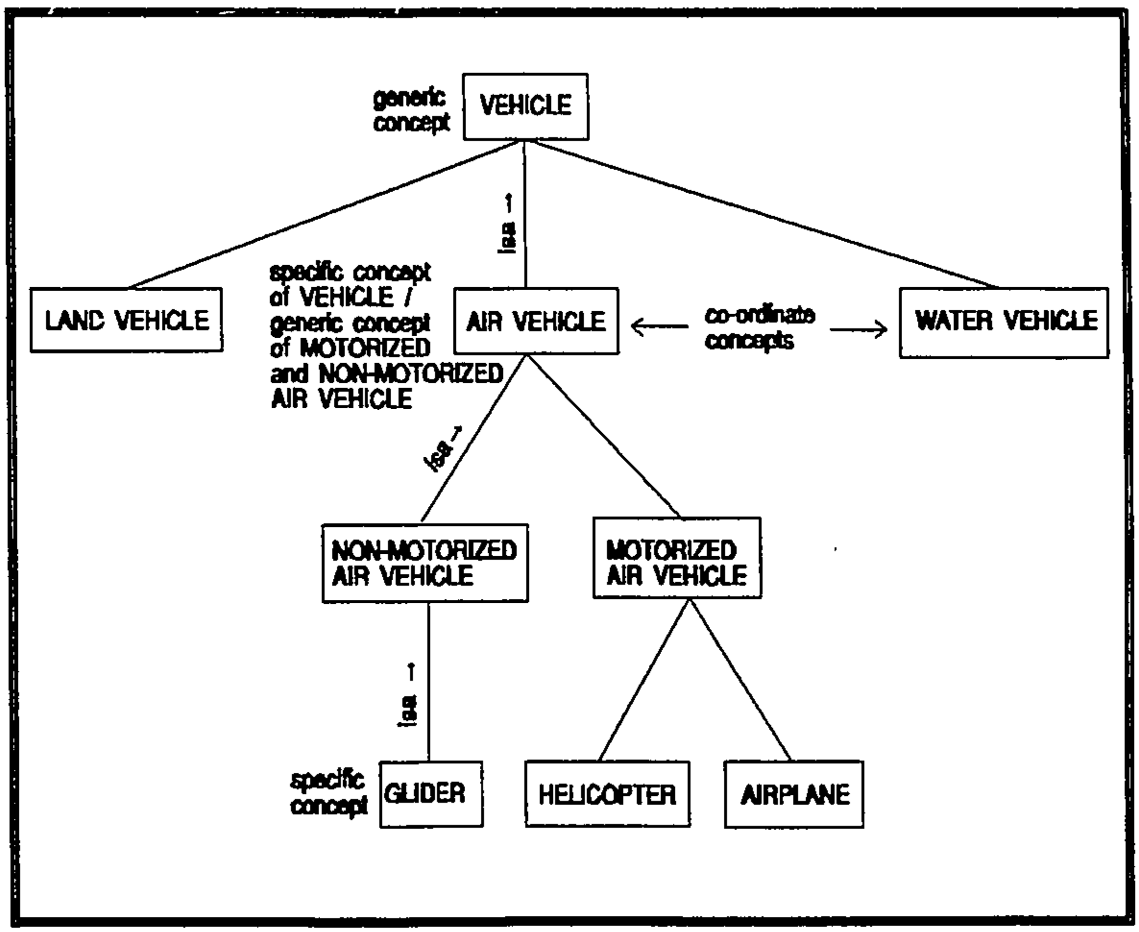


Figure 1-3. Classification showing the relationships between objects.

In summary, objects are grouped together in a class because they have at least one characteristic with the same value. One can continue to subclassify classes on the basis of different differentiating characteristics until all members of the class share exactly the same characteristics and have no other characteristics. A differentiating characteristic is one that is used to classify a generic concept into specific concepts, each of which have a different value for this characteristic. All the specific concepts at a given level of the classification are said to be co-ordinates.

1.2 MULTIDIMENSIONALITY

We have established that classification involves uniting like objects on the basis of a common characteristic, while separating unlike objects. Therefore, it logically follows that what is like or unlike depends on the characteristic that is chosen. If more than one characteristic can be used to distinguish between the objects (i.e., if there is more than one differentiating characteristic), then the objects can be classified in more than one way.

Throughout this thesis, we will use the term *dimension* to designate a classification that has been produced by classifying objects on the basis of one particular differentiating characteristic. In other words, a dimension represents one way of classifying a group of objects. In a case where there are two differentiating characteristics to choose from, the objects can be classified in two ways; therefore, there are two dimensions. A class which has been given only one dimension is said to be *unidimensional*, while a class which has more than one dimension is *multidimensional*.

The representation of the class VEHICLE that was shown in figure 1-2 was unidimensional; VEHICLE was classified solely on the basis of the characteristic **medium of transportation**. However, there are actually several ways in which VEHICLE can be classified, based on the different characteristics that a vehicle can have. The following list contains some of the characteristics that could be used to classify VEHICLE:

1) **medium of transportation**
2) **type of propulsion**
3) **type of load**[7]

Depending on which characteristic is chosen, the classification will be different. Figure 1-4 shows three different dimensions or ways of classifying VEHICLE based on three different characteristics that VEHICLE can have.

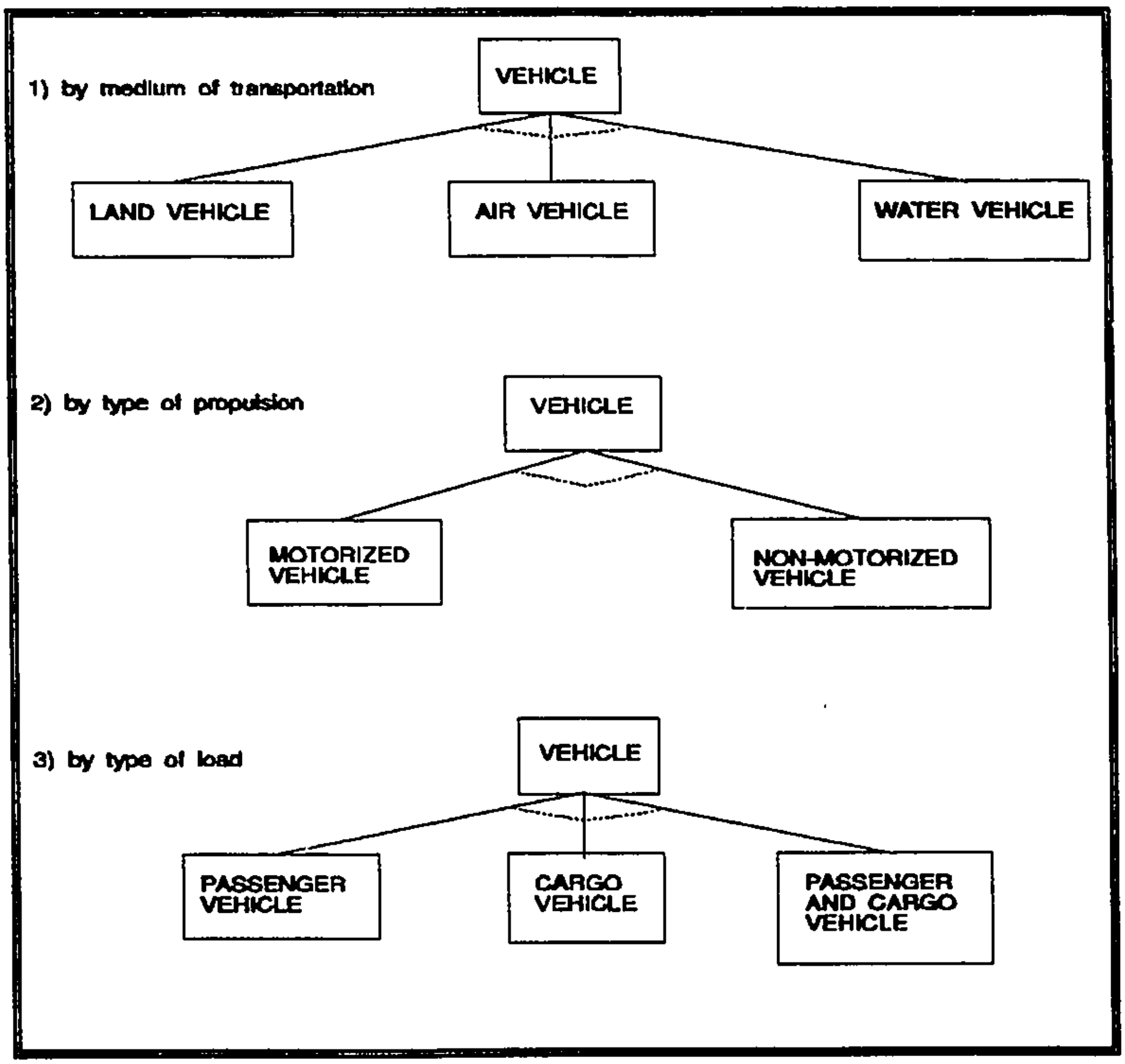


Figure 1-4. Three unidimensional classifications of VEHICLE.

[7] By **type of load**, we mean the type of load that the vehicle is *primarily* intended to carry (e.g. passengers, cargo).

However, if all three ways of classifying VEHICLE are considered simultaneously, then the representation is multidimensional. Figure 1-5 shows a fairly simple example of a multidimensional representation for VEHICLE. As the figure indicates, some objects, such as AIRPLANE or CAR, can be members of several dimensions.

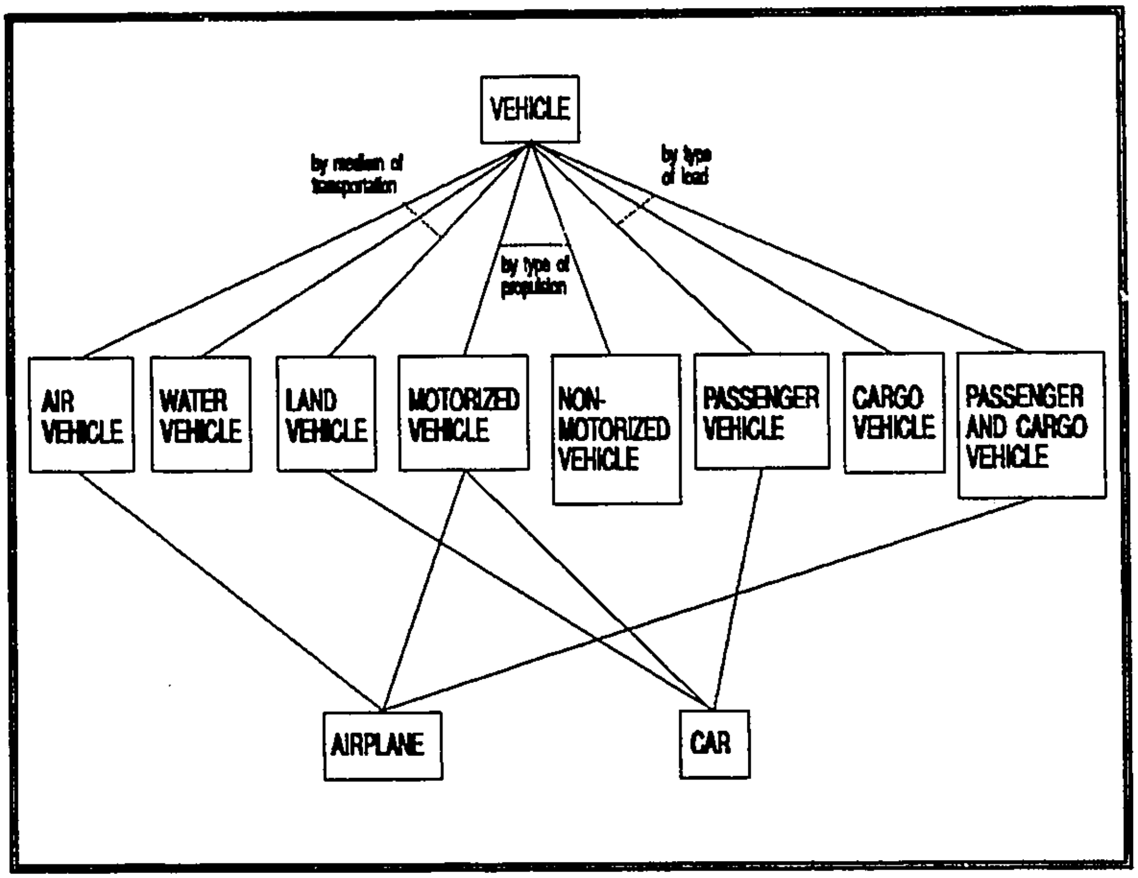


Figure 1-5. A multidimensional classification of VEHICLE.

Unfortunately, multidimensionality is not always simple and neat; sometimes it can be very complicated and messy. Figure 1-6 shows a more complex multidimensional representation of the class VEHICLE. As we can see, multidimensionality can occur at various levels of the hierarchical classification; the higher the number of occurrences of multidimensionality within a subject field, the messier and more complex the graphical representation of that subject field.

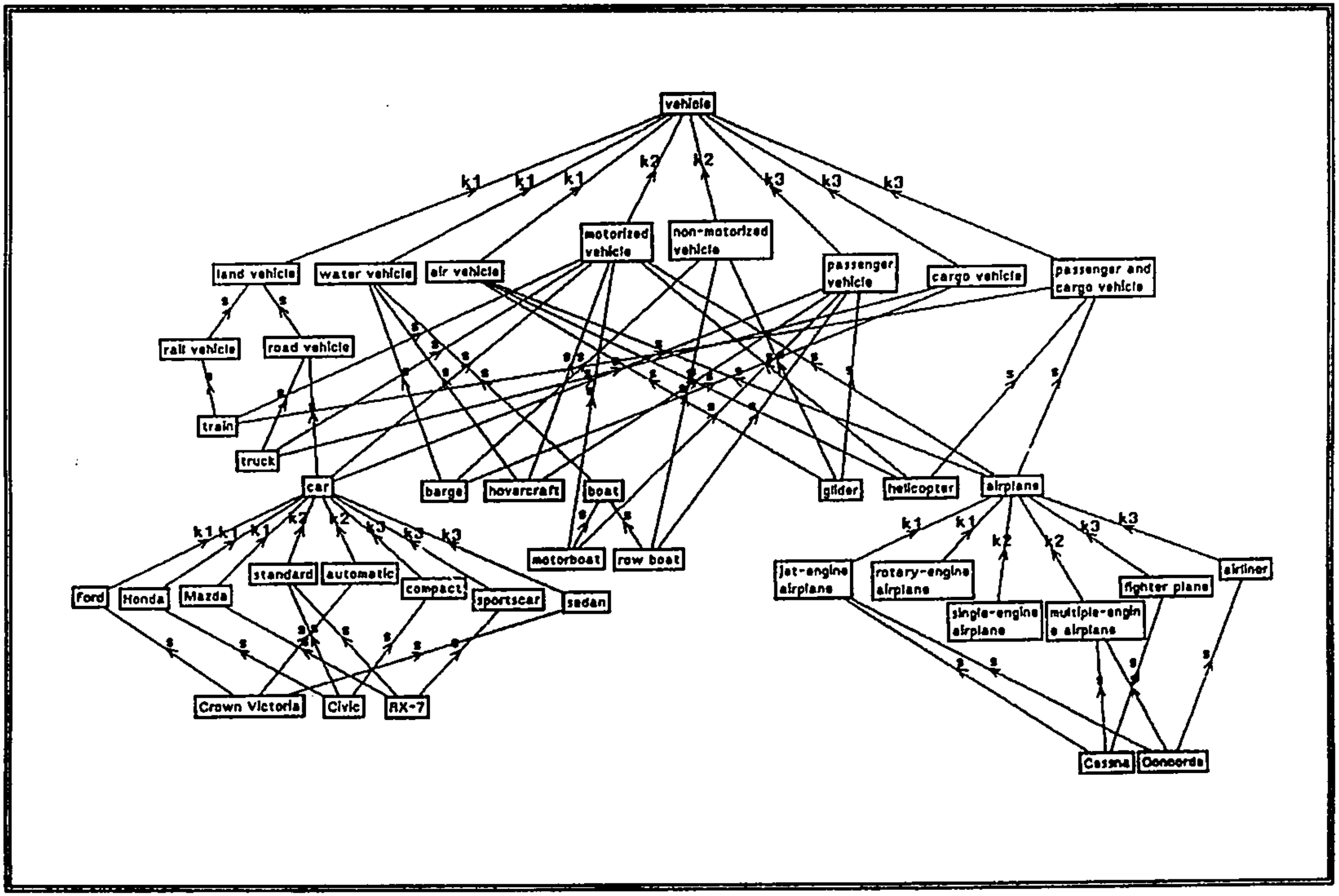


Figure 1-6. A complex multidimensional representation of VEHICLE, drawn using CODE.

Figure 1–6 was drawn using CODE (Conceptually Oriented Description Environment), a knowledge engineering tool designed to help users *manage* knowledge.[8] CODE assigns a different notation to each dimension (k1, k2, k3, etc.); the differentiating characteristic that matches each notation can be found in another part of the system. At the first level, VEHICLE has been classified according to three different characteristics: k1 (i.e. dimension 1) represents a classification of VEHICLE on the basis of the characteristic **medium of transportation**, k2 represents a classification according to the characteristic **type of propulsion**, and k3 represents a classification based on the characteristic **type of load**.

At a deeper level of the classification, the class CAR also has three dimensions. In this case, k1 represents a classification on the basis of the **make** of the car, k2 represents a classification on the basis of **type of transmission**, and k3 represents a classification according to the **style** of the car. A third multidimensional representation is shown under the class AIRPLANE. Obviously, the more occurrences of multidimensionality that appear in a representation, the more complex the representation becomes. This is particularly true when the multidimensionality occurs at different hierarchical levels.

In summary, we can say that multidimensionality is the phenomenon that occurs when an object can be classified in more than one way, on the basis of different differentiating characteristics. Each way of classifying the object is considered to be a different dimension. When more than one dimension is reflected in a single classification, the representation is said to be multidimensional. Multidimensionality can occur at any level of a classification, and at more than one level of the same classification; the complexity of the representation of a classification increases with the number of occurrences of multidimensionality that arise within the classification.

---

[8] CODE will be described in considerable detail in chapters 5 and 6.

1.2.1. What causes multidimensionality?

Multidimensionality occurs when a object can be classified in more than one way. Why is it that objects can be classified in more than one way? What causes one person to classify a certain object in one way, while another person classifies the same object in a different way? What causes the same person to classify the same object in several different ways? We will briefly consider six possible causes of multidimensionality: viewpoint, opinion, scientific theory, purpose, changing organization of the reality and of the subject field describing it, and language and culture.

Viewpoint

People often see objects from a particular point of view, which causes them to attach a higher degree of importance to one characteristic over another. For example, a person who uses optical discs as a data storage medium would most probably consider OPTICAL DISC from the viewpoint of a user. Users are primarily interested in whether they can write information to the disc, or merely retrieve information that has been stored on the disc by the manufacturer. Thus, a user would classify OPTICAL DISC according to the characteristic **degree of writability**, as shown in figure 1–7.

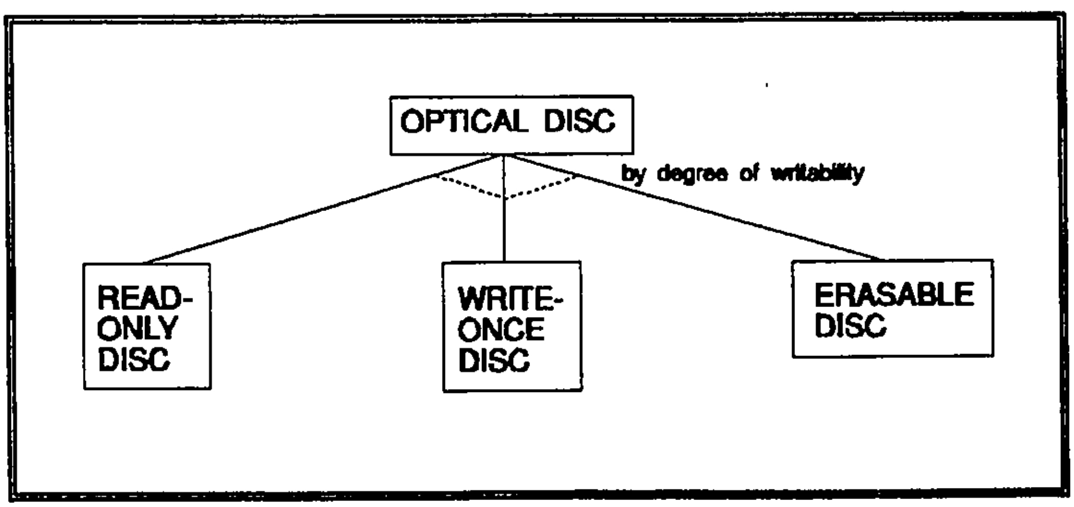


Figure 1–7. Classification of OPTICAL DISC from a user's viewpoint. *Source: Chen 1989b: 5.*

Meanwhile, manufacturers may very well look at optical discs from a different viewpoint: whether they are required to produce a pre-recorded disc (REPLICABLE DISC) or a blank disc (INDIVIDUAL DISC). Thus, manufacturers might classify OPTICAL DISC on the basis of the characteristic **replicability**, as shown in figure 1-8.

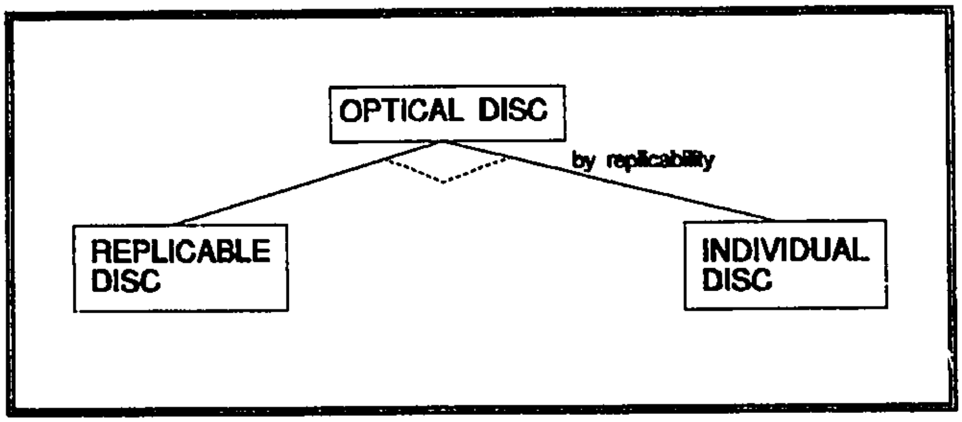


Figure 1-8. Classification of OPTICAL DISC from a manufacturer's viewpoint. *Source: Heimburger 1990: 32.*

Purpose

Another reason that people classify objects differently is because they are classifying the objects for a different purpose, and one particular classification might serve a specific need better than another classification. For example, as shown in table 1.1, *Le Bescherelle 1: L'Art de conjuguer* (*Le Bescherelle* 1980) classifies French verbs on the basis of their endings into -ER VERBS, -IR VERBS, and -RE VERBS, and further subclassifies the verbs within these groups. The ending of a verb determines how it will be conjugated. Therefore, in this book, verbs that are conjugated in the same manner have been grouped together, and only one representative verb per category has been fully conjugated. All other verbs are listed alphabetically at the back of the book; to find out how one of these verbs is conjugated, users are referred to the page where the representative verb for that category has been conjugated. For example, a complete conjugation is given for the verb *cuire* (a member of the category of verbs ending in –uire); users who wish to know how to translate the verbs *traduire*, *détruire*, or *instruire* would be referred to the conjugation given for *cuire*. One purpose of this classification is to help teach people rules for conjugating different categories of French verbs; if one knows the rules, one is able to conjugate almost any verb[9] that belongs to one of the categories.

FRENCH VERBS

–er verbs	–ir verbs	–re verbs
–er	–ir	–endre
–cer	–enir	–erdre
–ger	–tir	–attre
–écer	–llir	–aindre
–ier	–mir	–aire
–ayer	–rir	–aître
–oyer	–vir	–oire
–eter	–uir	–oudre
–éger	–oir	–ivre
–e(.)er	–seoir	–uire
etc.	etc.	etc.

Table 1.1. FRENCH VERBS classified on the basis of verb endings. One verb for each of the categories is fully conjugated in the book; the other verbs are listed alphabetically at the back of the book, and the user is referred to the conjugation of the representative verb for that category. *Source: Le Bescherelle 1980.*

Meanwhile, a second book of French verb conjugations, *French Verbs* (Kendris 1990), is a pocket book for travellers that contains alphabetical entries with complete conjugations for the 300 most frequently used verbs in the French language. This book is primarily designed to help travellers communicate, not to teach them the rules of conjugation. Hence, the entire conjugation of the 300 most common verbs in the French language are presented in alphabetical order. As illustrated in table 1.2, the complete conjugation is provided for all the verbs in the book, even if they belong to the same verb ending category.

[9] In fact, *Le Bescherelle* claims to conjugate 12,000 verbs.

aimer (to like, to love)	*Impératif:* aime, aimons, aimez *Participe présent:* aimant *Participe passé:* aimé
INDICATIF	SUBJONCTIF
Présent j' aime tu aimes il aime nous aimons vous aimez ils aiment	*Présent* j' aime tu aimes il aime nous aimions vous aimiez ils aiment
Imparfait j' aimais tu aimais il aimait nous aimions vous aimiez ils aimaient	*Passé* j' aie aimé tu aies aimé il ait aimé nous ayons aimé vous ayez aimé ils aient aimé
Passé composé j' ai aimé tu as aimé il a aimé nous avons aimé vous avez aimé ils ont aimé	*Futur simple* j' aimerai tu aimeras il aimera nous aimerons vous aimerez ils aimeront
Plus-que parfait j' avais aimé tu avais aimé il avait aimé nous avions aimé vous aviez aimé ils avaient aimé	*Conditionnel* j' aimerais tu aimerais il aimerait nous aimerions vous aimeriez ils aimeraient

ajouter (to add)	*Impératif:* ajoute, ajoutons, ajoutez *Participe présent:* ajoutant *Participe passé:* ajouté
INDICATIF	SUBJONCTIF
Présent j' ajoute tu ajoutes il ajoute nous ajoutons vous ajoutez ils ajoutent	*Présent* j' ajoute tu ajoutes il ajoute nous ajoutions vous ajoutiez ils ajoutent
Imparfait j' ajoutais tu ajoutais il ajoutait nous ajoutions vous ajoutiez ils ajoutaient	*Passé* j' aie ajouté tu aies ajouté il ait ajouté nous ayons ajouté vous ayez ajouté ils aient ajouté
Passé composé j' ai ajouté tu as ajouté il a ajouté nous avons ajouté vous avez ajouté ils ont ajouté	*Futur simple* j' ajouterai tu ajouteras il ajoutera nous ajouterons vous ajouterez ils ajouteront
Plus-que parfait j' avais ajouté tu avais ajouté il avait ajouté nous avions ajouté vous aviez ajouté ils avaient ajouté	*Conditionnel* j' ajouterais tu ajouterais il ajouterait nous ajouterions vous ajouteriez ils ajouteraient

Table 1.2. FRENCH VERBS classified alphabetically. The complete conjugation is given for each of the 300 verbs in the book, even if more than one verb is conjugated in exactly the same way, as is the case with the verbs *aimer* and *ajouter*. *Source: Kendris 1990.*

Opinion

Different classifications can also result from different opinions.[10] For example, in the cookbook *Basics & Beyond: An Adventure in Cooking* (MacDonald and O'Connell 1986), the authors have classified recipes into some of the following categories:

[10] In some cases, an opinion results from a viewpoint or purpose; however, we feel that an opinion does not necessarily have to be based on a viewpoint or purpose, but can be based, for example, on personal feelings or intuition. Therefore, we have included *Opinion* as a distinct cause of multidimensionality.

Appetizers and Beverages
Soups and Sauces
Entrées

Meanwhile, the editors of another cookbook, *Good Housekeeping's Cookery Book* (1967), have classified their recipes into some slightly different categories:

Beverages
Appetizers and Soups
Meat, Poultry and Fish
Pasta and Rice
Sauces

If we compare the two lists, we can see that the *Good Housekeeping* editors have paired SOUPS with SAUCES, probably because they are both prepared in a similar fashion, while the authors of *Basics and Beyond* have grouped APPETIZERS and SOUPS together because they are generally served before the main meal. Another difference is that the *Basics and Beyond* authors have grouped all the ENTRÉES together in one category, while the *Good Housekeeping* editors have classified entrées into meat dishes (i.e. MEAT, POULTRY and FISH) and non-meat dishes (i.e. PASTA and RICE). These classifications were slightly different because the personal opinions of the authors and editors were slightly different.

Scientific theory

Many differences in classification result from different scientific theories. For example, in the field of biology, there are three different and mutually exclusive methods of classifying living organisms.[11] Each of these methods produces a different classification.

Figure 1–9 depicts a classification of animals produced following the method of evolutionary systematics. This method classifies organisms on the basis of shared evolutionary history and *morphological* (i.e. structural) similarity. In this example, CROCODILES and BIRDS are not grouped together because they have significant morphological differences.

[11] The three methods of biological classification will be discussed in greater detail in section 2.2.1.

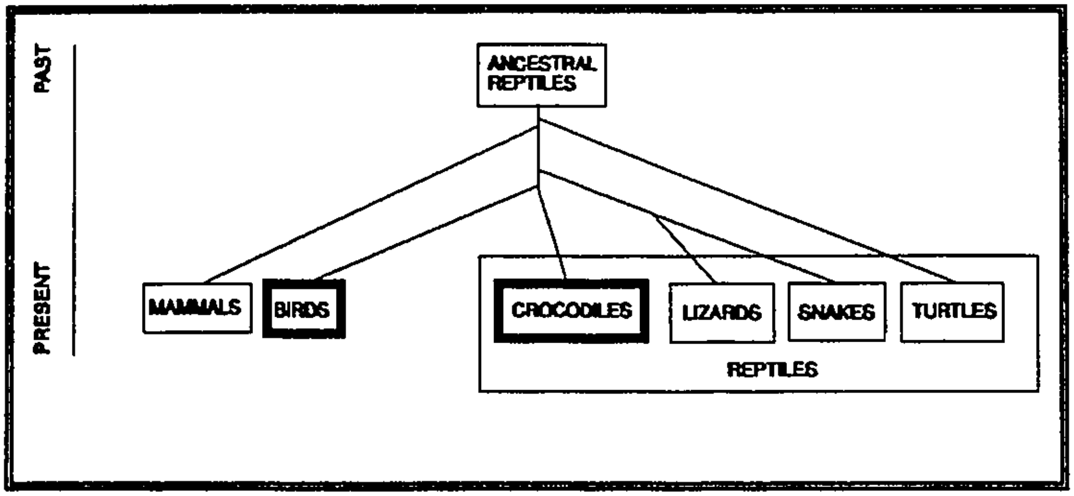


Figure 1-9. Classification of animals produced according to the scientific method of evolutionary systematics. *Source: Dorit 1991: 546.*

A classification of the same animals produced according to the method of cladistics is shown in figure 1-10. Using the method of cladistics, organisms are classified strictly on the basis of their evolutionary history, without taking morphological similarity into account. In this example, CROCODILES and BIRDS *have* been grouped together because they evolved from a common ancestor.

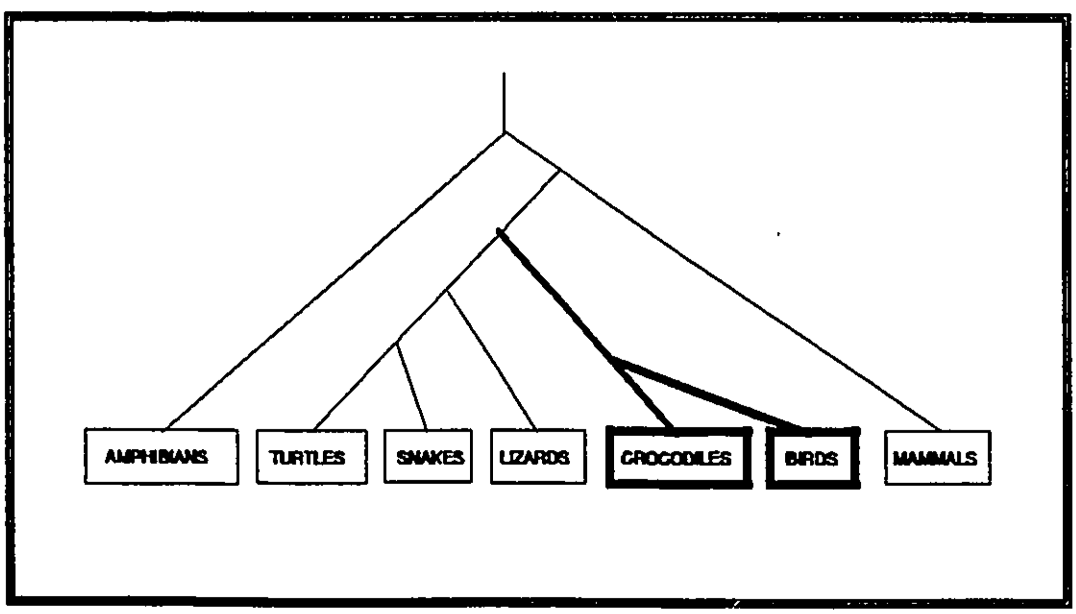


Figure 1-10. Classification of animals produced according to the scientific method of cladistics. *Source: Dorit 1991: 546.*

Changes in the reality of a phenomenon and in the subject field describing it

Different classifications can also result from changes in the reality of a phenomenon in the world (e.g. a given technology), and the consequent changes in the subject field describing this phenomenon. For example, as shown in figure 1-11, before the advent of a process called PHOTOTYPESETTING, the two main categories of typesetting were HOT TYPESETTING and COLD TYPESETTING. HOT TYPESETTING referred to any typesetting process that used type characters cast from molten metal. All other typesetting processes, such as STRIKE-ON TYPESETTING, became known as COLD TYPESETTING.

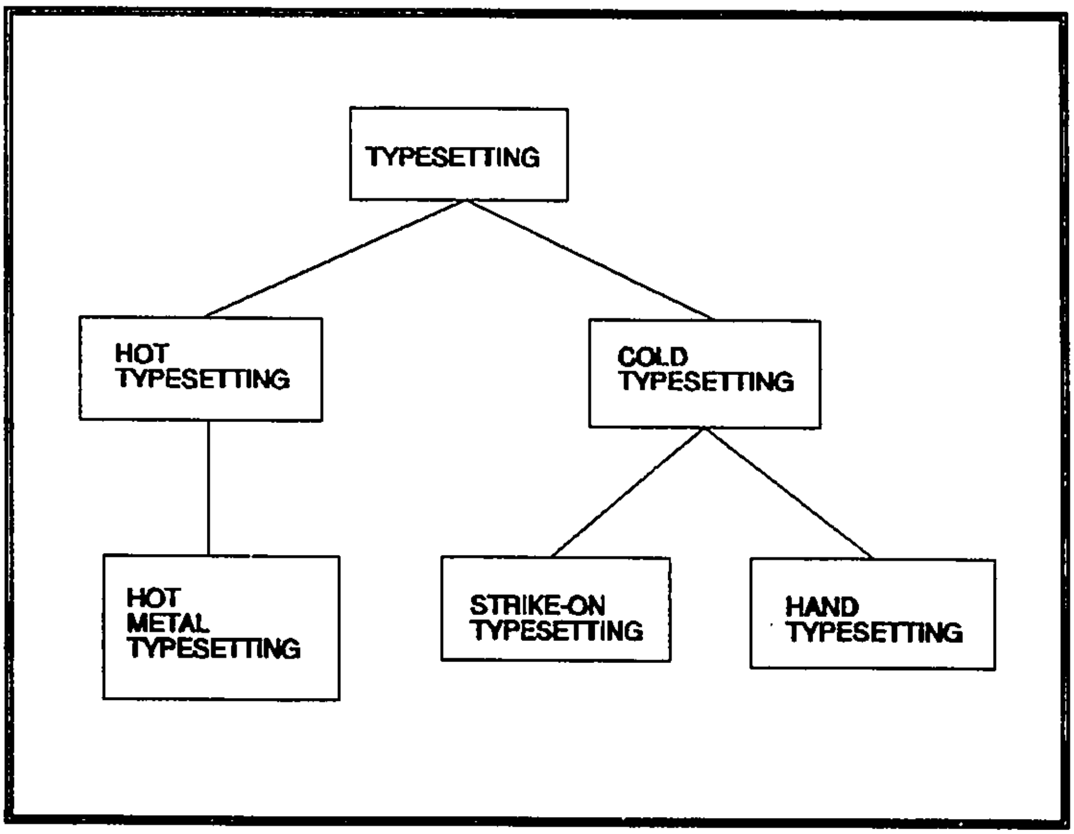


Figure 1-11. A former classification of TYPESETTING.

When PHOTOTYPESETTING and computerized typesetting techniques were developed, they were first classified as kinds of COLD TYPESETTING because no molten metal was involved. As we can see in figure 1-12, these techniques have now replaced HOT TYPESETTING altogether. Since all typesetting methods are now so-called COLD TYPESETTING, the distinction between HOT TYPESETTING and COLD TYPESETTING is no longer relevant; nevertheless, the term *cold typesetting* is still occasionally used to describe modern typesetting methods.

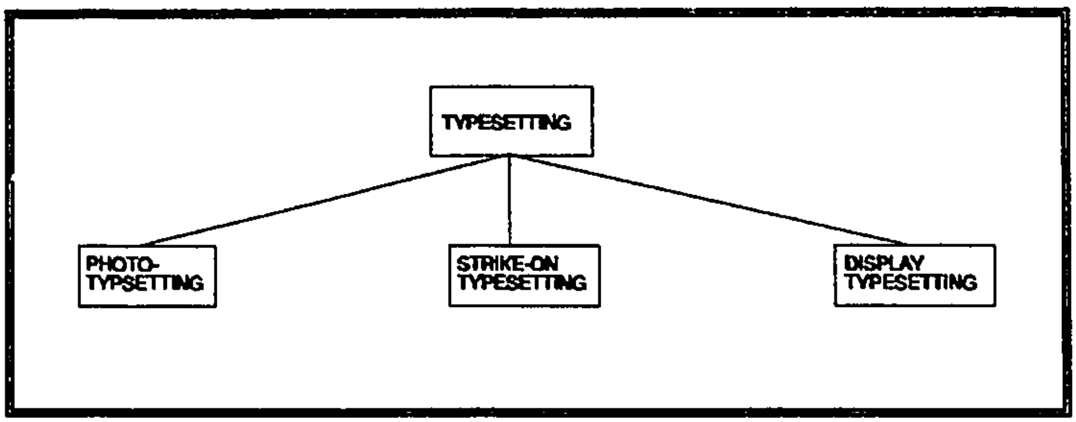


Figure 1-12. A more common classification of TYPESETTING.

Language and culture

Finally, different classifications can also be observed in different languages and cultures. With regard to language, it is well-known that knowledge structures do not have simple one-to-one correspondences from one language to the next. In the words of Mounin (1963: 73), "Tout système linguistique renferme une analyse du monde extérieur qui lui est propre, et qui diffère de celle d'autres langues..." For example, in English, flowing bodies of water are classified primarily according to the characteristic **size**, producing words such as *river*, *stream*, *creek*, etc. In French, however, flowing bodies of water are classified not only according to the characteristic **size**, but also according to the characteristic **destination**. Therefore, in French, there are two words, *rivière* (a river that empties into another river or a lake) and *fleuve* (a river that empties into the sea), covering the same semantic extension that is represented in English by the word *river*.

Different classifications can occur on a cultural level as well. For instance, while North Americans classify CATTLE as a type of livestock to be used for food, people from other cultures might classify CATTLE primarily as a type of money, or perhaps as a means of transportation.[12] The cultural question is also affected by time: not only do concepts differ from culture to culture, but they may differ from period to period within the same culture. This temporal issue was discussed in the previous section (cf. section 1.2.1 (*Changes in the reality of a phenomenon and in the subject field describing it*)).

The underlying causes of interlinguistic and intercultural differences have been the subject of much debate, and are beyond the scope of the present thesis.

In summary, we have seen that multidimensionality has a variety of causes, including viewpoint, purpose, opinion, scientific theory, changes in the reality of a phenomenon and in the subject field describing it, and language and culture.

[12] This example was inspired by Brian Harris in a personal communication.

PART II

CHAPTER TWO

AN OVERVIEW OF CLASSIFICATION AND MULTIDIMENSIONALITY IN VARIOUS DISCIPLINES

2.0 CLASSIFICATION FOR GENERAL AND SPECIFIC PURPOSES

Now that we are familiar with the basic concepts underlying classification, we will discuss what types of people are concerned with classification and why. On a very general level, everybody is concerned with classification. In the words of Buchanan (1979: 10), we use classification to "cope with the multitude of unorganized impressions we receive by way of our senses; we can use the pattern to impose order on chaos, 'placing' what we see, hear, feel, smell and taste, within it." Thus, classification is an important part of our everyday lives and we see examples of it all around us. For instance, a grocery store is organized so that all products of the same class are together: meat products, dairy products, fresh produce, etc. Each class is then subclassified: fresh produce is classified into vegetables and fruits; fruits are classified into bananas, oranges, and apples; apples are classified into Macintosh apples and Delicious apples; Delicious apples are classified into red Delicious apples and golden Delicious apples; etc. Without classification, customers would find it very difficult to locate the product they need in a store.

In addition to being important in our everyday lives, classification plays a significant role in some very specialized disciplines. One of the objectives of this thesis is to develop a set of guidelines for handling multidimensionality in a terminological knowledge base. To do this, we will begin by discussing the role of classification in terminological work. We will then examine the classification methods used in the disciplines of biology and library science, to see if any of these techniques might be incorporated into our set of guidelines.

While biology itself may not, at first glance, seem particularly pertinent to the subject of this thesis, quite the opposite is true: living organisms were among the *first* objects that people ever tried to classify; hence, biology was one of the first fields to be concerned with classification. The methods used in biological classification have been meticulously developed over a great number of years.

Once people had begun classifying objects in the natural world, they found it necessary to begin classifying their knowledge about these objects. The field of library science is primarily concerned with classifying knowledge as it is found in documents, a concern that is shared by terminologists.

2.1 CLASSIFICATION AND MULTIDIMENSIONALITY IN TERMINOLOGY

Picht and Draskau (1985: 30) observe that "There is a close relationship between terminology and the science of classification, which plays a central role in conceptual relations and systems of concepts." Concepts themselves form the cornerstone of terminology work. Picht and Draskau (1985: 36) go on to comment that "In the theory of terminology, there is widespread agreement that the concept occupies a central position...." Sager (1990: 22) gives one possible explanation of the process of concept formation. According to Sager, concept formation is a process of grouping and ordering all the objects we perceive into abstract categories. First, we identify individual objects as having certain shared characteristics, and then we abstract some of these characteristics in order to arrive at types of objects. Sager gives the example that we identify certain animals as having a number of common features which we group under the concept label "cats". Sager continues by saying that in a further stage of ordering, we may then group the already abstract types of objects into broader classes, e.g. we may group cats with dogs as animals, and set up separate categories for wild and domestic animals. An important distinction is thus created between the individual objects which we perceive and the abstract categories, i.e. the concepts which represent them.

We can see, from Sager's example, that classification is important in concept formation. It also plays a significant role in establishing concept systems in the subject field. Sager (1990: 28) points out that:

> In terminological theory, it is accepted that concepts should be ordered according to some conceptual classification scheme and presented in a systematic order. In order to do this, concepts are characterized by the relationships they form with neighbouring concepts.

Similarly, the ISO (ISO 1087: 4) defines a system of concepts as a "structured set of concepts established according to the relations between them, each concept being determined by its position in this set." As discussed in section 1.1.2, relations between concepts can be of many types; however, for the purpose of this thesis, we will concentrate on generic relations. A concept system formed by generic relations can be represented as an inverted tree structure, as was shown in figure 1-3.

2.1.1 Classification in bilingual thematic terminological research

Now that we have identified the importance of the concept in terminology work, and explained the notion of a concept system, we will proceed to examine how terminologists use classification in the various stages of a bilingual thematic terminological research.[13] In this type of research, there are five main stages involved in the preparation of a term record:

---

[13] There are two main types of terminology work: *term research*, which concentrates on a single term, and *thematic research*, which designates the activity of collecting and describing (as exhaustively as possible) all the terms relevant to a specific field or subfield. This work can be carried out in one or more languages. For the purpose of this thesis, I will focus on bilingual thematic research as it is the most common type of terminology work practised in Canada.

introductory reading; selection of documentation; scanning; analysis of data; and actual preparation of the term records.[14]

Introductory reading

When terminologists are assigned a new project, their first task is to familiarize themselves with the subject field. In Canada, most terminologists are language specialists, typically trained in linguistics or translation, and not true subject–field experts.[15] For this reason, they usually start by acquiring some general knowledge about the subject field through introductory readings in encyclopedia articles, basic textbooks, etc. As they familiarize themselves with the field, terminologists are able to pick out enough general conceptual characteristics from their readings that they can begin to classify the subject field into major subdivisions. Sometimes, the introductory readings contain tree diagrams showing classifications of the subject field (e.g. the classifications of optical disc discussed in section 3.1); however, different authors may present different classifications of the same subject field (cf. section 3.1), in which case, the terminologist is faced with multidimensionality. When no classification is explicitly presented in the literature, terminologists themselves often sketch out a classification of the subject field based on the information found in the readings. Terminologists may even blend the two methods, taking a classification that is presented and augmenting it with additional information from other sources.

The end result is usually a sketch of the major subdivisions of the subject field in the form of a hierarchical inverted tree diagram. This diagram represents the *concept system* of

---

[14] The stages for conducting a thematic research project that are outlined in this thesis have been compiled from the following sources: Auger and Rousseau (1978), Cole (1987), Dubuc (1985), Meyer (1992b), Picht and Draskau (1985), Rondeau (1984), and Sager (1990).

[15] Although many terminologists may become very well versed in a particular subject field in the course of their work, for the purpose of this thesis, I consider a *true subject–field expert* to be someone who has received specialized training in a particular field, e.g. molecular biology, graphic arts, electrical engineering.

the subject field. Figure 2–1, taken from Rondeau (1984: 203) shows a partial representation of the concept system for the subfield of industrial printing. The production of such tree diagrams during the introductory reading stage is recommended by several authors.

Dubuc (1985: 53) advocates:

> La fréquentation sommaire des titres retenus complétera l'initiation au domaine entreprise par le terminologue aux premières étapes de la recherche. Il est maintenant en mesure de mettre en place l'arbre de domaine qui va structurer sa recherche, en répartissant le domaine principal en quelques sous–domaines et en identifiant les domaines connexes auxquels la recherche pourra toucher.

Similarly, Auger and Rousseau (1978: 17) attest:

> Il s'agit donc pour le terminologue d'établir une liste structurée de sous–domaines à l'intérieur desquels il peut classer les activités propres au domaine à traiter. Ces sous–domaines peuvent s'organiser sous la forme d'un arbre qui schématise la structure hiérarchique du domaine.

Finally, Rondeau (1984: 70) states:

> Au cours d'une première étape, le champ notionnel du domaine (ou sous–domaine) est exploré et découpé en structure hiérarchisée à laquelle on donne généralement le nom *d'arbre du domaine*.

Although it is recommended that terminologists create representations of the subject field, as pointed out by Rondeau (1984: 72), among others, "l'établissement d'un arbre de domaine n'est pas une tâche facile, même pour le terminologue chevronné...." Rondeau advises that terminologists should proceed from the generic to the specific, and not be afraid to go down as many levels as necessary. Rondeau also warns against trying to make the tree structure symmetrical as some branches of the tree will require more development than others. Finally, Rondeau observes that some concepts will have more than one place on the tree, adding that terminologists will have to decide whether to repeat the concept or whether to have it appear only once. As we argue in section 2.3.3, there are advantages to assigning each concept only one place in the classification: when concepts are repeated, the classification becomes very cumbersome and confusing. While we are not in favour of repeating concepts

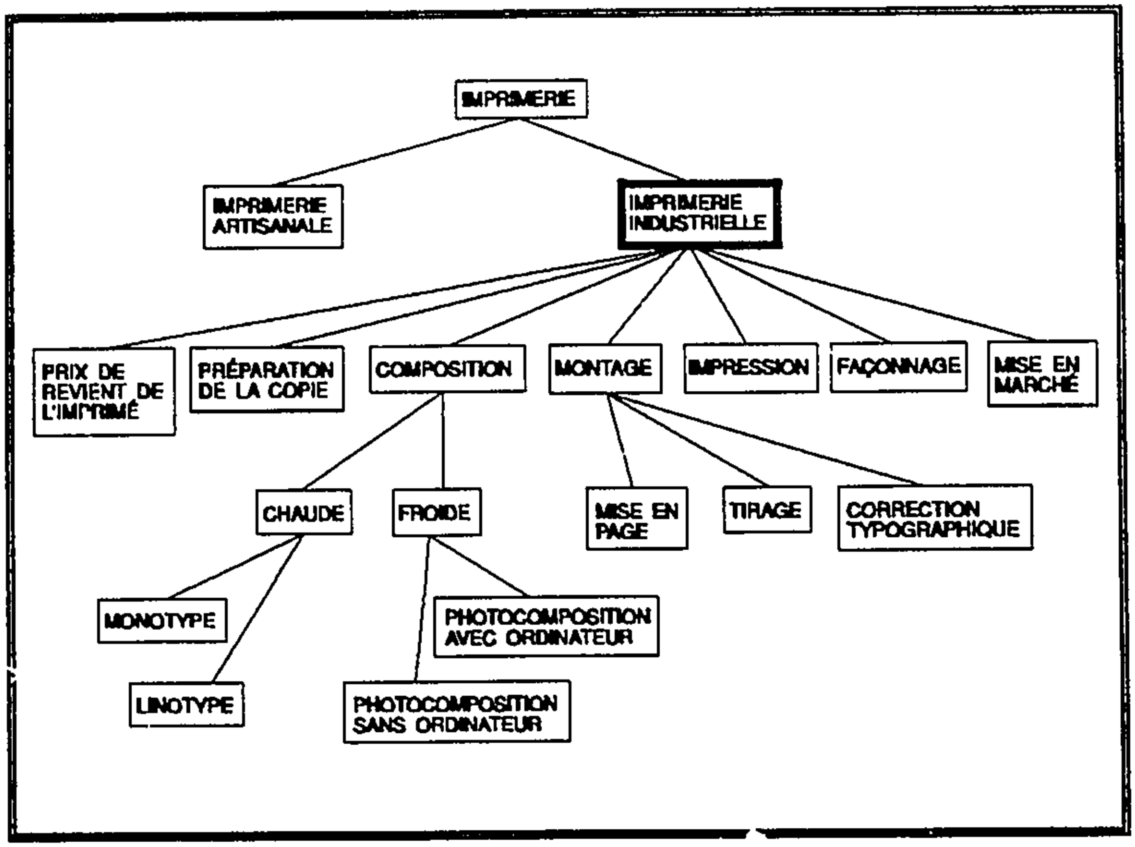


Figure 2-1. A partial representation of the concept system for the subfield industrial printing. *Source: Rondeau 1984: 203.*

in a classification, we feel that it is important for terminologists to recognize such repetition of concepts in their documentation as an indication that multidimensionality exists in the subject field (cf. section 2.3.3).

Even though it may not be an easy task, sketching out the concept system is of extreme importance. Picht and Draskau (1985: 92) stress that "The system of concepts in terminology is not a goal in itself, nor an intellectual pastime. On the contrary, it is an indispensable aid in the elaboration of a terminology." At the first stage of a thematic terminological research, sketching out a rough representation of the concept system helps terminologists to delimit the subject field to be treated. It allows them to identify the boundaries and major subdivisions of the subject field and to see the place of some of the main concepts in the subject field, as well as their relations to other concepts. (Rondeau 1984: 72; Auger and Rousseau 1978: 17; Picht and Draskau 1985: 92). This knowledge proves very useful in the documentation search and selection, which is discussed in the following section.

Selection of documentation

Once terminologists have produced a preliminary classification of the subject field, their next step is to select the documentation that they will use as a their main source of knowledge for the project. Although terminologists do value communication with subject-field experts, texts remain their principal source of information.

In the words of Rondeau (1984: 70), "Les limites du corpus à dépouiller sont fixées en fonction de l'arbre du domaine." The rough sketch of the concept system prepared during the introductory reading stage can now help terminologists to select a documentary corpus in two ways. Firstly, the names of subfields and key concepts identified in the concept system provide specific entry points into the documentation. These are the terms that can be used to do subject searches in library databases or card catalogues, or to identify pertinent sections of more general documentation using the index or table of contents.

Secondly, since terminologists tend to work from the general to the specific (i.e. top-down in a hierarchy), and to collect documentation dealing with one subfield at a time, having a basic idea of the hierarchical structure of the field enables them to concentrate their efforts on one particular subfield and to select documentation dealing with both general and specific concepts in that subfield.

Scanning

Once terminologists have selected their documentation, they proceed to *scan* it. Scanning is a process whereby the documents are carefully examined, and potential terms and their contexts are extracted. As they scan documents, terminologists inevitably learn more about the concept system and they are able to fill out their original rough sketch, whether on paper, or in their minds, or both.

At this stage of the thematic research, concept systems once again prove invaluable to terminologists. Auger and Rousseau (1978: 17) submit that "L'arbre du domaine constitue un plan de dépouillement du corpus." This idea is expanded by Dubuc (1985: 53), who states that representation of the concept system enables terminologists to evaluate the pertinence of the terminological units, allowing them to 1) identify and eliminate terms that fall outside the boundaries of the subject field, and 2) identify lacunae in the list of terms. In the first case, a terminologist conducting a thematic research in the subfield optical storage media might come across the term *magneto-optical disc* in the documentation being scanned. However, an examination of the concept's characteristics and the concept system shows that a magneto-optical disc actually belongs to the subfield magneto-optical storage media, which, along with optical storage media and magnetic storage media, is a branch of the larger subject field storage media. Therefore, the term is not directly pertinent to the subfield in question and need not be treated. In the second case, the concept system can help a terminologist to pick out concepts for which terms have not yet been identified in the documentation. For example, a terminologist working in the subfield of typesetting might find that one possible way of classifying typesetters is on the basis of whether they use digital technology. Typesetters which do use digital technology could be grouped under the collective term *digital typesetters*. Those which do not use digital technology form a second group, but the terminologist may not have found a precise term to designate this group; this would represent a terminological lacuna. Once a lacuna has been identified, the terminologist can prepare to do more research to see whether a term might indeed exist for the concept.

Rondeau (1984: 70), Sager (1990: 56), and Picht and Draskau (1985: 92) nicely sum up the usefulness of the concept system representation in the scanning stage by stating that it acts as an indicator of the exhaustiveness of the search for terms.

Analysis of data

The next stage of a bilingual thematic terminological research is to analyze the data (i.e. terms and contexts) gathered in the scanning stage. At this point, the terminologists' goal is to achieve the depth of understanding necessary to define the term,[16] establish interlingual equivalence, and identify synonymy. Terminologists carefully analyze the various contexts in which the terms have been found in order to identify the characteristics[17] of each concept. These characteristics will be used to define the term, and will be compared with those of potentially related terms (e.g. target–language equivalents, synonyms) in order to determine conceptual matches.

Defining a term. Probably the most important application of concept analysis is definition construction. To accurately define a concept, it is absolutely essential for terminologists to have a clear understanding of the concept's characteristics. As explained in more detail by Eck (1992), the terminologist can compare the characteristics of co–ordinate concepts (i.e. concepts which rank at the same hierarchical level in the concept system) to establish the differences required for creating an intensional definition, the most common type of definition in terminology work. The intensional definition has two main parts: genus and

[16] Some terminologists do not handcraft their own definitions; instead, they take them from contexts gathered in the scanning stage. Even so, it is necessary for the terminologist to have a good understanding of the concept's characteristics in order to choose the most accurate definition.

[17] Here we are using the term *characteristic* in the general sense of 'attribute,' and not in the narrower sense of 'differentiating characteristic' as discussed in section 1.1.1.

differentia. When terminologists define a concept in this way, they refer to 1) the generic class to which the concept belongs, and 2) the characteristic(s) which differentiates this concept from other members of the same class. For example, an intensional definition for the concept **rifle** could be: a rifle is a *firearm* (genus) with a *spirally grooved bore* (differentia). *Firearm* represents the generic class to which **rifle** belongs. Other members of this class, such as **pistol** and **revolver**, are co-ordinate concepts of **rifle**, but none of them have of them have a spirally grooved bore, and this is what makes them different from a **rifle**.

Establishing interlingual equivalence. In bilingual terminology work, concept analysis is crucial to establishing interlingual equivalences (Auger and Rousseau 1978: 40; Picht and Draskau 1985: 92). Dubuc (1985: 72) asserts that in order to determine interlingual equivalence, it is necessary to establish a "crochet terminologique," which he defines as "les descripteurs communs aux contextes accompagnant les vedettes d'une fiche terminologique." In order to establish a "crochet terminologique," terminologists must examine the characteristics of the concepts in both languages and compare them to see if they match up. For example, in order to establish the equivalence between the English term *pit* and the French term *microcuvette* in the subfield of optical storage media, a terminologist would have to compare the characteristics of the concepts that are designated by the two terms. An analysis of the following two contexts reveals a "crochet terminologique" composed of six common characteristics, which are shown in table 2.1.

English context: "When the laser emitted from the writing head hits the surface of the disc, it melts a specific area in the recording layer. The result of this process is an oblong depression called a *pit*. However, unlike a phonograph record, the track is neither a groove nor a continuous line but only marks forming a spiral of a broken line. The oblong pits are 0.12 microns (i.e. micrometers or millionth meters) deep and about 0.6 microns wide. The minimum length of a pit is 0.9 microns. The maximum length is 3.3 microns. The distance between the tracks in the spiral is 1.6 microns." (Elshami 1990: 3)

French context: "La couche métallique réfléchissante du CD-ROM comporte des *microcuvettes* de 0,12 micron de profondeur sur 0,6 de large, disposées sur une piste unique en spirale, selon un pas de 1,6 micron. La longueur de ces microcuvettes est variable." (Dargery 1989: 122).

English	French
· located on recording layer of disc	· la couche métallique réfléchissante du CD-ROM les comporte
· 0.12 microns deep	· 0,12 micron de profondeur
· 0.6 microns wide	· 0,6 [micron] de large
· minimum length is 0.9 microns, maximum length is 3.3 microns	· la longueur est variable
· track ... forming a spiral	· disposées sur une piste unique en spirale
· distance between the tracks in the spiral is 1.6 microns	· selon un pas de 1,6 micron

Table 2.1. Characteristics which constitute the "crochet terminologique" for the terms *pit* and *microcuvette*, in the subfield of optical storage media.

Identifying synonyms. Identifying synonyms also requires a very careful comparison of conceptual characteristics to determine if the characteristics are indeed identical for the terms in question. Picht and Draskau (1985: 131) observe that:

> ... in defining or delineating the concepts in relation to one another, ... synonymy is confirmed or what was initially taken for synonymy is disproved, even in the face of generations of dictionaries which make contrary claims.

Thus, in a similar fashion to that used to establish interlingual equivalence, synonymy is established by matching the characteristics of concepts that are designated by different terms in the same language. When concepts differ in only a few minor characteristics, they may be designated as pseudo-synonyms (e.g. one concept may have one more characteristic than another, and thus be more specific).

Preparation of term records

The final stage of a thematic terminological research is the actual preparation of the term records. To prepare a term record, the terminologist simply compiles the relevant data gleaned from the analysis into a series of term records (one record per concept). There are no definitive rules as to what information must appear on a bilingual term record; however, it is generally accepted that the record will contain at least the entry terms, the subject field, relevant grammatical information (e.g. part of speech, gender, usage notes), a definition, synonyms (if there are any), and sources.

Unfortunately, for the most part, traditional term records provide minimal, if any, indication of the relations between concepts in the subject field or of an individual concept's place within the concept system.[18] In the words of Sager (1990: 147), "This [relations] is the most controversial and least defined category of information." However, it is encouraging to note that, in a section entitled *New trends in compilation* (Sager 1990: 160), Sager has included *conceptual relationships* as a category of information to be recorded on a term record. Sager adds that technological advances will also make it easier to preserve actual representations of concept systems within term banks of the future. Indeed, this is already being done in the COGNITERM term bank.[19]

In summary, we have seen that although the concept system underlying thematic terminological research is not explicitly represented on conventional term records,[20] this

[18] Dubuc (1985: 82) suggests that *marques logiques* can be used to express four basic relations on a term record: *générique-spécifique (G/S), cause-effet (C/E), partie-tout (P/T),* and *concret (réel)-abstrait (R/A).*

[19] For more details on the COGNITERM term bank, please refer to Meyer, Bowker and Eck 1992 a and b; and Meyer, Skuce, Bowker and Eck 1992 a and b.

[20] Information pertaining to concept relations, such as generic-specific relations, may be gleaned from definitions or contexts found on term records; however, this information is embedded within free natural language text and may not be immediately obvious.

classification is nonetheless essential to the preparation of a term record. Terminologists benefit from mapping out and refining the concept system and using it to guide them at each stage of their work; it can help them delimit the subject field, select their documentation, evaluate the exhaustiveness of their research, and establish definitions, interlingual equivalence and synonymy.

2.1.2 Why is multidimensionality pertinent to terminology?

In this section, we will consider two important reasons why multidimensionality is pertinent to terminology. Firstly, it allows terminologists to gain a deeper understanding of the entire subject field, and secondly, it reduces the risk of omitting concepts that belong to only one dimension of the subject field.

It is quite conceivable that multidimensionality may have other implications in terminology work, for example, in definition writing. This and other possible implications would be interesting subjects for future research, but will not be dealt with in the present thesis.

Understanding the subject field

Before terminologists begin producing term records, it is important that they have a clear understanding of the subject field in question, which entails recognizing the different ways of classifying the field. We established in section 2.1.1 (*Introductory reading*) that, as terminologists do their introductory reading, they often sketch out a rough outline of the subject field based on this reading. It was also noted in the same section that different authors may present different classifications of the same subject field, in which case the terminologist is faced with multidimensionality.

Clearly, it is much easier for terminologists to work with a single dimension; however, we feel that a unidimensional understanding of the subject field is much more superficial and simplistic than a multidimensional one. Picht and Draskau (1985: 69) echo this sentiment when, in their discussion on unidimensional representations,[21] they state "the advantage of this system is its clarity; its drawback is its limited expressivity, since, after all, only one of a number of possible classifications can be represented."

Until recently, terminologists have not had adequate tools for dealing with multidimensionality; however, recent developments in the field of artificial intelligence (AI) have provided tools that make it easier for terminologists to handle multidimensionality. One such tool, called *CODE (Conceptually Oriented Description Environment)*, will be discussed in detail in Part III of this thesis.

Even with the help of AI tools, multidimensionality can still be a confusing phenomenon. For the sake of clarity, as we will propose in section 2.2.3, terminologists may find it easier keep track of individual dimensions separately and work on understanding and developing them one at a time. The important thing is not to ignore any of the dimensions completely as this may have repercussions for the end user of a term bank or terminological publication, as discussed in the following section.

Omission of concepts

A very practical and extremely important reason for acknowledging multidimensionality in terminological research is to avoid omitting concepts that are relevant to the subject field at hand. The primary function of a bilingual term record is to provide translators with an equivalent for a term found in their source text. It is possible that the text being translated includes a term that appears in only one dimension of a subject field. If the

[21] Picht and Draskau (1985: 68-9) refer to unidimensional representations as *monodimensional* or *monohierarchic systems*.

translator consults a term bank or terminological publication which contains only terms belonging to a different dimension of the subject field, the translator will be out of luck.

For example, suppose a translator is translating a text in the field of optical storage media, and he or she requires a translation for the term *replicable disc*. If the terminologist had classified optical storage media only on the basis of the characteristic **degree of writability**, and had not considered the dimension based on the characteristic **replicability**, then the term *replicable disc* might not be found in the term bank or terminological publication.[22] It could be argued that terminologists would come across terms from various dimensions when scanning the documentation; however, there is a slight danger that they may not consult documentation related to a specific dimension if this dimension is ruled out in the early stages of the research. Since terminologists have no way of knowing which terms are of interest to any given end-user, they must take into account all possible dimensions in their classifications to ensure that their term records will meet the needs of the widest possible group of end-users.

2.1.3 How do terminologists deal with multidimensionality?

Having established that multidimensionality is indeed pertinent to terminology work, we will now consider how both terminology researchers and practising terminologists attempt to deal with this phenomenon.

Terminology researchers

Some terminological authors (Dubuc 1985, Auger and Rousseau 1978) have completely ignored the issue of multidimensionality. Others (Rondeau 1984, Picht and

[22] This problem is discussed in more detail in section 3.2.1 (*Problem 2*) as part of an extended example.

Draskau 1985, Sager 1990) recognize that multidimensionality exists, but they do little more than acknowledge this existence.

For example, Picht and Draskau (1985: 68, 70) acknowledge that:

> ... in reality, we all know that concepts can be ordered according to more than one ordering characteristic[23] at the same time, and this may be realized up to a certain point. But the limitations soon appear. ...when the system becomes so complex that it ceases to contribute to the elucidation of the relations and degenerates into a puzzle for cross-word enthusiasts.

This complexity is one of the main reasons that so little attention has been given to the issue of multidimensionality. Until recently, the only tools available to terminologists were pencil and paper. Unfortunately, a two-dimensional medium (i.e. paper) is not conducive to representing multidimensional classifications. As Sager (1990: 9) observes:

> A great deal of attention has ... been devoted to the structure of conceptual systems and the best way of representing them on paper. In this area in particular, conventional thinking is still largely dominated by pen and paper processing techniques and by the relatively simple relationships that suffice for structuring documentation thesauri and respond well to human attribution and processing.

Later in his book, Sager (1990: 147) drives home the point that sophisticated tools are necessary to deal with this problem:

> Full conceptual systems based on the sets of relationships which are found to be the most informative and effective for any one subject field have hitherto been developed only on paper. Complex and multidimensional conceptual systems can only be represented and managed by computer; it was therefore impossible until quite recently to have anything other than rather prescriptive and narrow conceptual systems.

Perhaps terminology researchers will give more consideration to the issue of multidimensionality in future works, now that tools are available to help them handle this phenomenon more easily.

---

[23] Picht and Draskau use the term *ordering characteristic* in the same way that we use the term *differentiating characteristic* (cf. section 1.1.1).

Practising terminologists

Practising terminologists do not generally concern themselves with attempting to construct multidimensional representations because, as previously noted, they usually do not have adequate tools to do so. They may, as mentioned in Rondeau (1984:72) deal with multidimensionality by repeating those concepts that belong to more than one dimension (cf. section 2.3.3).

Also, since terminologists do not generally include representations of concept systems in term banks or terminological publications,[24] they may include terms without indicating specifically how they fit into the concept system. In other words, terminologists may recognize that a term is relevant to a subject field, but be unable to pinpoint its exact location in their unidimensional representation; however, they can still include the term in their term bank or terminological publication because they do not usually provide a representation of the concept system. For example, as discussed in section 1.2.1 (*Changes in the reality of a phenomenon*) and illustrated in figures 1–11 and 1–12, HOT TYPESETTING and COLD TYPESETTING were originally the two main categories of typesetting. Since then, computerized typesetting techniques have been developed, and the distinction between HOT TYPESETTING and COLD TYPESETTING is no longer relevant. Nonetheless, the term *cold typesetting* is still occasionally used to describe modern typesetting methods. A conscientious terminologist would therefore want to include the term *cold typesetting* in the vocabulary,[25] even though it would be difficult to work the concept into the (unidimensional) representation of the concept system.

---

[24] The Finnish terminologists at the *Tekniikan Sanastokeskus* are a notable exception as they include tree-structure diagrams, which indicate the relations between concepts in a subject field, in both their term bank and their terminological publications.

[25] Indeed the term *cold typesetting* is found in *Terminology Series: Graphic Arts* (Neal and Paradis 1986), and in TERMIUM III, the terminology bank of the Department of the Secretary of State of Canada.

2.2 CLASSIFICATION AND MULTIDIMENSIONALITY IN BIOLOGICAL TAXONOMY

Classification plays a very important role in the field of biology. Jeffrey (1989: 1) notes that one of the first tasks of biology was to make meaningful generalizations about living organisms so that useful knowledge could be passed on from person to person. For example, early in human history, it was found useful to know which plants were edible and which were poisonous, which animals were dangerous and which were good to hunt for food. It was soon noticed that living organisms possessed certain consistent features, such as being edible or being poisonous, by which they could be reliably identified and classified into constantly and recognizably distinct groups.

Since that time, biologists have identified well over a million species of animals and plants, and according to Hickman (1986: 326), thousands more are identified each year. Thus, we can see that there are many, many different kinds of organisms with a vast range of size and structure. Obviously it would be impossible to remember all the characteristics of each individual organism; therefore, it is necessary to have some means of grouping the organisms and summarizing their characteristics. To meet this need, biologists have developed formal systems for classifying and naming organisms. In the words of Jeffrey (1989: 2):

> ...such systems are hypotheses of the way in which variation in the living world is ordered and are used in biology for the storage, retrieval and communication of information and for the making of reliable predictions and generalizations.

Thus, we can see that a formal classification system enables biologists to compare the characteristics of a potentially new species against those of previously identified species, and thereby determine its place in the classification.

2.2.1 *Classification methods*

The modern science of identification and classification is called *taxonomy*.[26] One of the most important features of taxonomic classification is that it is hierarchical. There is a set of seven major taxonomic categories to which groups of plants or animals are assigned for the purposes of classification. These categories are employed in a conventional order which must be strictly adhered to.[27] Figure 2-2 shows the taxonomic hierarchy, the lowest category of which is the category *species*.[28] The ultimate goal of the biological taxonomist is to assign each living organism to a particular species.

In order to do this, the taxonomist must first identify *characters*[29] which can be used to describe an organism adequately and place it in the proper species. Biological taxonomists make an important distinction between *homologous* characters and *analogous* characters. Homologous characters are ones that have a common evolutionary origin, whether or not they have a similar function. For example, the forelimbs of dogs and humans are homologous, as are those of birds and whales. In contrast, analogous characters have a common function, but were not derived from a common ancestor. An example of analogous structures are the wings of an insect and those of a bird.

---

[26] Several authors (Jeffrey 1989: 1, Hickman 1986: 330, Mitchell 1988: 436) note that the terms *taxonomy* and *systematics* are frequently, though incorrectly, used as synonyms. According to Hickman (1986: 330), taxonomy is the science concerned with naming and classifying organisms, while systematics is a broader science, including both taxonomy and phylogeny, which is the study of the evolutionary history of organisms.

[27] The seven major taxonomic categories are as follows, in descending order: kingdom, phylum, class, order, family, genus, and species.

[28] A species is defined by Jeffrey (1989: 2) as "a series of similar individuals showing certain common features distinct from other such series." He adds that in sexually reproducing organisms, individuals of a species are reproductively isolated from individuals of other species.

[29] Biologists use the term *character*, to designate what we have been calling a *characteristic*. Hickman (1986: 331) defines a character as "any feature or attribute that can be described, measured, weighed, pictured, counted, scored, or otherwise defined about an organism."

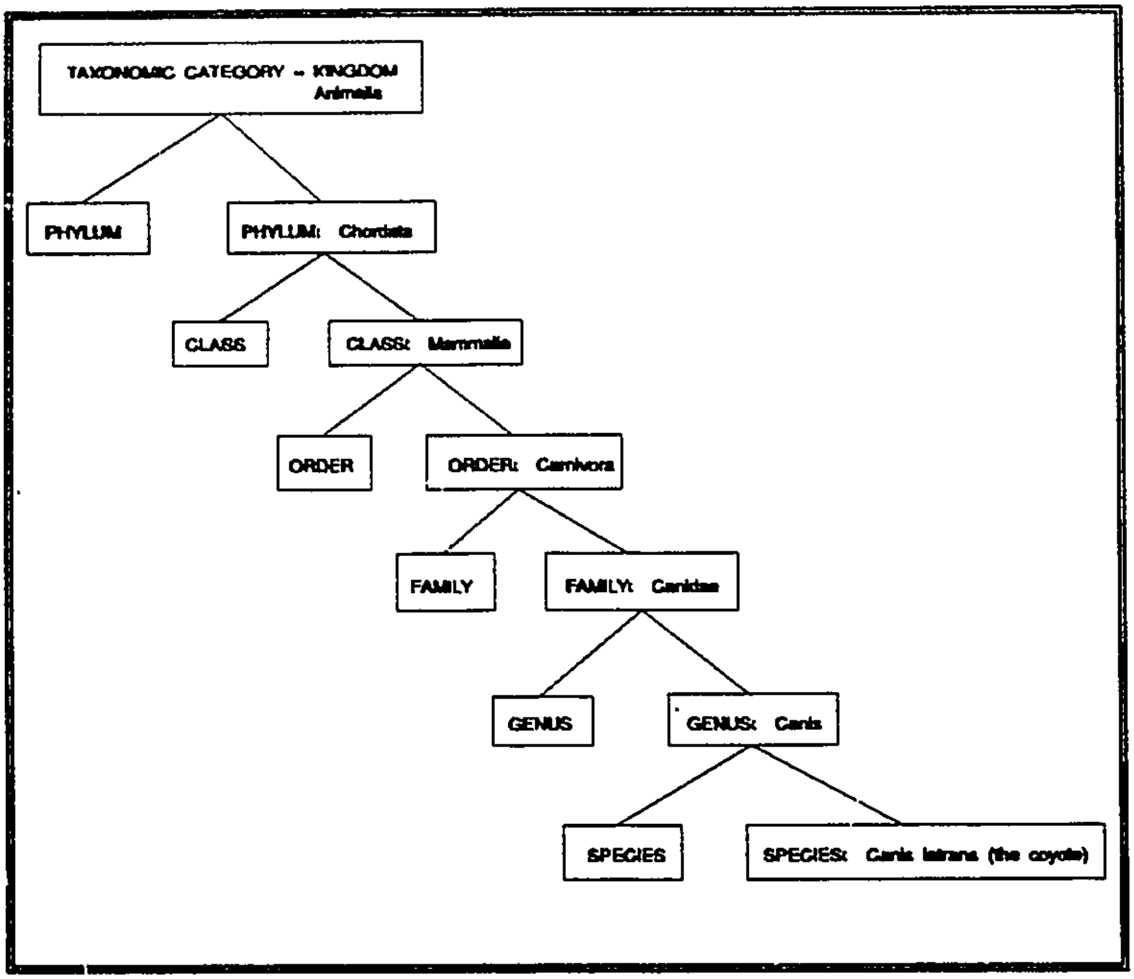


Figure 2-2. A taxonomic hierarchy showing the taxonomic categories. *Source: Gage Canadian Dictionary 1983: 215.*

There are currently three main approaches to taxonomic classification: *evolutionary systematics*, *phenetics*, and *cladistics*.

Evolutionary systematics

Evolutionary systematics is the traditional approach; it has dominated taxonomic classification since evolutionary theory was first developed by Darwin. Evolutionary systematics attempts to reconstruct evolutionary history as closely as possible by tracing the actual ancestor–descendent relationships between organisms. Species are grouped on the basis of shared *phylogeny* (i.e. evolutionary history) and *morphological* (i.e. structural) similarity. The relationships between the species are graphically depicted in the form of a *phylogenetic tree*, where the branches represent the appearance of new species. A geological time scale is usually included to indicate at what point in history the species evolved.

Figure 2-3 shows a phylogenetic tree representing a classification of animals that was made following the methods of evolutionary systematics. Note that CROCODILES and BIRDS have been separated into different classes. Biologists know that CROCODILES and BIRDS share a more recent common ancestry than do CROCODILES and other living REPTILES; nevertheless, evolutionary systematists have grouped CROCODILES with REPTILES because CROCODILES share with REPTILES important biological characteristics, such as being cold blooded, that they do not share with BIRDS. Thus we can see that this division has been made on the basis of both phylogenetic relationships and morphological similarities.

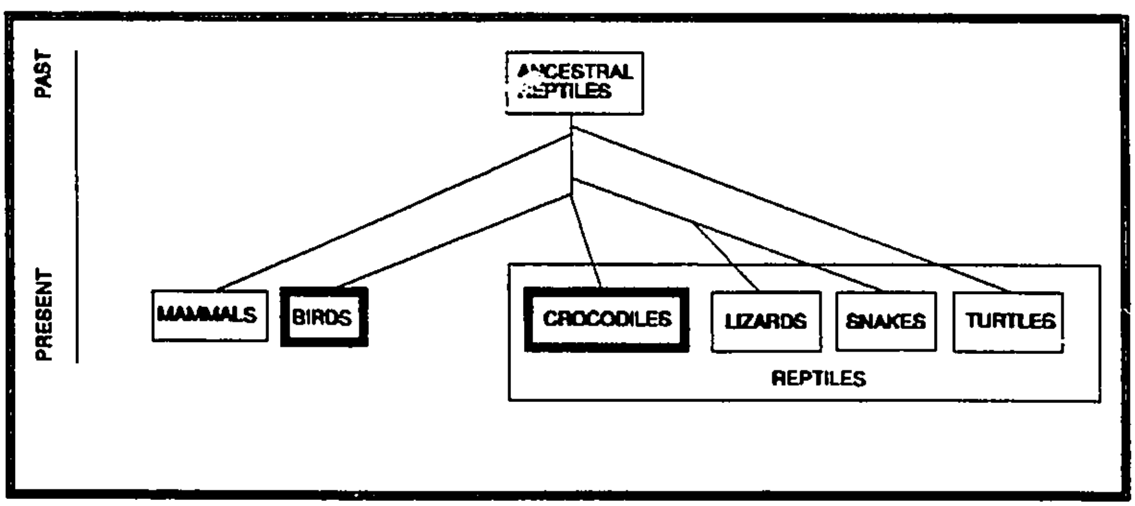


Figure 2-3. A phylogenetic tree representing the classification of animals made following the methods of evolutionary systematics. *Source: Dorit 1991: 546.*

Phenetics

In contrast with evolutionary systematics, which is based on both the phylogenetic relations and the morphological similarities of organisms, phenetic classification groups organisms solely on the basis of their total similarities and dissimilarities; it is an explicitly quantitative approach. In a phenetic classification, as many arbitrarily chosen, equally weighted characters as possible[30] are selected, coded, and fed into a computer. These characters can be either homologous or analogous because, as pointed out by Dorit (1991: 547), pheneticists believe that, provided enough species are examined, similarity due to common ancestry will always outweigh analogical similarity. The computer analyzes the characters; species are then grouped on the basis of their overall similarity, which is assumed to be a reliable indication of their common ancestry. In cases where extensive convergence between two groups of organisms has occurred, a phenetic classification might group them together, despite their dissimilar ancestry. For example, some pheneticists might well group BIRDS and BATS together since they are both flying vertebrates, regardless of the fact that they evolved from different ancestors.

Once the computer has analyzed and grouped the species, a *phenogram* is constructed, which shows the percentage of similarity between species. A phenogram is a graphic representation of a grouping of organisms on the basis of overall similarity; it makes no explicit attempt to reflect the evolutionary history of the groups involved. Figure 2–4 shows a phenogram with a percentage scale.

Cladistics

The third and most recent approach to scientific classification is known as cladistics or *phylogenetic systematics*. Cladistics attempts to reconstruct the course of evolution by

---

[30] According to Hickman (1986: 330), well over 100 characters are chosen.

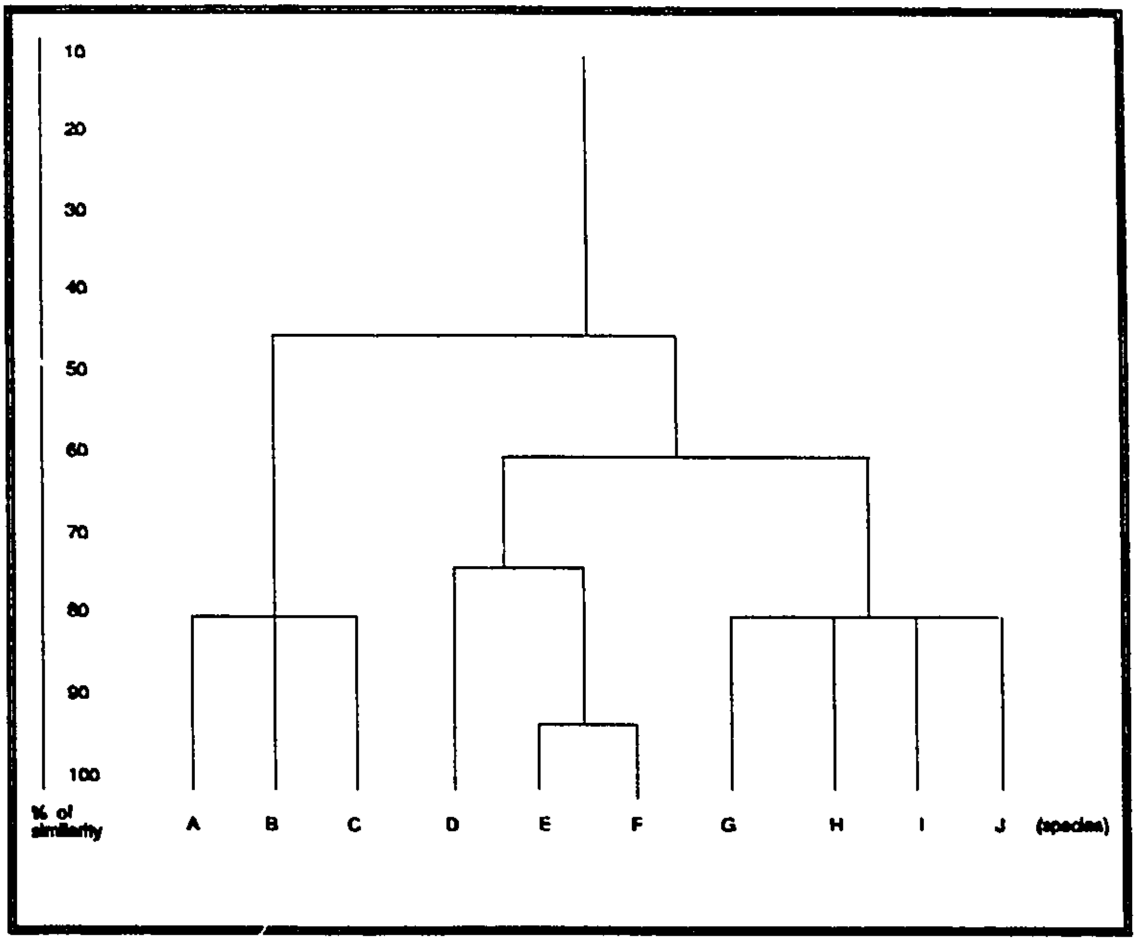


Figure 2–4. A sample phenogram with a percentage scale. *Source: Goto 1982: 56.*

grouping organisms in terms of the relative recency of common ancestry. This approach bases classification exclusively on phylogeny; morphological differences are not taken into account. When determining the recency of common ancestry, cladists use only homologous characters, which they divide into two subtypes: *primitive characters* and *derived characters*. A primitive character is one which is shared with a distant common ancestor (e.g. having DNA as a genetic material), while a derived character is one that is shared with a more recent common ancestor (e.g. having hair).

In practice, cladists are interested in studying the evolutionary relationships of certain groups of organisms, for example, two species within a genus. Cladists then select another group (i.e. a third species), which they believe is only distantly related to the groups of interest; this distantly related group is called the *outgroup*. Next, a variety of homologous characters are chosen that will be used to compare among the more closely related groups and the outgroup. If the outgroup has been correctly selected, all the characters that it has in common with the related groups may be considered primitive; that is, they are characters possessed by the common ancestor of both the outgroup and the related groups. Characters shared by the related groups but not possessed by the outgroup are derived. The greater the number of shared characteristics, the more closely the groups are related.

Cladists record the results of their classification in a *cladogram*, which resembles a phylogenetic tree with no time line; this is because a cladogram is intended to show only relative degree of relationship and not actual historical events. The cladogram in figure 2-5 shows a classification of animals where AMPHIBIANS is the outgroup. Note that, in contrast to the evolutionary systematic classification that we saw in figure 2-3, CROCODILES and BIRDS have been grouped together since they share certain derived characteristics. This classification has been based solely on phylogeny; the morphological differences between CROCODILES and BIRDS have not been taken into consideration.

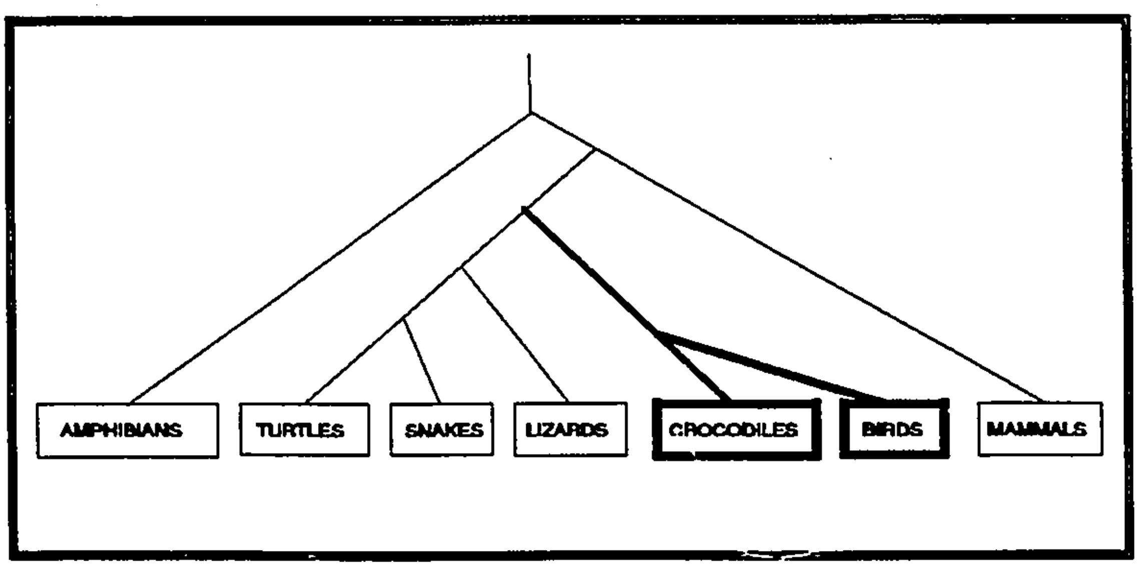


Figure 2-5. A cladogram representing the classification of animals according to cladistic methods. ***Source: Dorit 1991: 546.***

2.2.2 How do biological taxonomists deal with multidimensionality?

Although biologists do recognize the existence of multidimensionality in biological classification, as is demonstrated by the existence of three mutually exclusive classification methods, practising taxonomists are generally not concerned with the theories behind classification methods. Rather, they are given a taxonomy to use as their frame of reference and they are expected to work within it.

Taxonomists take such a black and white approach to classification because their main purpose in classifying organisms is to place them in a particular species. They are not overly concerned with how other taxonomists classify the organism; rather, their primary concern is to fit the organism into its proper place in their own classification, according to the rules of their chosen method.

2.2.3 What can terminologists learn from biological classification methods?

Even though taxonomists do not represent multidimensionality in their individual classifications, they do acknowledge its existence, and terminologists can learn several useful techniques from biological classification methods. Both biological and terminological classification are based on the hierarchical relations between objects. Taxonomists have recognized the usefulness of systematically representing these hierarchical relations in graphical form. Biology textbooks are filled with tree-structure diagrams[31] representing classifications of living organisms; in fact, taxonomists prefer to represent their classifications graphically instead of textually. Although terminologists also acknowledge the importance of the relations between objects in a classification, the majority of term banks and terminological publications which they produce do not include tree-structure diagrams to show these

[31] These diagrams are either phylogenetic trees, phenograms, or cladograms, depending on the classification approach favoured by the author of the book.

relations; instead, terms are listed alphabetically, along with definitions and equivalents. This is a logical format because it allows users to have quick, easy access to any given term; however, as pointed out by Picht and Draskau (1985: 129):

> The principal disadvantage of alphabetical ordering lies in the 'arbitrary' ordering of concepts, for the alphabet prevents a conceptually coherent representation. The relationships between concepts cannot be expressed — yet these constitute a central element in the process of understanding and assessment.

We feel that there is no reason why, in principle, a graphical representation of the subject field (or parts of the subject field) could not be included to complement the alphabetical listing.[32] After all, when terminologists begin learning about a new subject field, they themselves often sketch out the relationships between objects in the field in the form of tree diagrams. If this type of representation is helpful to terminologists, there is little doubt that it would also be helpful to those consulting the terminological publication or term bank.

Secondly, terminologists and taxonomists take the same general approach to classification: both disciplines carefully identify and analyze characteristics in order to classify objects. However, taxonomists are more subjective in their choice of characteristics; in theory, every taxonomist could produce a classification of living organisms based on those characteristics that he or she feels are important for distinguishing between the organisms. The choice of characteristics is largely influenced by the taxonomist's preferred classification method. Hence, biological classifications are always unidimensional. Terminologists are not afforded the luxury of choosing which characteristics they will use to classify the object. Because their principal goal is to provide definitions and equivalents for all the terms in the given subject field, ideally, terminologists ought to represent all possible dimensions, which entails considering all characteristics.

---

[32] We say *in principle* because in practice it is very difficult, messy, and space consuming to represent multidimensionality in glossaries, and it may be technically impossible, or at least very expensive, to do so in many existing term banks.

Nonetheless, we feel that terminologists could perhaps learn something from the taxonomists' unidimensional approach. Since multidimensionality can be a confusing issue, terminologists might be wise to start by working with only a single dimension. In our limited experience, we have found that during the introductory reading, one dimension tends to stand out as being more common or important than the others. It is possible that if terminologists choose this dimension and work on understanding it first, they will avoid the confusion and information overload that may come with trying to understand all dimensions simultaneously. Terminologists may find it easier to go back and work on developing a second dimension, and then a third, etc., once they have a clear understanding of the most common dimension. When they clearly understand and have developed unidimensional representations of each the individual dimensions in the subject field, terminologists will likely find it easier to integrate the unidimensional representations into a single multidimensional representation. The possibility of being able to represent multiple dimensions, and yet work with one dimension at a time, is provided by artificial intelligence tools, such as the CODE system (cf. chapters 5 and 6), which offer mechanisms for temporarily hiding unwanted dimensions. This feature is appealing both to the terminologists conducting the thematic research, and the end-users consulting the completed TKB.

2.3 CLASSIFICATION AND MULTIDIMENSIONALITY IN LIBRARY SCIENCE

Classification is of prime importance in the discipline of library science. Library scientists are principally concerned with classifying knowledge as it appears in documents. They must organize the documents in the library in such a way that their contents will be accessible to users. In the words of Shera (1976: 70), "classification, the arrangement of the books as they stand on the shelves, has long been thought to be the key to subject access to the library's store...." Early classification systems consisted of large and somewhat random groupings of documents in the major subject fields, but as libraries grew in size, these groupings became inadequate. And so, in the late nineteenth and early twentieth centuries, classification theorists began designing classification systems that would serve the growing needs of the library.

2.3.1 Classification methods

In the field of library science, the most important part of the classification process is the classification of subjects[33] into appropriate groups. If all documents dealt with a simple[34] subject, the classification process would be fairly easy. There would be a separate class for each subject; all documents dealing with a certain subject would belong to the same class, and would therefore be kept in the same location. For example, there would be a separate class for each of the following subjects:

sex
children
psychology
teaching

Unfortunately, as pointed out by Hunter (1988: 6), among others, not all documents deal with a simple subject. In fact, most documents deal with a complex subject, which is a combination of simple subjects. Examples of complex subjects would be:

psychology of sex
child psychology
teaching children about sex
teaching psychology
teaching child psychology

---

[33] In the field of library science, the *subject* is the topic (or topics) dealt with in a document.

[34] The term *simple* is used in the library science literature in the sense of 'single' and not in the sense of 'easy to understand'.

When a document deals with a complex subject, it is much harder to classify. For example, does a document about teaching child psychology belong to the class TEACHING, or the class CHILDREN, or the class PSYCHOLOGY? Undoubtedly, every library user has had the experience of not knowing exactly where to look for a needed document, particulary when the document has a complex subject.

A library classification consists of three main parts: a *schedule*, which is a hierarchical list of the subjects; a series of *notations*,[35] which are codes assigned to subjects so that a document dealing with a given subject can be easily filed and retrieved; and, an *alphabetical index*, which lists the subjects in alphabetical order, along with their corresponding notations.

There are currently two principal methods for classifying documents in libraries: *enumerative classification* and *faceted classification*.

Enumerative classification

Traditional classification systems are enumerative: they attempt to list all possible subjects, both simple and complex, within the defined subject field or fields. Two well-known enumerative classification systems are the Dewey decimal classification, which is used in most public libraries in North America, Britain, and some other countries, and the Library of Congress classification, which is used in most university libraries in North America.

To begin an enumerative classification, the librarian must first decide on the body of knowledge to be classified.[36] This body of knowledge is taken as the first class, and it is then hierarchically divided into subclasses on the basis of differentiating characteristics, until every subject has been included. "In a properly designed hierarchical classification, each subject should have only one place where it fits into the system" (Hunter 1988: 35). The

[35] The notations must have a self-evident order so that users can find their way around the classification. Examples of self-evident order can be seen in the number system (e.g. 1 is followed by 2, which is followed by 3, etc.), and the alphabet (e.g. A is followed by B, which is followed by C, etc.).

[36] Most libraries attempt to classify all knowledge; however, specialized libraries, such as a law library or a music library, classify knowledge within a restricted subject field.

librarian must give careful consideration to the order in which the characteristics are applied because this will determine which classes have their documents kept together, and which have theirs scattered. For example,[37] figure 2–6 demonstrates that documents on VERTEBRATES will be found together, as will documents on INVERTEBRATES; however, documents on RESPIRATION or REPRODUCTION will be found in four different places. The order in which the characteristics are applied should be chosen on the basis of the clients' needs. The order shown in figure 2–6, in which the animals are divided first on the basis of **scientific classification** and later according to **physiological process**, is probably the most helpful order in a general library; however, if the library were a zoological research library whose clients were primarily interested in animal processes rather than the animals themselves, then it would be more helpful to apply the characteristics in the reverse order.

Up to this point in the thesis, we have seen hierarchical classifications represented in the form of tree structures, such as the one shown in figure 2–6. Librarians, however, tend to draw up schedules in the form of a list, using indentations to represent hierarchical relations. Table 2.2 shows a schedule containing the same information found in the tree structure diagram in figure 2–6.

[37] The example used in figure 2–6 and tables 2.2 – 2.4 is derived from an example given in Buchanan (1979: 26).

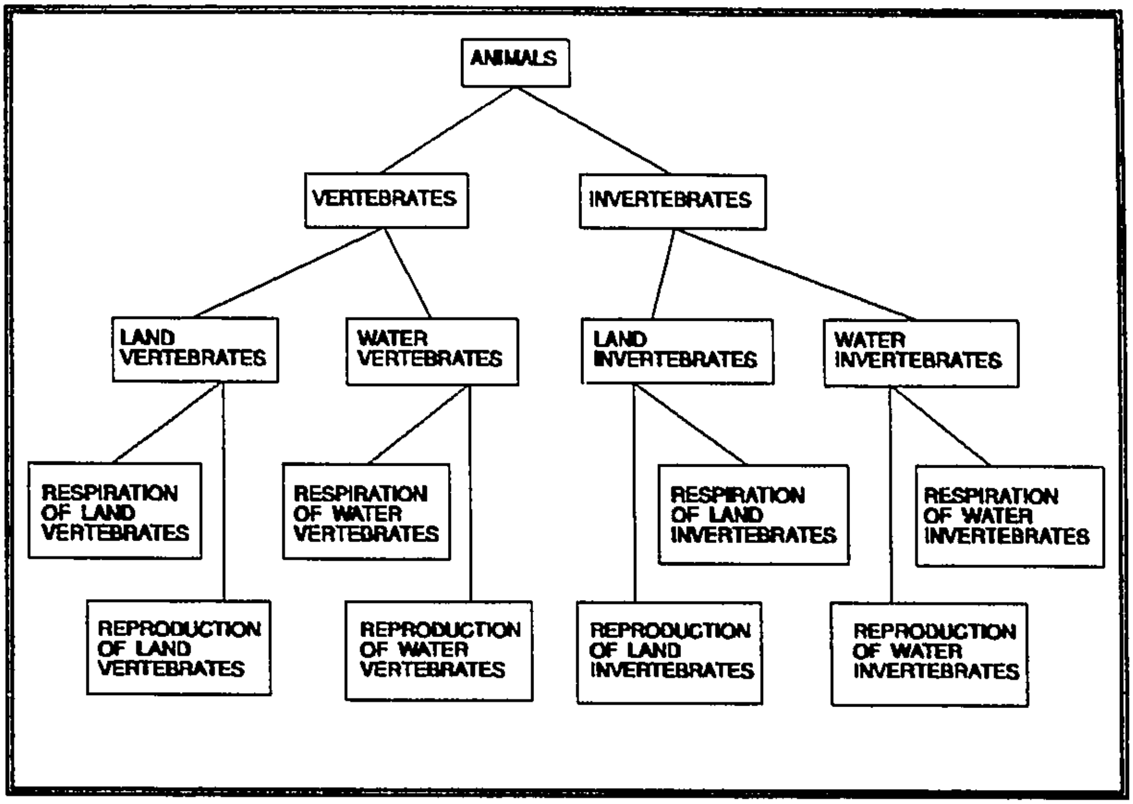


Figure 2-6. An enumerative classification of ANIMALS.

ANIMALS
- VERTEBRATES
  - LAND VERTEBRATES
    - RESPIRATION OF LAND VERTEBRATES
    - REPRODUCTION OF LAND VERTEBRATES
  - WATER VERTEBRATES
    - RESPIRATION OF WATER VERTEBRATES
    - REPRODUCTION OF WATER VERTEBRATES
- INVERTEBRATES
  - LAND INVERTEBRATES
    - RESPIRATION OF LAND INVERTEBRATES
    - REPRODUCTION OF LAND INVERTEBRATES
  - WATER INVERTEBRATES
    - RESPIRATION OF WATER INVERTEBRATES
    - REPRODUCTION OF WATER INVERTEBRATES

Table 2.2. A schedule for an enumerative classification of ANIMALS.

Once the subject matter has been fully classified, the librarian will then assign a notation to each subject, thus allowing it to be represented by a code or classification number. When a notation is attached to a document, it acts as an address which makes it easy to file and retrieve the document. Therefore, it is helpful for the notations to have a logical order, such as that of numbers or letters, which will permit users to find their way around the filing arrangement easily. As table 2.3 demonstrates, subclasses inherit the complete notation of their generic classes, and then an additional symbol is added.

1	ANIMALS
11	VERTEBRATES
111	LAND VERTEBRATES
1111	RESPIRATION OF LAND VERTEBRATES
1112	REPRODUCTION OF LAND VERTEBRATES
112	WATER VERTEBRATES
1121	RESPIRATION OF WATER VERTEBRATES
1122	REPRODUCTION OF WATER VERTEBRATES
12	INVERTEBRATES
121	LAND INVERTEBRATES
1211	RESPIRATION OF LAND INVERTEBRATES
1212	REPRODUCTION OF LAND INVERTEBRATES
122	WATER INVERTEBRATES
1221	RESPIRATION OF WATER INVERTEBRATES
1222	REPRODUCTION OF WATER INVERTEBRATES

Table 2.3. An enumerative classification of ANIMALS with notations.

Finally, the librarian compiles an alphabetical index of the subjects within the classification so that users can easily find the corresponding notation and thus ascertain the location a particular document. Table 2.4 shows an alphabetical index for the enumerative classification of ANIMALS.

ANIMALS			1
INVERTEBRATES			12
LAND:	INVERTEBRATES		121
	VERTEBRATES		111
REPRODUCTION:		LAND INVERTEBRATES	1212
		LAND VERTEBRATES	1211
		WATER INVERTEBRATES	1222
		WATER VERTEBRATES	1122
RESPIRATION:		LAND INVERTEBRATES	1111
		LAND VERTEBRATES	1121
		WATER INVERTEBRATES	1221
		WATER VERTEBRATES	1121
VERTEBRATES			11
WATER:	INVERTEBRATES		122
	VERTEBRATES		112

Table 2.4. An alphabetical index for an enumerative classification of ANIMALS.

Problems with enumerative classification. Although enumerative classification is a straightforward procedure, it does pose some problems. Because the classification includes complex subjects, such as RESPIRATION OF LAND VERTEBRATES, the simple subjects of which they are composed are repeated every time they appear in a complex subject. In the example shown in figure 2-6, the concept VERTEBRATES appears in the classification in seven different places: once as a simple subject and six times as part of a complex subject. This type of repetition makes the schedule long and complex, which in turn increases the probability that a subject will be accidentally omitted from the classification. One final problem is that only subjects which are actually listed in the schedule can be classified. Since new subjects cannot be accommodated, regular revision of the classification may be required.

Faceted classification

In an effort to remedy the problems inherent in enumerative classification, librarians developed a system known as *faceted classification*. In contrast to enumerative classification, which attempts to list every possible subject, whether simple or complex, faceted classification lists only single concepts and allows them to be combined to form complex subjects. An example of a faceted classification system is the London Classification of Business Studies. Developed in London, England, in 1965, this classification has established an international reputation and is used in a number of business libraries throughout the world.

When developing a faceted classification, the librarian must first analyze all subjects (i.e. all their documentation) to determine what individual concepts are contained in each subject. Table 2.5 shows a possible list of subjects on the left, and a list of the concepts that comprise these subjects on the right.[38]

Subjects	Concepts
The Respiration of Land Vertebrates	respiration
The Reproduction of Invertebrates	land
Water Vertebrates	vertebrate
	reproduction
	invertebrate
	water

Table 2.5. Subjects analyzed into their constituent concepts.

Once the subjects have been analyzed into their constituent concepts, the librarian must then group the concepts into *facets*[39] on the basis of common characteristics. Table 2.6 illustrates how concepts are grouped into facets according to shared characteristics.

[38] The examples of faceted classification shown in tables 2.5 – 2.10 were developed by the author of this thesis following methods outlined in the library science literature.

[39] Facets are the equivalent of what we have been calling dimensions. The term *facet* was introduced by S. R. Ranganathan, a prominent classification theorist, who pointed out that each class represented a facet or aspect of a subject.

Facet 1: physiological process	**Facet 2: animals by habitat**	**Facet 3: animals by scientific classification**
· reproduction	· land animals	· invertebrates
· respiration	· water animals	· vertebrates

Table 2.6. Concepts grouped into facets on the basis of shared characteristics.

After the concepts have been grouped into facets, the librarian must then determine a citation order. To form a complex subject from single concepts, these concepts must be combined. The order in which concepts from the different facets are combined is called the *citation order*. The citation order is extremely important: it causes documents on subjects contained in the first or *primary facet* to be grouped together, while documents on subjects contained in the other facets are scattered. The principal factor in determining the citation order should be user need. As we saw in the section on enumerative classification, a general library would most likely keep documents on animals together, thus **animals by scientific classification** would have to be the primary facet, whereas a library for zoologists specializing in animal physiology would probably group together documents on each process, in which case, **physiological process** would become the primary facet. Table 2.7 shows a citation order in which documents on animals of the same scientific classification would be kept together.

1	☞	2	☞	3
animals by scientific classification		animals by habitat		physiological process

Table 2.7. Citation order with ANIMALS BY SCIENTIFIC CLASSIFICATION as the primary facet.

Once a citation order has been determined, each concept within each facet is assigned a notation. As we discussed in section 2.3.1, the notation is a code that will provide an address for the document. Therefore, the notations should have a logical order, such as that of numbers or letters, which will permit users to find their way around the filing arrangement easily. Table 2.8 illustrates how notations are assigned to each concept within each facet.[40]

Facet 1: animals by scientific classification	**Facet 2: animals by habitat**	**Facet 3: physiological process**
1 invertebrates	1 land animals	1 reproduction
2 vertebrates	2 water animals	2 respiration

Table 2.8. Concepts grouped into facets with numerical notations.

Notations for simple concepts can then be combined according to the previously determined citation order to construct complex subjects. It is important to note that if a subject does not contain a concept from each facet, a different notation, such as a 0 (zero), must be used as a place marker so the citation order is not disrupted. Table 2.9 illustrates how complex subjects can be formed by combining simple subjects.

RESPIRATION OF LAND INVERTEBRATES =	1 (from facet 1) + 1 (from facet 2) + 2 (from facet 3) = **112**
REPRODUCTION OF VERTEBRATES =	2 (from facet 1) + 0 (no concept from facet 2) + 1 (from facet 3) = **201**

Table 2.9. Facet notations combined to represent complex subjects.

[40] The same notations can be used within each facet because the position of the notation in the citation order determines which concept is represented by each notation.

Finally, the librarian must compile an alphabetical index of the subjects within the classification so that users can easily find the corresponding code and thus ascertain the location of a particular document. According to Hunter (1979: 27), the index for a faceted scheme "contains simple subjects only, and gives the class numbers for single terms..., but not for compounds of these...." Table 2.10 shows an alphabetical index for our sample faceted classification.

INVERTEBRATE	100
LAND (ANIMAL)	010
REPRODUCTION	001
RESPIRATION	002
VERTEBRATE	200
WATER (ANIMAL)	020

Table 2.10. An alphabetical index for a faceted classification of ANIMALS.

2.3.2 How do library scientists deal with multidimensionality?

Unlike biologists, library scientists do make an effort to represent multidimensionality. It is done with a cataloguing system, which allows the multiple dimensions to be represented through cross-referencing. Shera (1976: 70) makes some interesting comments about multidimensionality in library science. Shera first observes that everyone approaches an array of books in a different way and with a different purpose, adding that "books are not a single unit of knowledge... books are multidimensional." Shera goes on to state that "the only way we have been able as yet to introduce additional dimensions into the book stock is through the use of entries in a card catalogue or bibliography, which can be multiplied as dictated by the intellectual dimensions of the book."

For example, a document dealing with the complex subject *Teaching Child Psychology* can be filed in only one place, even though it deals with three simple subjects. The physical location of the document would be determined by the user needs of the library. It might, for instance, be filed under the notation for PSYCHOLOGY, but it could be cross-referenced in a card catalogue[41] under both CHILDREN and TEACHING. As long as library users are willing to consult the catalogue before searching the shelves, they will be able to locate information dealing with any dimension of a complex subject.

In enumerative classification, multidimensionality is represented through the card catalogue, and also through the repetition of concepts. In the classification, only one characteristic is used at each level to divide classes into subclasses. Each of these subclasses is then individually subclassified on the basis of a second characteristic. As we saw in figure 2–6, this results in repetition of concepts at the lower levels of the classification. This type of repetition is a fairly inelegant way of representing multidimensionality; it results in a complex and cumbersome classification. We feel that it is more logical to represent multidimensionality by introducing the different differentiating characteristics at the same level, but *in different dimensions*.[42] Figure 2–7 shows a modified version of figure 2–6 (cf. section 2.3.1) in which unnecessary repetition has been eliminated by introducing the characteristics in different dimensions.

Faceted classification also represents multidimensionality both through the card catalogue and in the actual classification. Different ways of classifying a subject can be represented in different facets. For example, in table 2.6, animals were classified into two different facets on the basis of two different characteristics: 1) according to **habitat**, and 2)

[41] In many libraries today, the computerized database has replaced the card catalogue.

[42] This process is explained more thoroughly in Chapter 3 as part of an extended example.

according to **scientific classification**. These two different ways of classifying animals represent two different dimensions based on two different characteristics.

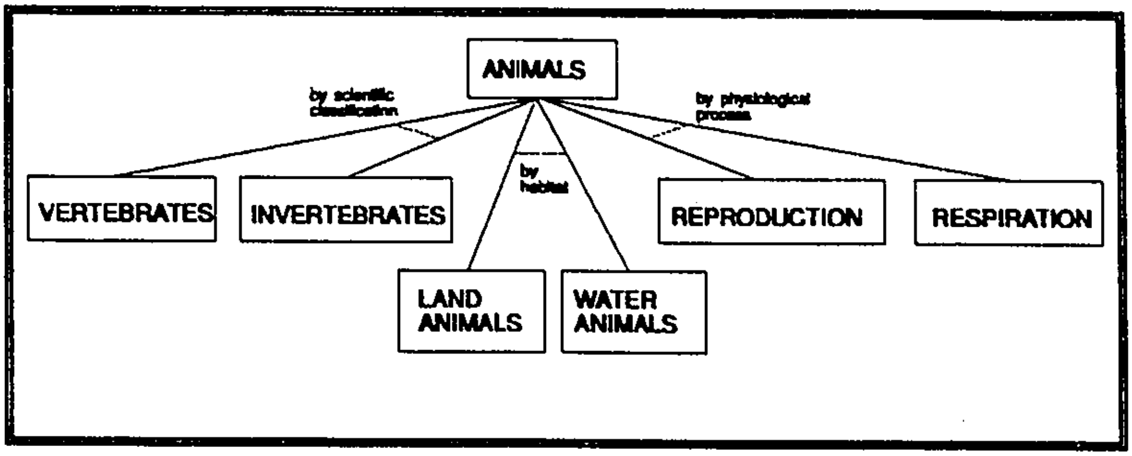


Figure 2–7. The elimination of repetition through the use of multiple dimensions.

2.3.3 What can terminologists learn from library science classification methods?

Library scientists and terminologists take the same general approach to classification: both work hierarchically, and both analyze characteristics very carefully to determine the relations between concepts.

Library scientists have recognized the usefulness of systematically representing these concepts in such a way that the relations between them are explicit; they do not use tree-structure diagrams *per se*, but tend to draw up a schedule in the form of a notated indented list (cf. table 2.3), which is complemented by an alphabetical index (cf. table 2.4). Terminologists, although they are equally aware of the importance of the relations between concepts, do not tend to explicitly show these relations in their term banks or terminological publications. As mentioned in section 2.2.2, they usually list terms alphabetically, along with definitions and equivalents, thus allowing users to have quick, easy access to any given term. Yet it might be even more helpful to users if terminologists followed the methods of library scientists and listed the terms in a notated indented list with a complementary alphabetical index.[43] After all, when terminologists begin learning about a new subject field, they themselves often sketch out the relationships between concepts in the field, so there is little doubt that users of the terminological publications would also benefit from explicitly seeing the relationships between the concepts. Such indented lists may even be cheaper and more feasible than tree-structure diagrams for representing hierarchical relations because they require less space in a terminological publication, and are technically easier to produce since they are text-based and do not involve graphics.

Although terminologists are aware that many concepts can be classified on the basis of more than one characteristic, they do not generally represent multidimensionality in their term banks or terminological publications since they tend to present terms alphabetically. In contrast, library scientists do attempt to represent multidimensionality in their classifications. In enumerative classification, multidimensionality is expressed through the repetition of concepts, while in faceted classification, different ways of classifying a concept are represented in different facets. The method of repeating concepts to express multidimensionality is inelegant and results in a confusing and cumbersome classification; this is one of the reasons that the system of faceted classification was developed. Terminologists should be aware that if, in the course of their reading, they encounter a classification containing such duplication, it may indicate the presence of multidimensionality (cf. figures 2-6 and 2-7). A better way to represent multidimensionality could be achieved by adopting the approach used in faceted classification. Terminologists could introduce the different ways of classifying a concept in different dimensions, similar to the way in which library scientists use different facets. In fact,

[43] Some terminological publications, such as the British Computer Society's *Glossary of Computing Terms* (1989), do present information in a *systematic arrangement*. The ISO (ISO 1087: 11) defines *systematic arrangement* as an "arrangement of entries in an order based on a system of concepts." According to Picht and Draskau (1985: 33) and Sager (1990: 36), this method of representing terms systematically is starting to become more popular; however, at present, the format for the majority of terminological publications remains purely alphabetical.

Sager (1990: 36) briefly discusses faceted classification and observes "There is also a growing tendency to use such divisions [facets] in systematically structured glossaries where terms are grouped together because they share a characteristic of division of the broader genus."

Terminologists should also take note of the use of notations in library science. Although library scientists use notations as codes for filing and retrieving documents, terminologists could use them for identifying dimensions, concepts or characteristics. This will be discussed in more detail in section 3.2.2.

Finally, library scientists take the trouble to rank the characteristics used to make the classifications in order of importance, which is based on the needs of their clients. A similar type of ranking might be useful for terminologists. Since multidimensionality can be confusing, we feel that terminologists would benefit by working on one dimension at a time; therefore, it might be useful to rank the dimensions and develop them in order of importance. The dimensions can be ranked according to the frequency of occurrence in the documentation, and possibly the number of terms in the dimensions. It would also prove helpful to consult with subject-field experts when ranking the dimensions, as they may be able to provide some valuable insights as to the relative importance of the dimensions in question.

In our own experience, we have found that some terms do appear less frequently in the documentation, such as the terms *cold typesetting* (cf. section 1.2.1) and *replicable disc* (cf. section 1.2.1), and they have been subsequently placed in lower-ranking dimensions.

Concluding remarks for chapter 2. The insights culled from the investigation of classification methods used in the disciplines of terminology, biology, and library science provide a good starting point for the development of a set of guidelines for handling multidimensionality. In the next chapter, we will complement this research with some practical experimentation.

CHAPTER THREE

AN EXTENDED EXAMPLE DEMONSTRATING MULTIDIMENSIONALITY

3.0 MULTIDIMENSIONALITY IN SOURCE DOCUMENTATION

Now that we have discussed classification in general, and the specific classification problem of multidimensionality, we are going to look at an extended example of a multidimensional representation. This example derives from various sources which were gathered as part of the COGNITERM project (cf. section 5.1) to establish a terminology of optical storage technology.[44] We believe these sources, and the type of information presented in them, are representative of the types of documents that are used by most terminologists conducting terminological research projects. Hence, the types of problems contained in these sources are typical of the types of problems with which terminologists are generally faced.

For the extended example, we will begin by examining three different ways of classifying OPTICAL STORAGE MEDIA. We will then identify problems that arise in the various classifications, and propose possible solutions to the problems.

3.1 THREE CLASSIFICATIONS OF OPTICAL STORAGE MEDIA

Classification 1

The first classification we will examine is taken from Chen (1989b: 5), and is depicted in figure 3-1. Chen has classified OPTICAL STORAGE MEDIA on the basis of the characteristic **degree of writability** into READ-ONLY MEDIA, WRITE-ONCE MEDIA, and ERASABLE MEDIA. This classification was obviously made from the point of view of a user (cf. section 1.2.1); it is based on whether a user can write to the media many times

[44] A more detailed discussion of the findings related to the optical storage technology project can be found in Meyer, Bowker and Eck 1992 a and b; Meyer, Skuce, Bowker and Eck 1992 a and b.

(ERASABLE MEDIA), only once (WRITE-ONCE MEDIA), or not at all (READ-ONLY MEDIA).

At the second level of the classification, READ-ONLY MEDIA, WRITE-ONCE MEDIA, and ERASABLE MEDIA have each been subclassified principally[45] according to the characteristic **physical form**. This classification was based on whether the media took the form of a disc, a card, film, tape, or paper.

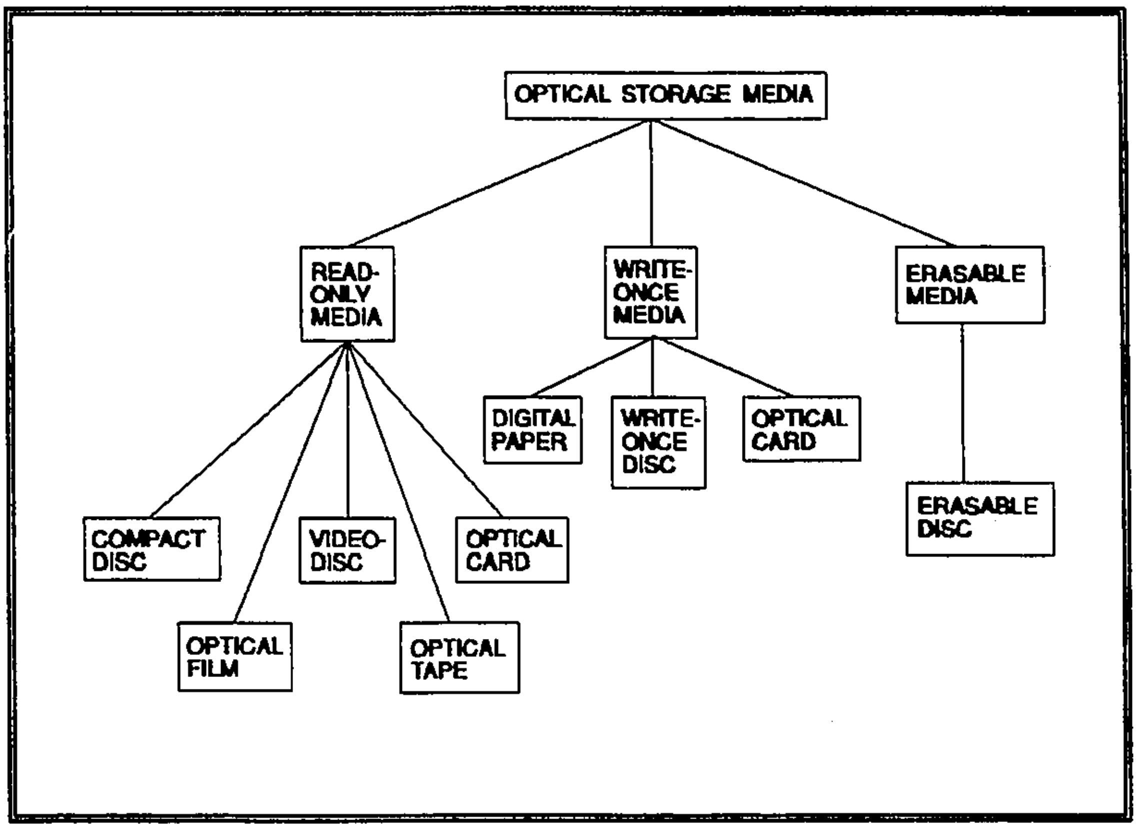


Figure 3-1. OPTICAL STORAGE MEDIA: Classification 1. *Source: Chen 1989b: 5.*

[45] We say *principally* because, as we will see in section 3.2.1 (*Problem 1*), Chen has erroneously classified some concepts on the basis of two characteristics at this level.

Classification 2

Compare Chen's classification with the next example, shown in figure 3–2, which is taken from Heimburger (1990: 32). We can see that this classification is different from Chen's because OPTICAL STORAGE MEDIA has been classified principally[46] according to the characteristic **physical form**. As we saw in figure 3–1, this characteristic was not used to differentiate between different types of media until the second level of Chen's classification.

Another observation we can make about Heimburger's classification is that the discs have been classified from a manufacturer's point of view (cf. section 1.2.1 (*Viewpoint*)), hence the differentiating characteristic is **replicability**. A REPLICABLE DISC is one that is produced by a manufacturer with all the necessary information already stored on it, whereas an INDIVIDUAL DISC is produced as a blank disc and the user is responsible for storing information on it.

At the second level of the classification, we see that Heimburger has classified REPLICABLE DISC according to the characteristic **size**,[47] while INDIVIDUAL DISC has been classified according to the characteristic **degree of writability**. Meanwhile, OPTICAL CARD and OPTICAL TAPE have not been classified any further.

Classification 3

The final example is actually a classification of OPTICAL DISC, which is a subclass of OPTICAL STORAGE MEDIA. This example is taken from Saffady (1988: x), and is shown in figure 3–3. Saffady has classified OPTICAL DISC on the basis of the characteristic **degree of writability** into READ-ONLY DISC, WRITE-ONCE DISC, and ERASABLE DISC. Saffady then proceeded to classify READ-ONLY DISC by **size**, while WRITE-ONCE DISC and ERASABLE DISC were not classified any further.

[46] Once again, we say *principally* because, as we will see in section 3.2.1 (*Problem 1*), Heimburger has erroneously classified some concepts on the basis of two characteristics at this level.

[47] A compact disc is 4.72 inches in diameter, while a videodisc is generally 12 inches in diameter.

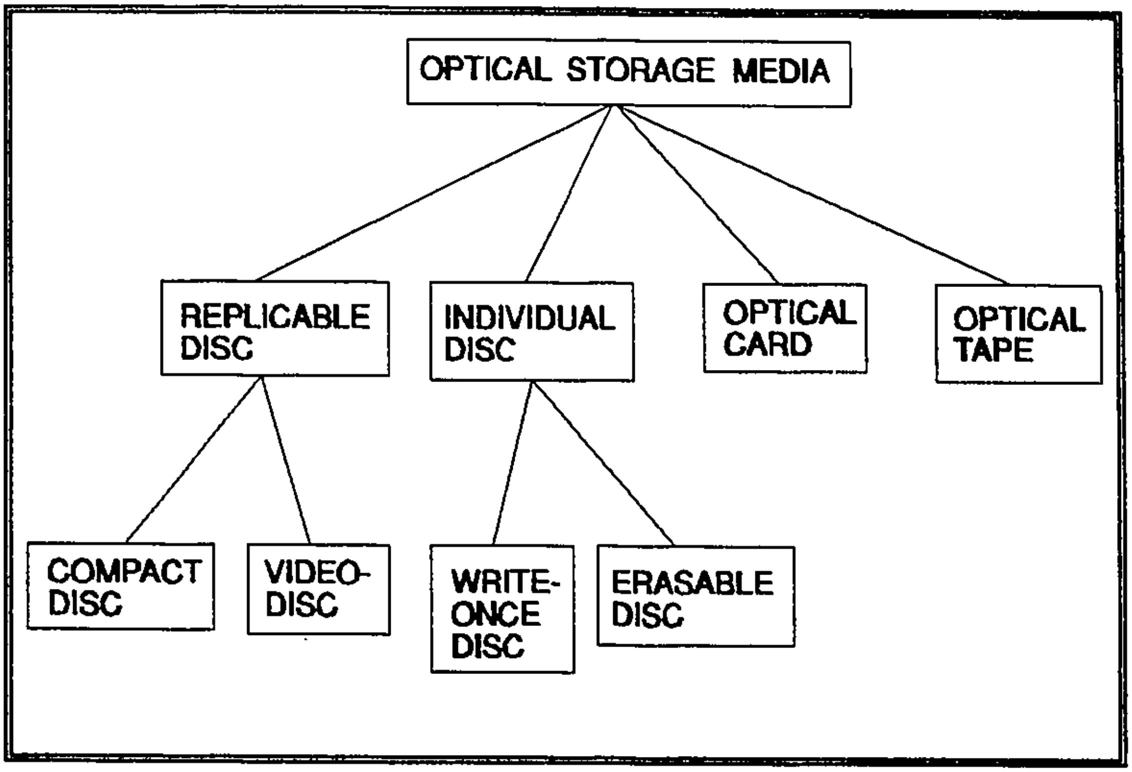


Figure 3–2. OPTICAL STORAGE MEDIA: Classification 2. *Source: Heimburger 1990: 32.*

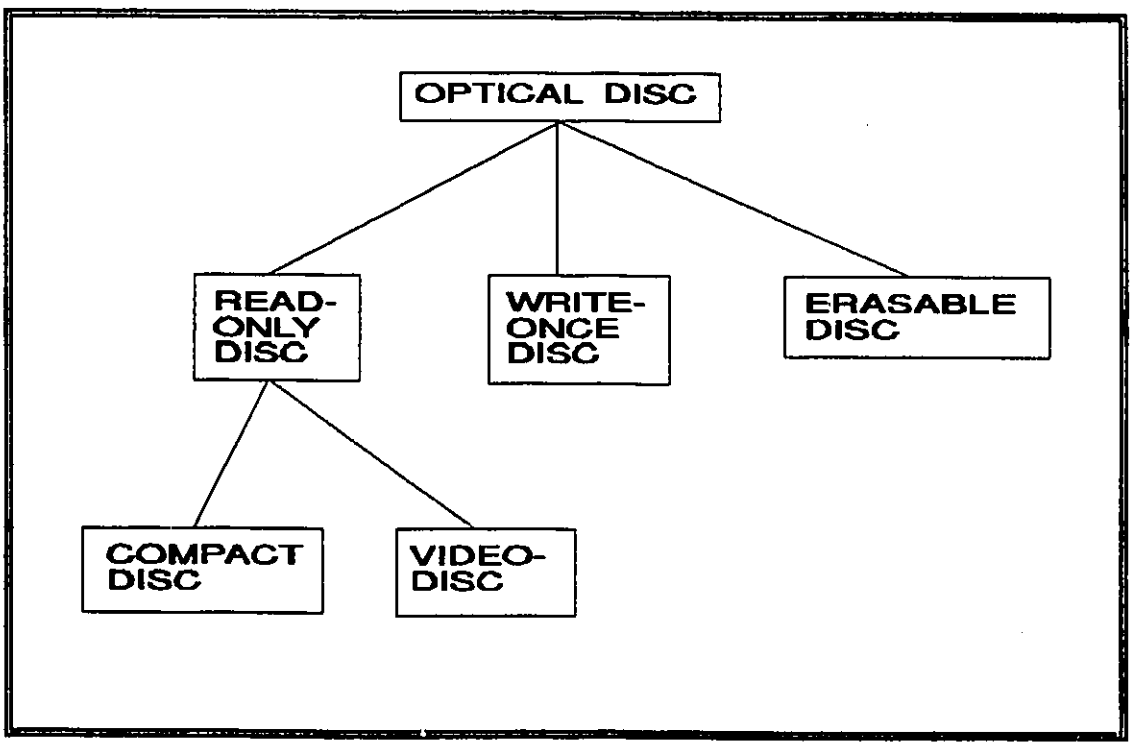


Figure 3–3. OPTICAL DISC: Classification 3. *Source: Saffady 1988: x.*

3.2 PROBLEMS AND PROPOSED SOLUTIONS IN THE CLASSIFICATIONS OF OPTICAL STORAGE MEDIA

In the three preceding classifications of OPTICAL STORAGE MEDIA, we can identify two main categories of problems: classification problems and notation problems. Each of the problems will be discussed in turn, and possible solutions will be presented.

3.2.1 Classification problems

Two types of classification problems arise in the three examples studied. The first is the occurrence of a phenomenon which we have termed *false multidimensionality*, and the second is the unidimensional nature of all the representations.

Problem 1: False multidimensionality

The first classification problem is one which we have termed *false multidimensionality*. Looking once again at Chen's example in figure 3-1, we can determine, although it is not explicitly indicated, that the class READ-ONLY MEDIA has actually been subclassified on the basis of two different characteristics: 1) **physical form** and 2) **size**. In other words, the subclasses COMPACT DISC and VIDEODISC have the value ***disc*** for the characteristic **physical form**, and are therefore different from the other forms of media with the values ***tape***, ***film***, and ***card***. However, in addition to differing from other types of media by their **physical form**, COMPACT DISC and VIDEODISC also differ from each other on the basis of the characteristic **size**: compact discs are 4.72 inches in diameter, while videodiscs are 12 inches in diameter.

This type of simultaneous classification, or *false multidimensionality*, is undesirable for several reasons. On a very simple level, it causes conceptual confusion — it goes against the basic principle of classification, i.e. that a concept should only be classified according to one characteristic at any one level.[48] It is worth noting that none of the classifications seen in biology or library science showed evidence of such simultaneous classification.

A second and more serious consequence of false multidimensionality is that a concept may be omitted, in this case, the concept OPTICAL DISC. When a concept is classified by two characteristics simultaneously, only the final result of the classification is shown in the representation. Chen showed the place of discs in the classification only after they had been classified on the basis of both **size** and **degree of writability**. As a result, the concepts COMPACT DISC and VIDEODISC appeared in the classification, but their generic concept, OPTICAL DISC, was not explicitly reflected. In terminology work, the omission of a concept is very serious because it results in the omission of a term.

A third reason why false multidimensionality should be avoided is that it could be confused with true multidimensionality. It is important to understand the difference between multidimensionality and false multidimensionality. We established earlier (cf. section 1.2) that multidimensionality is a phenomenon which occurs when a concept can be classified according to more than one characteristic at the same level; however, what we see in Chen's classification is *not* multidimensionality, but rather a fundamental error in classification. In a multidimensional representation, a concept can be classified on the basis of more than one characteristic at the same level, *but not within the same dimension*. In other words, all the concepts at a given level in dimension *A* are classified on the basis of the same single characteristic, while all the concepts at this same level in dimension *B* are classified on the basis of another single characteristic. Each dimension can be viewed as a complete hierarchical classification that reflects one particular way of classifying the entire subject field.

---

[48] In a multidimensional representation, a concept can be classified according to more than one characteristic at the same level, *but not within the same dimension*.

Chen's classification does *not* show multidimensionality. Chen has made an error in classification which has resulted in the phenomenon false multidimensionality. False multidimensionality occurs when a concept is classified at one level on the basis of two characteristics that should be applied within the same dimension, but at two *different* levels.

A second occurrence of false multidimensionality can be seen in Heimburger's classification (cf. figure 3–2). Here, the discs have been classified simultaneously according to **physical form** and **replicability**, instead of at two different levels. Once again, this has resulted in the omission of the concept OPTICAL DISC.

Solution. False multidimensionality can be avoided by classifying concepts on the basis of only one characteristic at each level of the classification; in this way, there is no conceptual confusion, and no missing concept. Both Chen and Heimburger classified discs on the basis of two characteristics simultaneously, and their diagrams reflected the classification of the discs only after both characteristics had been applied. What they should have done was classify the concepts based on one characteristic at a time, and represent the concepts produced at each step of the classification process. In other words, when they were classifying media by **physical form**, Chen and Heimburger should have noted that ***disc*** is the value of the characteristic, regardless of the **size** or the **replicability** of the disc. When OPTICAL STORAGE MEDIA is classified on the basis of **physical form**, one of the concepts produced is OPTICAL DISC. OPTICAL DISC can then be subdivided according to another characteristic, for example **size**, to produce the concepts COMPACT DISC and VIDEODISC, or according to **replicability**, to produce the concepts REPLICABLE DISC and INDIVIDUAL DISC. Figure 3–4 shows revised versions of Chen's and Heimburger's classifications where OPTICAL STORAGE MEDIA has been classified on the basis of only one characteristic at each level of the classification. Both of these classifications include the concept OPTICAL DISC, which was missing from the original Chen and Heimburger representations.

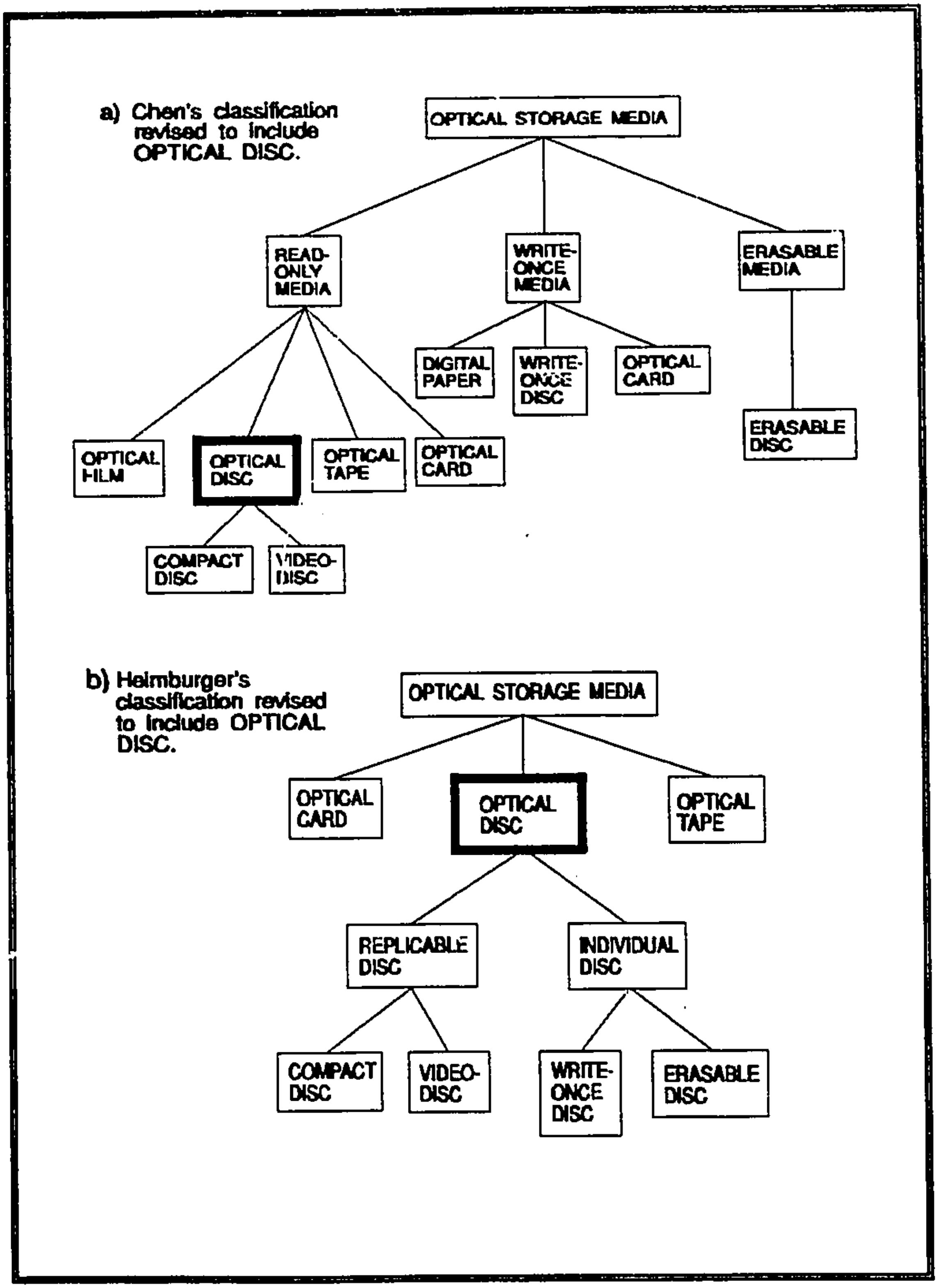


Figure 3–4. Classifications of OPTICAL STORAGE MEDIA on the basis of one characteristic per level.

Problem 2: Unidimensionality

While a unidimensional representation may suit some purposes (e.g. biological classification), it poses two problems for terminologists: 1) missing concepts (and hence missing terms), and 2) repetition. A unidimensional classification is an incomplete classification, which results in an incomplete representation of the subject field and may cause conceptual confusion or lead to incomplete definitions. Most importantly, an incomplete classification will likely have missing concepts, which in turn results in missing terms. Furthermore, a unidimensional classification may contain repetition of concepts, which can also cause confusion.

Consider the example taken from Heimburger, shown in figure 3-2. Heimburger has classified OPTICAL STORAGE MEDIA primarily on the basis of the characteristic **physical form**. In this example, the classification of OPTICAL STORAGE MEDIA on the basis of **degree of writability** was not represented, hence, the concepts READ-ONLY MEDIA, WRITE-ONCE MEDIA, and ERASABLE MEDIA were not included in the classification.

The problem of missing concepts was also present in the example taken from Chen, shown in figure 3-1. Chen has classified discs, albeit incorrectly, according to the characteristics **physical form** and **size**. The discs were not classified on the basis of **replicability**, and thus the concepts REPLICABLE DISC and INDIVIDUAL DISC are missing from the classification. A second problem, that of repetition,[49] was also present in Chen's classification. Instead of classifying OPTICAL STORAGE MEDIA on the basis of **physical form** in a separate dimension, Chen used this characteristic to individually subclassify each of the subclasses READ-ONLY MEDIA, WRITE-ONCE MEDIA and

[49] Repetition can refer to one of two things: sometimes, an actual concept is repeated (e.g. OPTICAL CARD appeared twice in Chen's classification); other times, the same characteristic is used to classify concepts in more than one part of the representation.

ERASABLE MEDIA. As a result, the concept OPTICAL CARD appeared twice in the classification, and several concepts with the value ***disc*** for the characteristic **physical form** appeared in different areas of the classification (e.g. ERASABLE DISC under ERASABLE MEDIA; WRITE-ONCE DISC under WRITE-ONCE MEDIA; and COMPACT DISC and VIDEODISC under READ-ONLY MEDIA). This type of repetition is unnecessary and can be avoided in a multidimensional representation.

Finally, Saffady's representation of OPTICAL DISC, shown in figure 3-3, is also unidimensional. Saffady has classified OPTICAL DISC according to the characteristic **degree of writability**, but has not reflected the **replicability** of the discs. Saffady has also chosen to subclassify only READ-ONLY DISC on the basis of the characteristic **size**, instead of introducing this classification in a separate dimension at the level of OPTICAL DISC, where the size of WRITE-ONCE DISC and ERASABLE DISC could also be reflected.

Solution. The classifications made by Chen, Heimburger, and Saffady were all unidimensional. In order for a representation to be multidimensional, *all* possible ways of classifying every concept must be reflected at each level of the classification. Instead of using a characteristic to individually subclassify groups of subclasses, as Chen used the characteristic **physical form** to independently subclassify the three subclasses READ-ONLY MEDIA, WRITE-ONCE MEDIA and ERASABLE MEDIA, the two characteristics, **degree of writability** and **physical form** should be introduced in separate dimensions at the same level of the classification. Figure 3-5 shows a multidimensional representation of OPTICAL STORAGE MEDIA, produced using the CODE system.

Although this solution is, in principle, very straightforward, it should be emphasized that a two-dimensional medium (i.e. paper and pencil) is not conducive to the representation of multidimensionality. This is especially true of very complex representations where multidimensionality occurs at several levels of the classification.

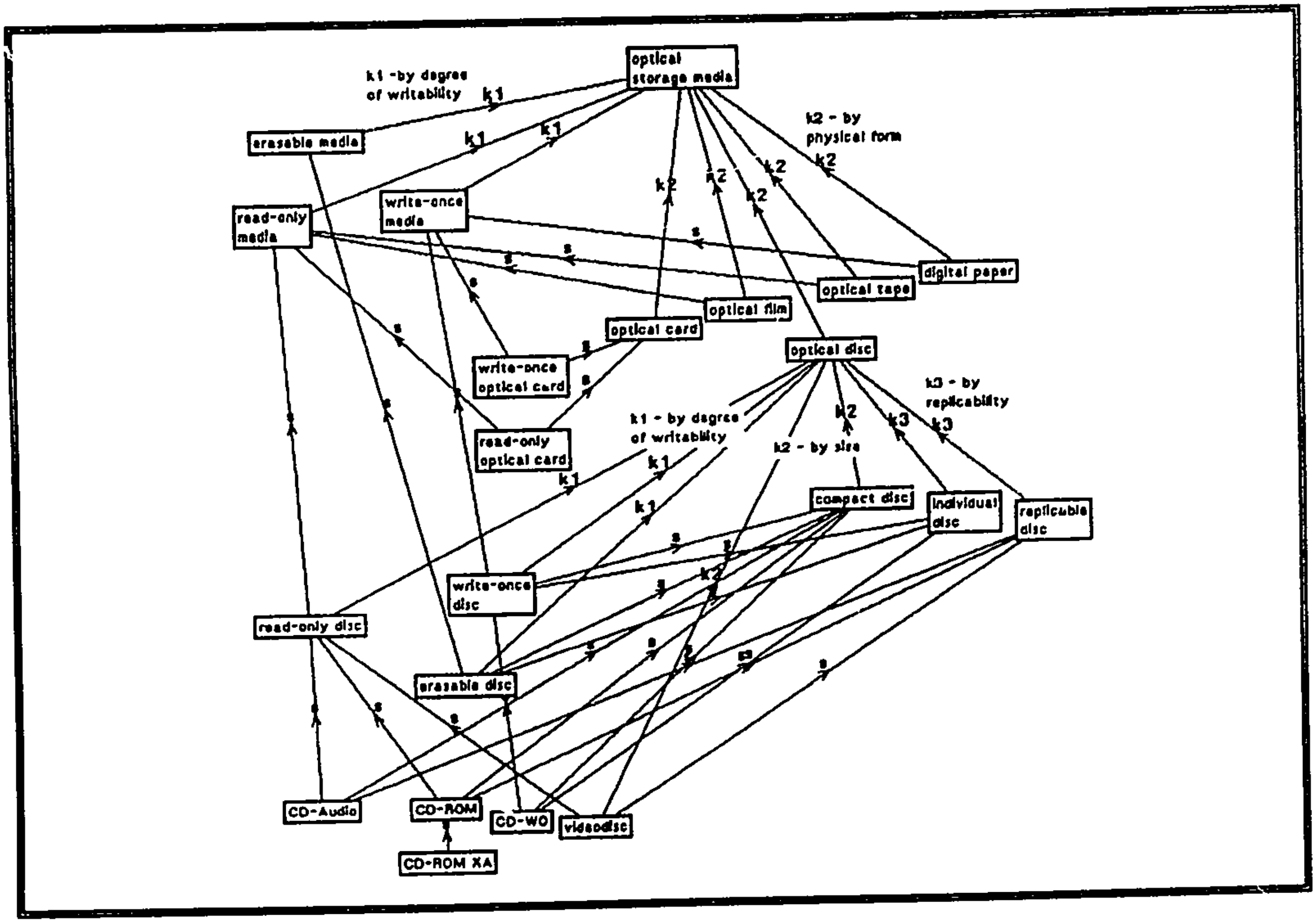


Figure 3-5. A multidimensional representation of OPTICAL STORAGE MEDIA, drawn using CODE.

3.2.2 Notation problems

The second category of problem that arises in the examples is related to notation. Two different notation problems were observed. Firstly, there was no notational differentiation made between different dimensions in the classification, and secondly, the characteristics that were used to classify the subclasses were not explicitly identified.

Problem 1: No notational difference between dimensions

When only one dimension is present in any given classification, there is no need to give it a notational marker; however, when the representation includes more than one way of classifying an object, we feel it is helpful to indicate which subclasses belong to each dimension. If the dimensions are not marked, the reader may be unaware that the subclasses at any given level have been classified according to different characteristics, and therefore, belong to different dimensions. This can lead to conceptual confusion on the part of the reader. In both Chen's (cf. figure 3-1) and Heimburger's (cf. figure 3-2) representations, there was no explicit indication that different dimensions were present in the representation.[50] In other words, the reader could not tell from the diagrams that classes were being subclassified on the basis of different differentiating characteristics.

For example, in Chen's classification (cf. figure 3-1), all the lines connecting the class READ-ONLY MEDIA to its five subconcepts look exactly the same. It is not immediately evident to the reader that OPTICAL CARD, OPTICAL TAPE and OPTICAL FILM have been classified on the basis of the characteristic **physical form**, while COMPACT DISC and VIDEODISC have been classified according to both **physical form** and **size**.

[50] Although we established in section 3.2.1 that both Chen's and Heimburger's classifications were essentially unidimensional, they both gave partial representations of different dimensions by using different characteristics to classify subclasses at the same level. However, they failed to indicate these instances with some sort of notation.

We see this same problem in Heimburger's classification (cf. figure 3–2), where no explicit distinction is made between the dimension in which OPTICAL CARD and OPTICAL TAPE have been classified according to **physical form**, and that in which INDIVIDUAL DISC and REPLICABLE DISC have been classified on the basis of both **physical form** and **replicability**.

Solution. There are several ways in which one could indicate that the classifications are being based on different characteristics. For example, as shown in figure 3–6, Finnish terminologists at the *Tekniikan Sanastokeskus* (the centre for technical terminology in Helsinki) distinguish between different dimensions of a classification by using thick lines, while the subclasses within a dimension are joined to the generic class with thinner lines. In figure 3–6, each of the thick lines represents a different dimension, which is obviously based on a different differentiating characteristic; however, these characteristics are not explicitly named on the diagram (Telesanasto Teleordlista 1991: 70).

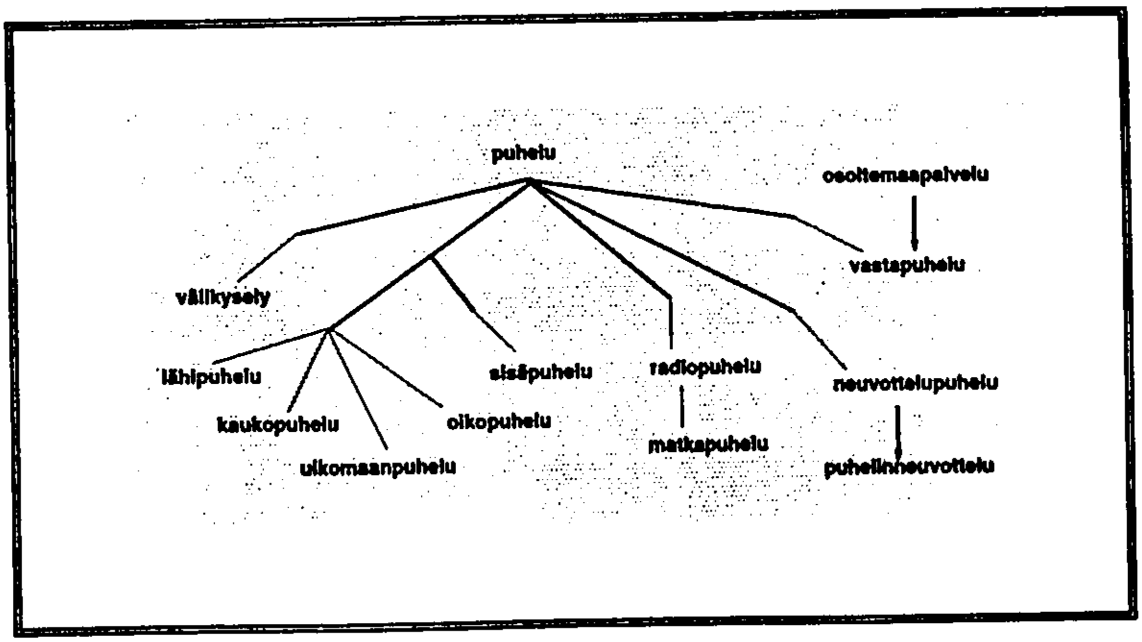


Figure 3–6. A variation in line thickness used to show different dimensions. *Source: Telesanasto Teleordlista 1991: 70.*

Another way to demonstrate that classifications have been made on the basis of different characteristics is to assign a different notational symbol to each dimension. This is the method used in the current version of the COGNITERM term bank[51] (COGNITERM 1992). Each subclass is marked with an alphanumeric notation;[52] subclasses with the same notation belong to the same dimension. In figure 3–7, the notation k1[53] represents the classification of OPTICAL STORAGE MEDIA on the basis of the characteristic **degree of writability**, and k2 represents the dimension classified according to **physical form**.

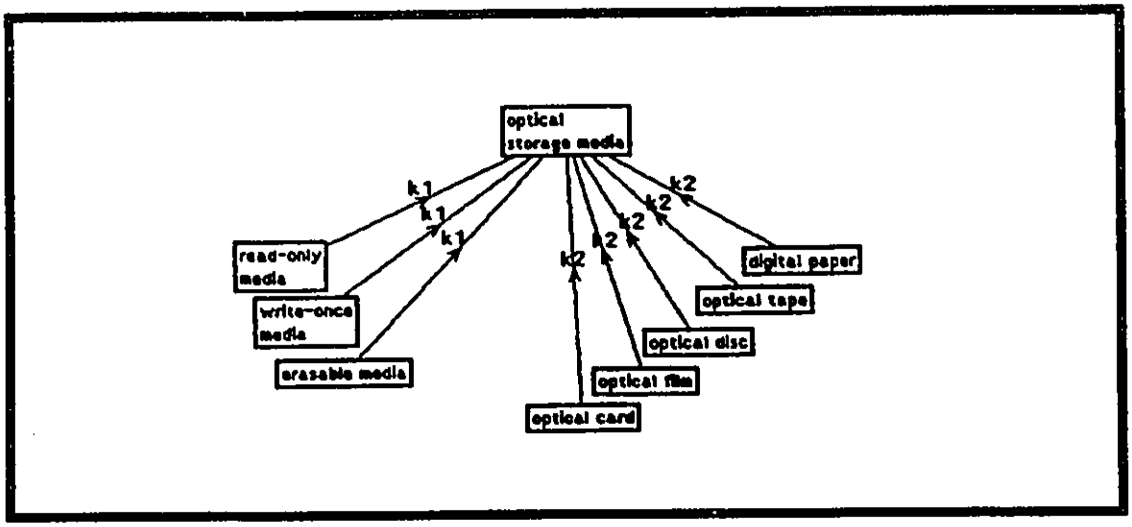


Figure 3–7. Notations used to show different dimensions. *Source: COGNITERM 1992.*

Yet another method that could be employed to visually indicate the different dimensions would be the use of different coloured lines. In figure 3–8, red lines represent the classification of OPTICAL STORAGE MEDIA based on the characteristic **degree of writability**, while black lines represent the dimension classified according to **physical form**.

[51] This method may change in the future, perhaps based on the findings of this thesis.

[52] The notation actually appears beside the line connecting the subclass to the generic class.

[53] The *k* stands for *kind*. Kinds are a feature of CODE, the system that was used to build the COGNITERM term bank. Kinds will be discussed in more detail in section 5.1.3 (*Kinds*).

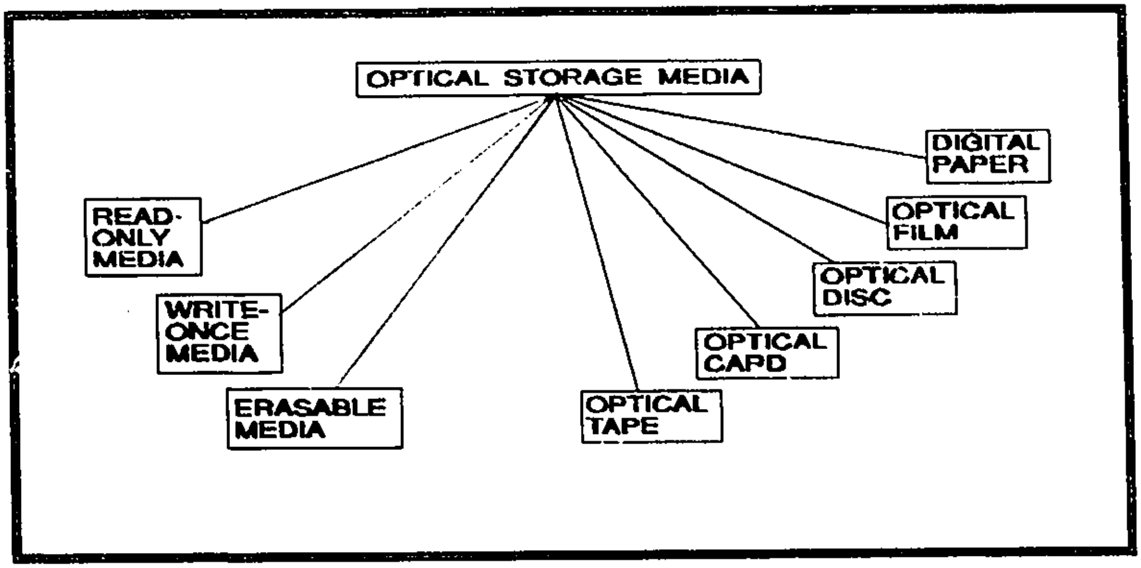


Figure 3-8. Different coloured lines used to show different dimensions.

Problem 2: Differentiating characteristics not identified

The second notation problem observed was that the characteristic which was used to classify the subclasses was not explicitly identified. Picht and Draskau (1985: 64) comment that "The types of relations and the classifying criteria must be clear and transparent. [...] Fuzzy information will lead to misinterpretations and render the value of the system suspect."

None of the three examples observed (Chen, cf. figure 3-1; Heimburger, cf. figure 3-2; Saffady, cf. figure 3-3) explicitly indicated which characteristic had been used to make the classification. While some of the characteristics are fairly transparent, such as Chen's classification of OPTICAL STORAGE MEDIA into OPTICAL DISC, OPTICAL CARD, OPTICAL FILM, OPTICAL TAPE, and DIGITAL PAPER on the basis of the characteristic **physical form**, other characteristics may be harder to identify. For example, it may not be immediately obvious to the reader that COMPACT DISC and VIDEODISC have been classified according to the characteristic **size**. Even when a concept can only be classified on the basis of one characteristic, we feel that it would be helpful to explicitly identify this characteristic for the reader.

We would like to emphasize that assigning names to characteristics is not always a simple matter, and a number of general guidelines to facilitate this process have therefore been established within the framework of the COGNITERM Project.[54] For instance, whenever possible, a characteristic's name should be an established term. If there is no generally accepted term to describe the characteristic, terminologists should try to choose a name that is relatively short, yet transparent and meaningful. If the chosen term is not completely clear, an explanation of the term should also be included in a place that is easily accessible to the user. For example, to the best of our knowledge, the term *degree of writability* does not appear in the documentation. Before coining this term, the terminologists had to consider very carefully exactly which criteria had been used to classify the concepts. It was decided that the classification had been based on whether or not the user could write to the media, and if so, how many times. Based on this description, the relatively short characteristic name **degree of writability** was coined, and it was accompanied by the following explanation: Degree of writability refers to whether the user can write to the media many times (erasable media), only once (write-once media), or not at all (read-only media). The appropriateness of characteristic names, whether selected from the documentation or coined by terminologists, should be verified by subject-field experts. The characteristic name, **degree of writability**, for example, was indeed judged appropriate by a subject-field expert.[55]

Solution. There are several possible ways to identify characteristics on a diagram. The first and most straightforward way, shown in figure 3-9, is to use a line to link those concepts classified according to the same characteristics, and then actually write the name of the characteristic beside them on the diagram. This works well if the diagram is large and has enough space in which to write the characteristic name.

---

[54] For more details on guidelines for naming characteristics, please refer to Meyer, Skuce, Bowker and Eck 1992b.

[55] The subject-field expert was Mr. Don Slaunwhite, a software developer in the CD-ROM division of Corel Systems Corp.

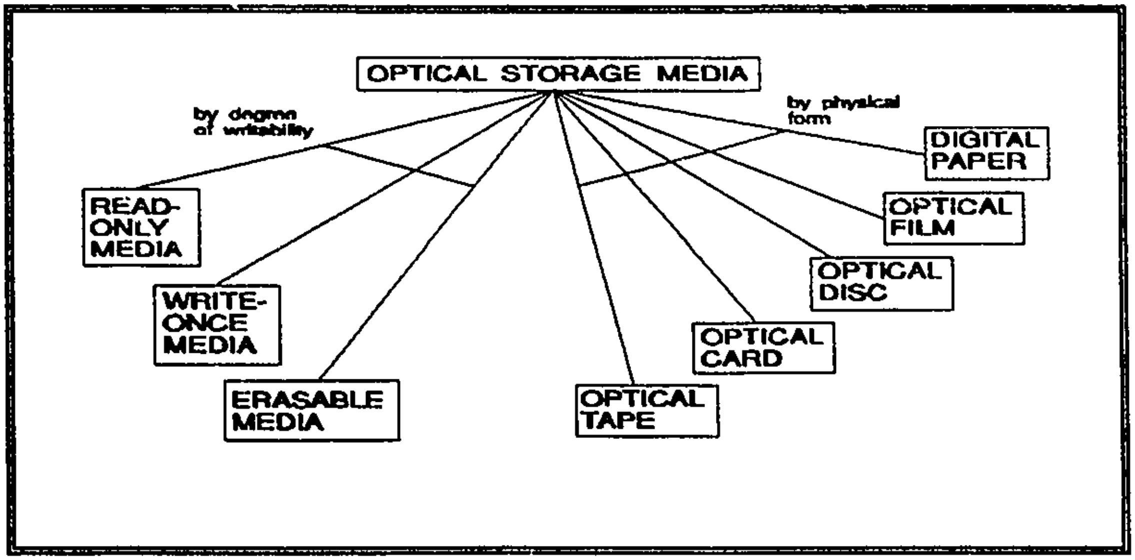


Figure 3-9. Representation indicating characteristic names.

On more intricate diagrams where space is limited, it may be more logical to insert a short notational symbol, such as a number, on the diagram itself, and then include a legend to indicate which characteristic is represented by each notation. An example of a representation combining notations with a legend is shown in figure 3-10.

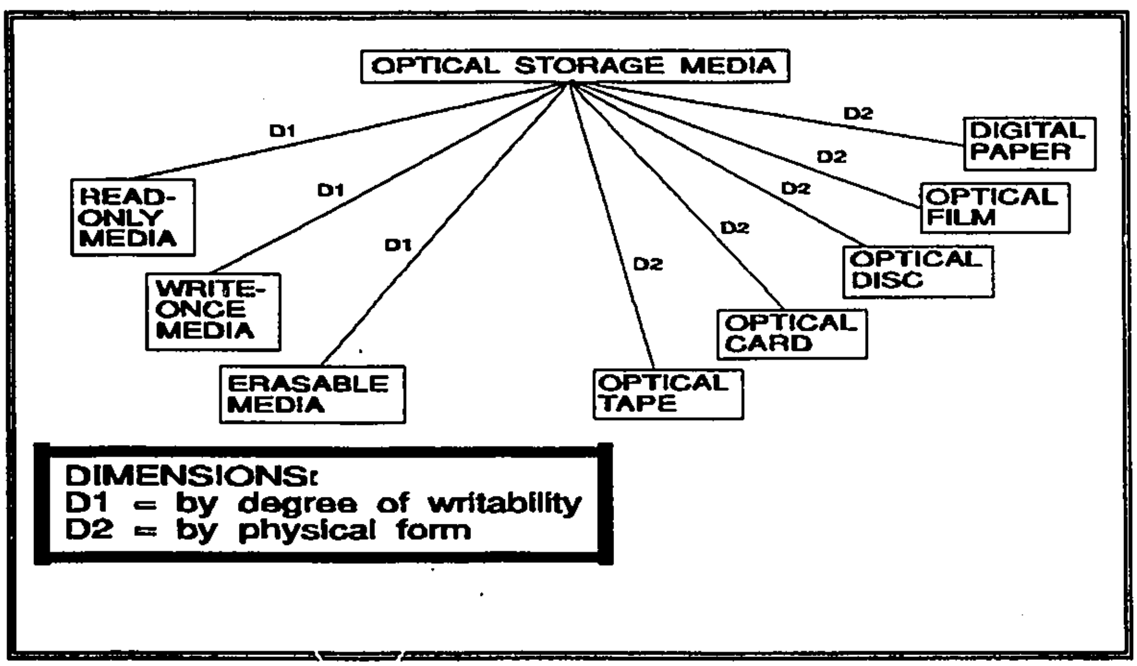


Figure 3-10. Representation with notations and a legend to indicate characteristics.

A third possibility is available to people working with computer systems that have hypertext capabilities. With hypertext, it is possible to have information stored that is hidden from the user's view until he or she requests the information. In the case of a classification representation, one could envisage the user simply positioning the cursor over the line connecting a given subclass to its generic class, and then clicking a button on the mouse. The characteristic name would then "pop up" on the screen in a text box. The user could read the information in the text box and then hide it again by clicking another button.

No matter which method is used, the essential point is that the representations would be more helpful to the user if the characteristic names were easily accessible, without the user having to refer to text not included on the diagram.

In summary, this example has served to identify some of the typical problems faced by terminologists in the documentation they use to conduct their terminological research. The problems fall into two general categories: classification problems and notation problems. It is evident from the example that some of these problems, and specifically that of representing multidimensionality, are extremely difficult to deal with using traditional paper-and-pencil techniques.

CHAPTER FOUR

INITIAL GUIDELINES FOR RECOGNIZING AND REPRESENTING MULTIDIMENSIONALITY

4.0 SOURCES FOR THE INITIAL GUIDELINES

In chapter 2, we examined the role of classification and multidimensionality in the disciplines of terminology, biology, and library science. In chapter 3, we presented an extended example of a multidimensional representation based on practical experimentation with a TKB on optical storage media. Based on the research carried out for these two chapters, we have been able to cull a variety of useful insights that we will now incorporate into an initial set of guidelines to help terminologists both *recognize* and *represent* multidimensionality.

4.1 RECOGNIZING MULTIDIMENSIONALITY

It has been previously established that many subject fields can be classified in more than one way for a variety of reasons, such as different viewpoints or opinions (cf. section 1.2.1). This being the case, terminologists should be able to recognize indications of multidimensionality, both in different sources and within one classification. Terminologists must also be able to identify and correct false multidimensionality.

The benefits of working with graphic representations of the concept systems were emphasized in the terminology, biology, and library science literature (cf. sections 2.1.1, 2.2.3, and 2.3.3). Indeed, working with a graphic representation will make it easier for terminologists to recognize multidimensionality. Sometimes, such representations can be found in the documentation; however, if no adequate representations are readily available, terminologists may find it helpful to construct their own based on the information they find in the documentation.

4.1.1 Recognizing multidimensionality in different sources

Although a given author may deal with only one possible classification of a subject field, terminologists should be able to detect indications of multidimensionality in different sources. These indications may be found in tree diagrams, tables of contents, lexical items, and by consulting with subject-field experts.

Tree diagrams

Different authors may each present a different classification, usually shown in the form of a tree diagram, of the same subject field. These classifications will be based on different differentiating characteristics (cf. sections 1.1.1 and 1.2), and may contain different concepts. For example, in section 3.1, we saw that Chen provided a tree diagram showing that OPTICAL STORAGE MEDIA had been classified according to the characteristic **degree of writability** of the media into READ-ONLY MEDIA, WRITE-ONCE MEDIA, and ERASABLE MEDIA (cf. figure 3-1). Meanwhile, Heimburger's tree diagram showed a classification of the same concept, OPTICAL STORAGE MEDIA, but this classification was based on the characteristic **physical form** of the media, and produced the subconcepts REPLICABLE DISC, INDIVIDUAL DISC, OPTICAL CARD, and OPTICAL TAPE (cf. figure 3-2).

Tables of contents

Different classifications can easily be identified from tree diagrams, such as those presented by Chen and Heimburger (cf. figures 3-1 and 3-2), but they can also be identified from divisions indicated in tables of contents. For example, in section 1.2.1 (*Opinion*), we saw that the tables of contents in two different cookbooks indicated two different classifications of recipes. In *Basics & Beyond: An Adventure in Cooking* (MacDonald and

O'Connell 1986), APPETIZERS AND BEVERAGES constitute one chapter, while SOUPS AND SAUCES make up another. Meanwhile, in *Good Housekeeping's Cookery Book* (1967), BEVERAGES alone comprise one chapter, as do SAUCES, while APPETIZERS AND SOUPS together form a third chapter.

Lexical Items

The vocabulary used to refer to concepts within different documents can also indicate that different classifications have been made, particularly when authors refer to different sets of contrasting lexical pairs. For instance, when discussing typesetting, one author may contrast *hot* with *cold typesetting*, while another author may contrast *phototypesetting* with *display typesetting*. Even if no diagrams are presented in either document and the table of contents is not very detailed, the fact that two authors contrast different kinds of typesetting is a good indication that the field of typesetting can be classified in more than one way.

Consultation with experts

Another method of uncovering different classifications of a subject field, or verifying that differing classifications already discovered are correct, is to consult a subject-field expert. For the COGNITERM project on optical storage technology, the multidimensional classification of OPTICAL STORAGE MEDIA (cf. figure 3.5) was indeed confirmed by an expert in the field.[56]

Once different unidimensional representations have been identified, they can later be amalgamated into a multidimensional representation. For example, as illustrated in figure 3-5 (cf. section 3.2.1), we have incorporated Chen's unidimensional classification of OPTICAL

[56] The subject-field expert was Mr. Don Slaunwhite, a software developer in the CD-ROM division of Corel Systems Corp.

STORAGE MEDIA according to the characteristic **degree of writability** of the media (cf. figure 3-1), and Heimburger's classification of the same concept based on the characteristic **physical form** of the media (cf. figure 3-2), into one multidimensional representation.

Terminologists should also remember that multidimensionality can occur at more than one level of a classification. For instance, in our work with the TKB, we discovered that not only was the classification of the concept OPTICAL STORAGE MEDIA multidimensional, but the classification of one of its subconcepts, OPTICAL DISC, was also multidimensional, as illustrated in figure 3-5. Terminologists should therefore consider the possibility that every concept may have a multidimensional representation, and not stop looking after one instance of multidimensionality has been found. To find different classifications of more specialized concepts, terminologists can again look in the tables of contents of specialized documents, examine the lexical items and diagrams contained in these documents, and consult subject-field experts.

4.1.2 Recognizing multidimensionality within one classification

In addition to recognizing different unidimensional classifications presented by different authors, terminologists must be able to recognize multidimensionality within one author's classification, even when it is not explicitly expressed. As discussed in section 2.3.2, repetition of concepts in a classification is actually an inelegant way of representing multidimensionality. Looking back to figure 2-6 (cf. section 2.3.1), we can see that the concept RESPIRATION appears in four different places in the classification of ANIMALS. This is because only one characteristic is used at each level to divide classes into subclasses. Each of these subclasses is then individually subclassified on the basis of another characteristic, which results in the repetition of concepts. An alternative to such repetition is to introduce characteristics at the same level, but in different dimensions. This is illustrated in the multidimensional representation of the classification of ANIMALS shown in figure 2–7 (cf. section 2.3.2).

4.1.3 Recognizing false multidimensionality

As well as recognizing true multidimensionality, terminologists must also be able to identify false multidimensionality. In section 3.2.1 (*Problem 1*), false multidimensionality was described as a classification error that occurs when a concept is classified at one level on the basis of two characteristics that should be applied within the same dimension, but at two *different* levels. An example of false multidimensionality is illustrated in Chen's classification (cf. figure 3–1 in section 3.1). Here, the class READ-ONLY MEDIA has been simultaneously subclassified on the basis of two different characteristics: 1) **physical form**, and 2) **size**. This incorrect classification can be rectified by first classifying READ-ONLY MEDIA on the basis of **physical form**, into the concepts OPTICAL DISC, OPTICAL FILM, OPTICAL TAPE, and OPTICAL CARD. The concept OPTICAL DISC can then be subclassified on the basis of **size**, into COMPACT DISC and VIDEODISC. A corrected version of Chen's classification can be seen in figure 3–4 a) in section 3.2.1 (*Problem 1*).

4.2 REPRESENTING MULTIDIMENSIONALITY

Once terminologists have identified exactly where in the subject field multidimensionality occurs, they must represent it. As we saw in section 3.2.1 (*Problem 2*), if all dimensions are not represented, then certain concepts (and hence, terms) may be omitted from the classification (and hence, from the term bank or terminological publication). For example, both Chen and Heimburger overlooked the concept OPTICAL DISC in their unidimensional representations (cf. figures 3–1 and 3–2 in section 3.2.1).

Although it is crucial to represent multidimensionality in order to have a complete classification of the subject field, it can be very confusing to work with multiple dimensions. To reduce the confusion, terminologists would be wise to follow the practice of taxonomists and work with only one dimension at a time. In order to know which dimension to work with, terminologists should follow the example set by library scientists of ranking the characteristics (and hence, dimensions). We have found, in our limited experience, that during the initial reading of the documentation in a subject field, one dimension seems to stand out as being more common or important than the others. As discussed in section 2.3.3, the dimensions can be ranked according to the frequency of occurrence in the documentation, and possibly the number of terms in the dimensions. Subject-field experts can also be consulted for help in determining the relative importance of the various dimensions. After ranking the dimensions, terminologists should begin working with the dimension that they consider to be the most essential to the subject field. Once they understand this dimension clearly and have completed it as exhaustively as possible, they can begin working on the dimension considered to be the next-most essential. This process of working through one dimension completely before moving on to the next one still holds true if multidimensionality occurs at different levels of the classification. A sample order for working on a multidimensional representation is illustrated in figure 4-1.

After the dimensions have been ranked, it is important that they be clearly distinguished with some form of visible notation (cf. section 3.2.2 (*Problem 1*)). In addition, it would be helpful to clearly indicate the characteristic underlying each dimension in a place that is easily accessible to the user (cf. section 3.2.2 (*Problem 2*)).

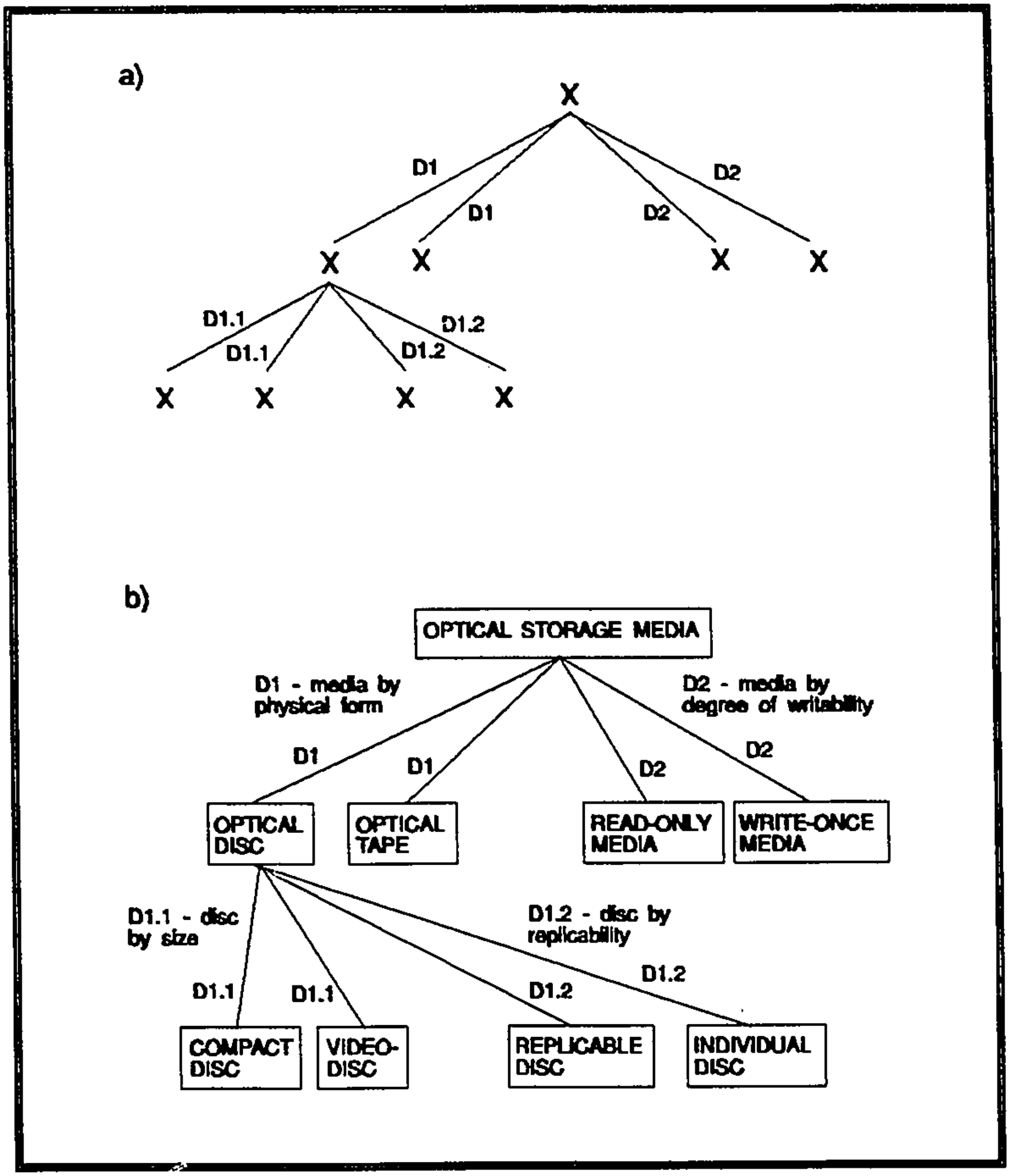


Figure 4–1. A suggested order for working through dimensions. Part a) shows a generic multidimensional representation. Terminologists should begin working on the concepts belonging to D1. Within D1, they should first complete D1.1 and then D1.2. Finally, they should move back up to D2. Part b) shows the same process for a partial classification of OPTICAL STORAGE MEDIA.

Finally, as mentioned in section 4.1, terminologists will benefit from working with a graphic representation of the concept system from the onset of the project. This graphic representation should be revised and updated throughout the project as terminologists learn more about the subject field.

4.3 NINE BASIC GUIDELINES

In summary, the information gleaned from the classification methods used in terminology, biology and library science, and from practical experimentation with the TKB on optical storage media, can be boiled down to the following nine basic guidelines.

We should explain that we have separated the guidelines for *recognition* and *representation* in order to be able to document them more clearly.[57] However, in actual fact, the procedure for handling multidimensionality is not strictly sequential; that is, terminologists do not recognize all aspects of multidimensionality, and then proceed to represent it fully. Rather, there is some overlap between the recognition and the representation: terminologists can begin to represent multidimensionality while they are in the process of recognizing it.

I. To recognize multidimensionality:

1) Work with a graphic representation, ideally a tree diagram, of the concept system. (If such a representation is not found in the documentation, create one.)

2) Consider all possible ways of classifying a subject field at all levels of the classification.

3) Recognize indications that multidimensionality exists (e.g. repetition of concepts, use of different contrasting lexical pairs).

4) Identify and correct false multidimensionality.

---

[57] Some of the guidelines, particularly guideline 9, contain elements that are important for both the recognition and representation of multidimensionality; however, to avoid unnecessary repetition, each guideline has been listed in only one place, according to whether it is most strongly associated with the recognition or the representation process.

II. To represent multidimensionality:

5) Rank the dimensions in order of importance.

6) Clearly distinguish the different dimensions on the graphic representation.

7) Clearly indicate the characteristic underlying each dimension in a place that is easily accessible to the user.

8) Revise the graphic representation of the concept system as more is learned about the subject field.

9) Work on one dimension at a time, completing it as exhaustively as possible before moving on to the next dimension.

PART III

CHAPTER FIVE

THE CODE SYSTEM

5.0 ARTIFICIAL INTELLIGENCE AND TERMINOLOGY

In the introduction to this thesis, and again in section 2.1.3, it was suggested that multidimensionality may have been largely neglected by terminologists because they did not have adequate tools that could aid them in dealing with this often complex phenomenon. However, recent developments in the field of artificial intelligence (AI), and particularly in the subfield of AI known as *knowledge engineering*, make the entire process of conceptual analysis more tractable. Since multidimensionality is one aspect of conceptual analysis, it is quite possible that knowledge engineering tools can assist in handling multidimensional concept systems.

Knowledge engineering is an activity that has grown out of the need to build computer systems that are *knowledge-based*. In the words of Meyer and Paradis (1991: 3):

> Knowledge-based systems, as opposed to traditional *algorithm-based* systems, are geared to solving problems in specialized fields of human expertise: as the term *knowledge-based* implies, these systems need to "know" what human experts know.

An important type of knowledge-based systems are the so-called *expert systems*. These systems are designed to function as expert consultants in areas such as medical diagnosis, financial decision making, and equipment repair.

In addition to expert systems, many other computational applications, such as machine translation, natural language processing, and database management, are already adopting a knowledge-based approach. As Meyer and Paradis observe (1991: 3): "According to a growing number of terminology researchers,[58] terminology banks of the future will also be knowledge-based to some degree." This observation is strengthened by the fact that there have already been two conferences devoted to terminology and knowledge engineering (TKE '87 and '90), and a third is planned for 1993 (TKE '93).

The central concerns of knowledge engineering are the *acquisition*, *formalization* and *refinement* of knowledge. Firstly, knowledge engineers, who are typically trained in computer science and are normally not true subject-field experts, must acquire information about the subject field in question. This may be problematic as information supplied by texts and even by human experts can often be unclear, inconsistent, or incomplete. Once the basic concepts have been acquired, they must be "engineered" into a form that will be suitable for the intended purpose. Finally, once formalized, the knowledge must be constantly refined; it is corrected or updated as the knowledge engineer's or expert's understanding of the subject field increases.

Clearly, the concerns of knowledge engineers are very similar to those of terminologists. Like knowledge engineers, terminologists are typically not true subject-field experts. Nevertheless, they must acquire knowledge about subject fields that can be very complex, formalize information about concepts in order to construct term records, and modify their knowledge about the subject field as they learn more from their research (cf. section 2.1.1), or as the subject field changes over time (cf. section 1.2.1). Since knowledge engineering has already made significant progress in developing computer aids to help with these tasks, terminology researchers would be wise to investigate the potential usefulness of existing knowledge engineering tools[59] for terminology purposes.

---

[58] Among others, these researchers include Juan C. Sager (1990), and Hans Czap and Christian Galinski (editors of the *Proceedings of the First International Congress on Terminology and Knowledge Engineering*, 1987).

[59] Knowledge engineering tools also go by various other names, including knowledge management tools, knowledge acquisition tools, and knowledge processing tools.

5.1 THE CODE SYSTEM

One such tool that has proved useful for terminology work is called *CODE* (*Conceptually Oriented Description Environment*). CODE has been under development at the Artificial Intelligence Lab of the University of Ottawa for over five years, and is currently being used in both commercial and research environments. Written in Smalltalk, the system runs on a UNIX, Macintosh, or 386 platform.

The purpose of CODE is to help the user create, debug, format, and retrieve knowledge from, a *knowledge base (KB)* that has both a textual and a graphical form (Skuce and Meyer 1990a: 189; Meyer and Paradis 1991: 4; Skuce 1992a: 2). The KB is organized into units called *conceptual descriptors* (CDs). In the words of Skuce (1992a: 6), "A CD can be thought of as a package containing the characteristics (i.e. knowledge) of some concept. Thus, a concept can be described, possibly even defined, by its associated CD."

Described by Skuce and Meyer (1991: 29–7) as "a general purpose knowledge management tool," CODE has two key features that many other knowledge management tools do not possess. Firstly, CODE is a *generic* tool. Whereas many tools of this nature are designed to handle only a specific type of knowledge (e.g. medical, legal),[60] CODE can be used to manage virtually any type of knowledge. Indeed, it has already been used for knowledge management in such diverse areas as software engineering, ontological design for knowledge-based systems, and terminology.[61] The second important feature of CODE that sets it apart from many other knowledge management tools is that CODE can be used by a non-computer expert. In other words, no knowledge of computer programming is required to use the system.[62]

---

[60] For example, the knowledge management system developed for the Botany Knowledge Base Project at the University of Texas at Austin is designed to handle only information related to the field of botany (cf. Porter *et al.* 1988).

[61] For more details about the use of CODE for these applications, please refer to Skuce (1992a), Skuce and Meyer 1990 a and b; Meyer, Bowker and Eck 1992 a and b; and Meyer, Skuce, Bowker and Eck 1992 a and b.

[62] The ONTOS system, developed at Carnegie Mellon University, uses a low level frame representation language. The user is required to understand the frame language and some Lisp as well (cf. Skuce 1992a: 25).

As previously mentioned, CODE has been used for a variety of knowledge management applications; however, for the purpose of this thesis, we will consider CODE in light of its use for managing specialized knowledge about concepts for terminology work. This particular application of CODE is currently being explored in the context of the COGNITERM Project, also based at the Artificial Intelligence Lab of the University of Ottawa. One objective of the COGNITERM Project is to use the CODE system to build a prototype term bank, called *COGNITERM*, which features a rich and formally structured knowledge component, and which is therefore described as a hybrid between a conventional term bank and a knowledge base, in other words a *terminological knowledge base* (*TKB*).[63]

5.1.1 System components

The CODE system[64] has three main components: graphs, CD views, and browsers. Knowledge entered into CODE can be accessed through each of these components types, which can be viewed in separate windows on the screen. Any number of windows may be open simultaneously, and individual windows may be sized or moved about the screen by the user at any time.

Graph

A graph is a graphical representation of all the concepts in a knowledge base and the relations between them; it gives users an overview of the entire subject field. A graph of an entire concept system may well cover an area much larger than the screen, but the user is able to scroll a graph using a mouse.

---

[63] For more details on the COGNITERM Project and the COGNITERM TKB, please refer to Meyer, Bowker and Eck 1992 a and b; and Meyer, Skuce, Bowker and Eck 1992 a and b.

[64] The description contained in this thesis is of *version 2* of the CODE system; version 4 is currently under development.

If users wish to view only a restricted portion of a graph, they may open a subgraph of a specific branch of the concept system. Part of a graph for the subject field optical storage technology, with a superimposed subgraph of the subfield read-only media, is shown in figure 5-1.

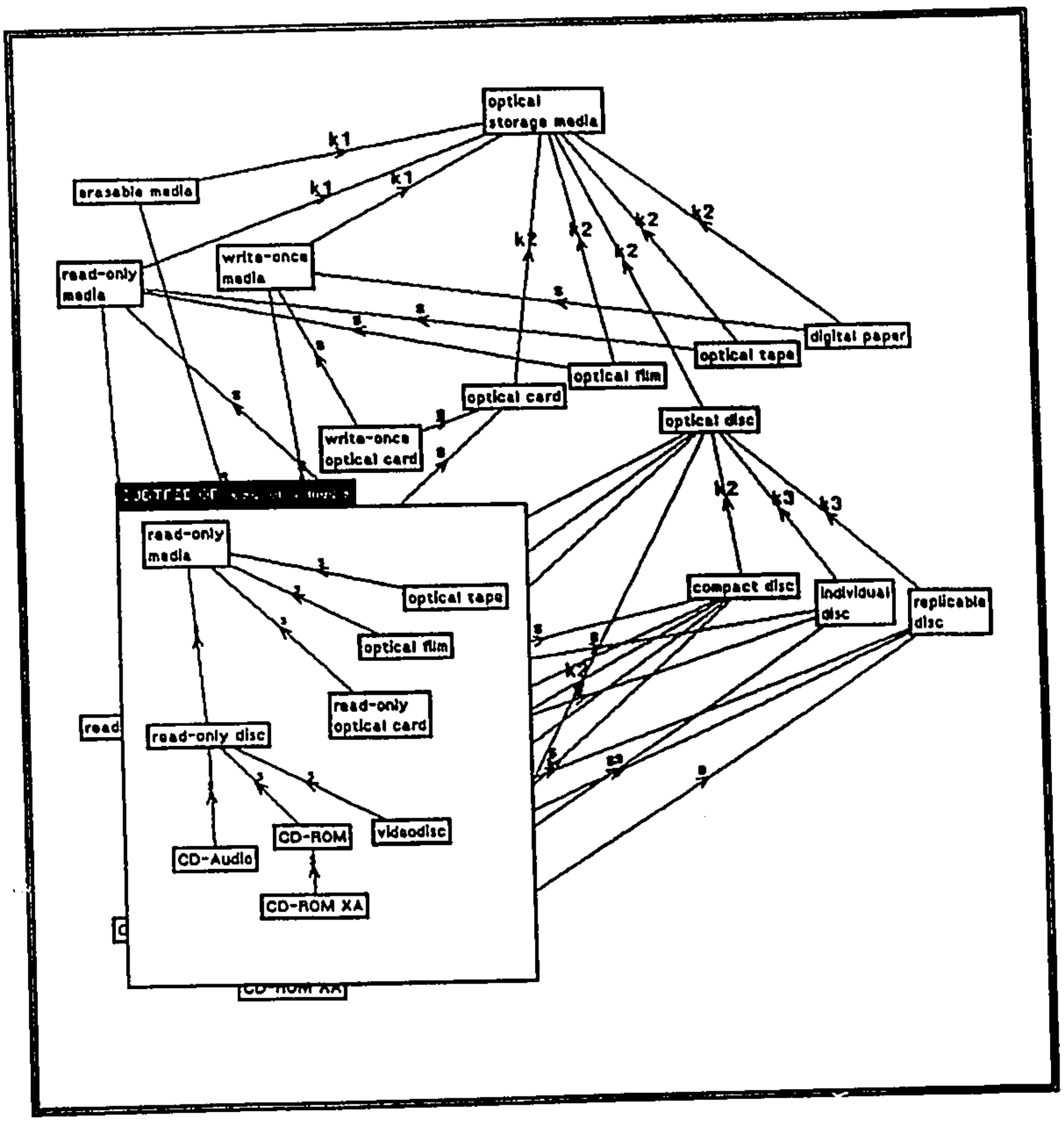


Figure 5-1. A portion of a CODE graph for the subject field optical storage media, with a superimposed subgraph of the subfield read-only media.

A second way that users can limit the amount of information on the screen is to temporarily hide concepts or subgraphs that they do not wish to view at a certain time. Users can hide concepts (or entire subtrees) by explicitly selecting them on a graph. Concepts can also be hidden in a more sophisticated manner by using CODE's masking facility, which will be discussed in more detail in section 5.1.3 (*Mask*).

CD view

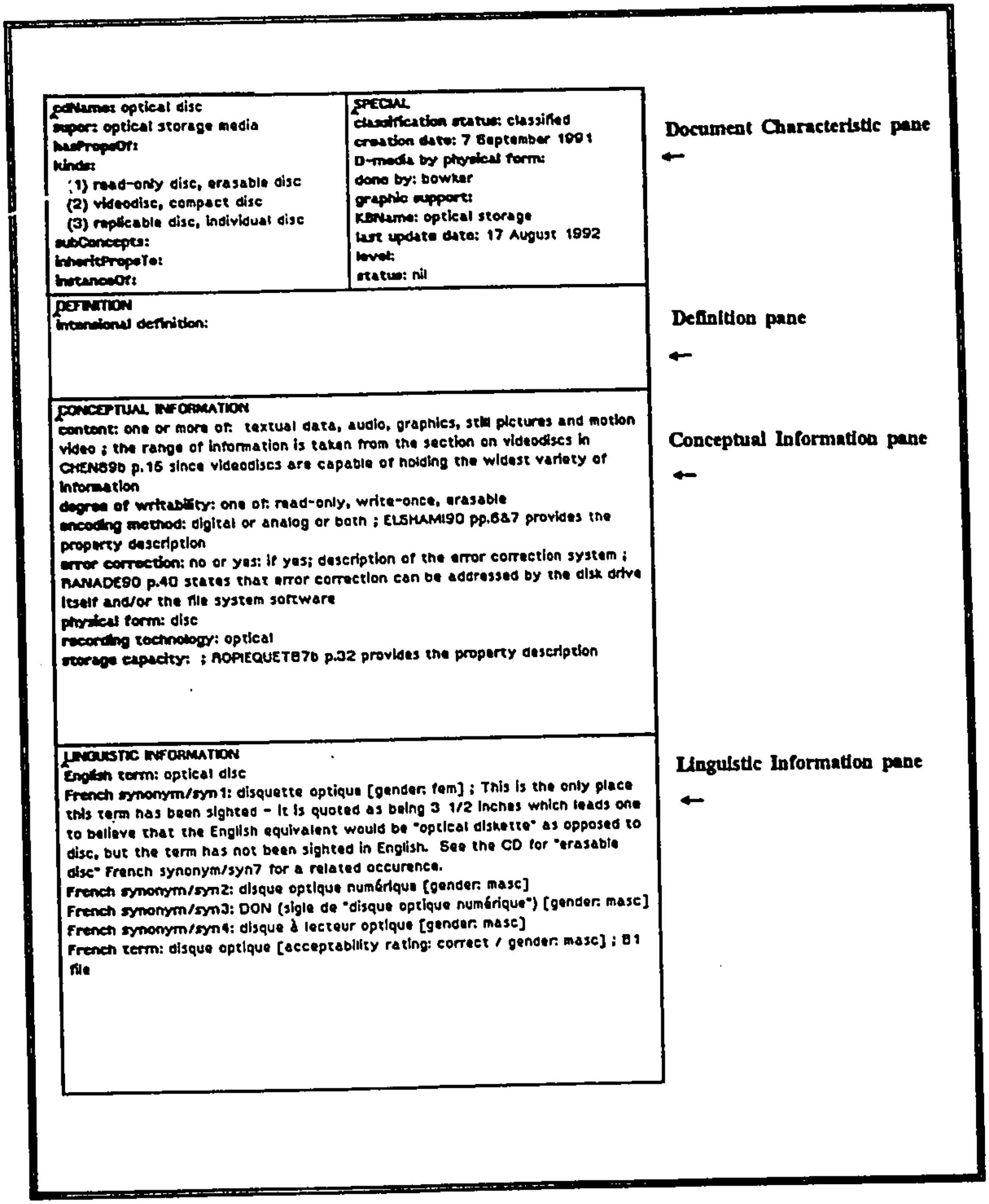
While a graph shows the overall picture of the subject field and lets users see how concepts interrelate in large knowledge structures, more detailed information about each individual concept that appears on a graph can be seen in a text-based, record-like structure called a CD view. A CD view has a number of partitions or *panes*, which represent different user-defined categories of information. As shown in figure 5-2, a CD view in the COGNITERM TKB, for example, can include a Conceptual Information pane, a Linguistic Information pane, and a Definition pane. The conceptual information is the knowledge base component of the TKB, the linguistic information is the term bank component, and the definition is seen as a kind of bridge between the two.[65] In CODE, information contained in a CD view is read-only; modifications are made in a browser, which will be discussed in the next section.

Browser

A browser, as its name implies, allows users to "browse" through all the information in the knowledge base, and also permits them to add, modify, or delete information.[66] Like a CD view, a browser is also partitioned into a series of panes. In figure 5-3 (a and b), the pane on the far left is called the CD list pane, and it lists all the CDs contained in the

[65] For more information about definitions in TKBs, please refer to Eck 1992.

[66] Information can also be added, modified, or deleted on a graph.



pdName: optical disc
super: optical storage media
hasPropsOf:
kinds:
(1) read-only disc, erasable disc
(2) videodisc, compact disc
(3) replicable disc, individual disc
subConcepts:
inheritPropsTo:
instanceOf:

SPECIAL
classification status: classified
creation date: 7 September 1991
D-media by physical form:
done by: bowker
graphic support:
KBName: optical storage
last update date: 17 August 1992
level:
status: nil

DEFINITION
intensional definition:

CONCEPTUAL INFORMATION
content: one or more of: textual data, audio, graphics, still pictures and motion video ; the range of information is taken from the section on videodiscs in CHEN89b p.16 since videodiscs are capable of holding the widest variety of information
degree of writability: one of: read-only, write-once, erasable
encoding method: digital or analog or both ; ELSHAMI90 pp.6&7 provides the property description
error correction: no or yes: if yes; description of the error correction system ; RANADE90 p.40 states that error correction can be addressed by the disk drive itself and/or the file system software
physical form: disc
recording technology: optical
storage capacity: ; ROPIEQUET87b p.32 provides the property description

LINGUISTIC INFORMATION
English term: optical disc
French synonym/syn1: disquette optique [gender: fem] ; This is the only place this term has been sighted - it is quoted as being 3 1/2 inches which leads one to believe that the English equivalent would be "optical diskette" as opposed to disc, but the term has not been sighted in English. See the CD for "erasable disc" French synonym/syn7 for a related occurence.
French synonym/syn2: disque optique numérique [gender: masc]
French synonym/syn3: DON (sigle de "disque optique numérique") [gender: masc]
French synonym/syn4: disque à lecteur optique [gender: masc]
French term: disque optique [acceptability rating: correct / gender: masc] ; 81 file

Figure 5-2. A CODE CD view for the concept OPTICAL DISC.

knowledge base. The CDs can be listed in either alphabetical or hierarchical order by clicking on the appropriate button (alpha or hiera) located directly below the CD list pane.

Like a complete graph, a browser contains a lot of information. Terminologists tend to work on one subfield at a time, and they can limit the number of CDs shown in the CD list pane by selecting the top concept in the subfield and then clicking on the hiera button once more. The list will be shortened to display only the selected CD and its superordinate and subordinate concepts. This is similar to opening a subgraph on a graph (cf. figure 5–1).

The top centre pane in a browser is the category list pane. This pane contains a list of different categories or types of information that are known about a concept. These categories are the same as those that appear in a CD view (i.e. Conceptual Information, Linguistic Information and Definition) (cf. figure 5–2).

The top right pane is called the characteristic list pane. It lists the names of all the characteristics possessed by the selected concept for the selected category of information. The value[67] of a selected characteristic is shown in the text pane located in the very bottom right corner of a browser. For example, in figure 5–3a, the selected concept is OPTICAL DISC, the selected category is *Conceptual Information*, and the selected characteristic is **encoding method** with a value of ***digital or analog or both***. In figure 5–3b, the selected concept is still OPTICAL DISC; however, the selected category is now *Linguistic Information*, the selected characteristic is **French term**, and the value of this characteristic is ***disque optique***.

Concepts, categories, and characteristics can be added, modified, or deleted in a browser by choosing the appropriate commands from mouse–driven menus. Any changes made in a browser will be automatically reflected in CD views and on graphs.

---

[67] The distinction between the *name* and the *value* of a characteristic was discussed in section 1.1.1.

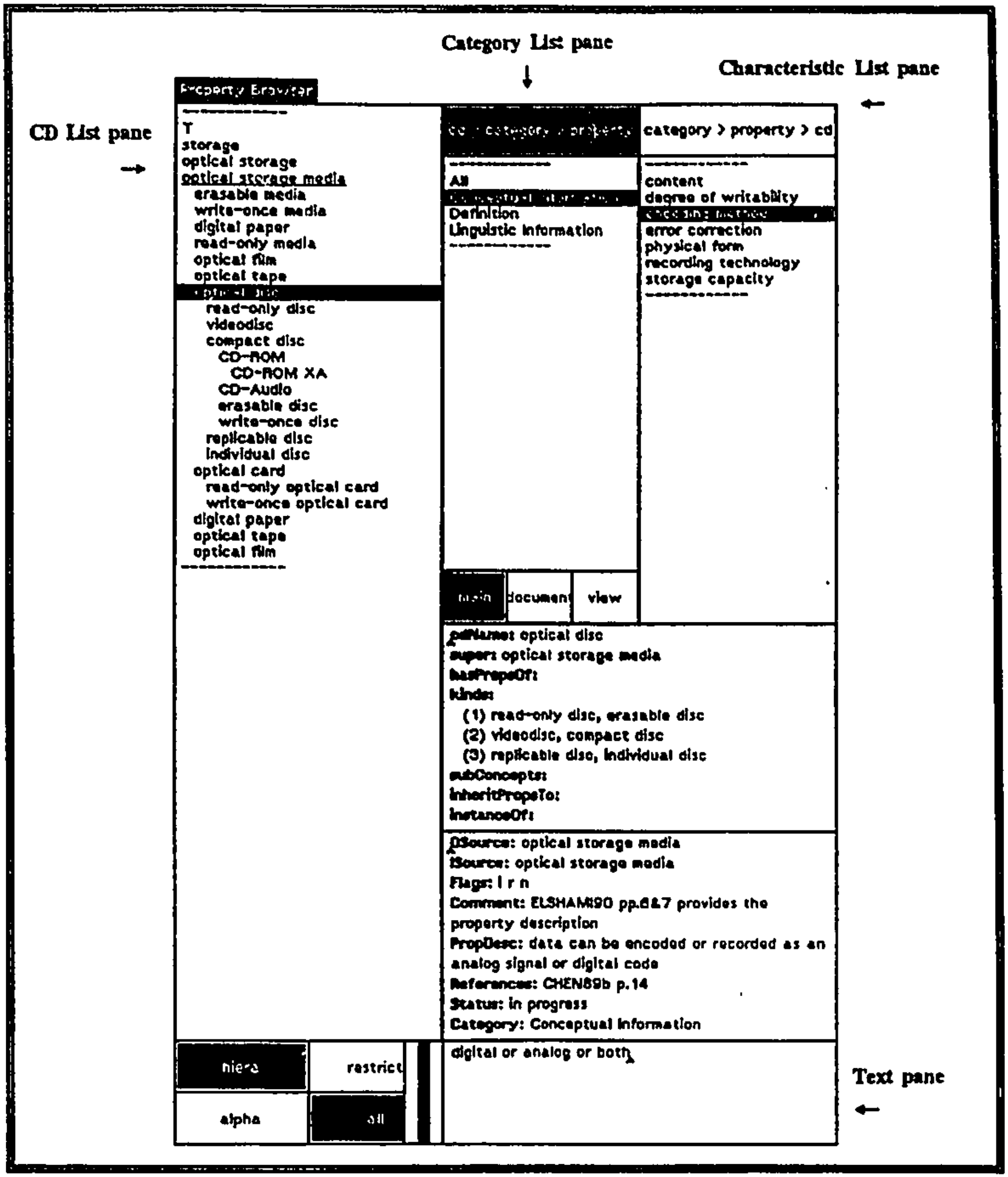


Figure 5-3a. A CODE browser with concepts listed hierarchically. The selected concept is OPTICAL DISC, the selected category is *Conceptual Information*, the selected characteristic is **encoding method**, and its value is ***digital or analog or both***.

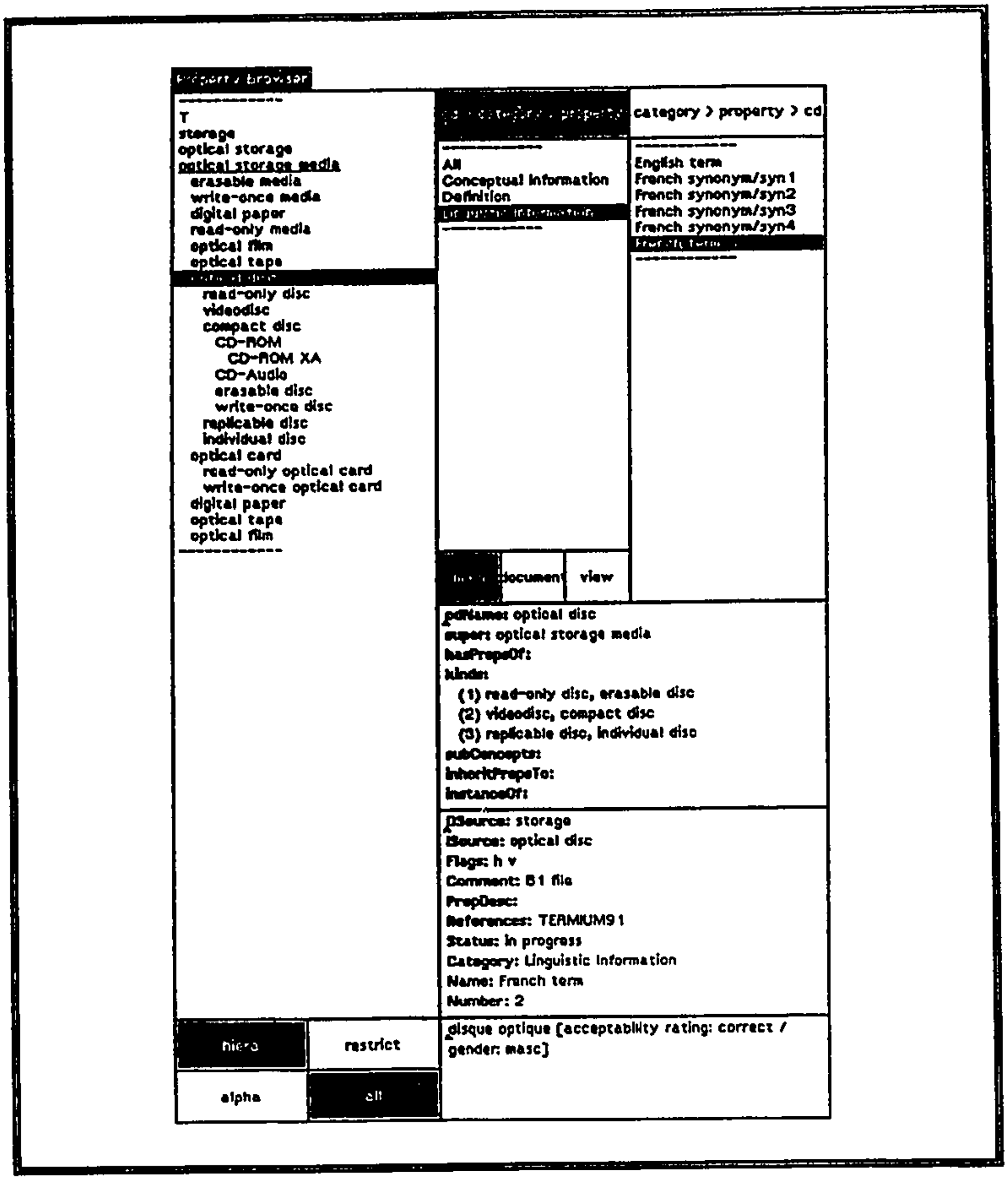


Figure 5-3b. Another CODE browser. In this browser, selected concept is still OPTICAL DISC, but the selected category is *Linguistic Information*, the selected characteristic is French term, and its value is *disque optique*.

5.1.2 CODE features that assist in good classification

Now that we are familiar with the basic components of the CODE system, we will examine some of CODE's features that assist in good classification. These features include explicit recording of all characteristics, inheritance, and detection of inconsistencies.

Explicit recording of characteristics

In CODE, the characteristics (both linguistic and conceptual) of any given concept are recorded in a simple format: **characteristic name:** ***characteristic value***. This means that the information is readily available, and users do not have to search through free natural language text to determine what characteristics or values are embedded there. Moreover, CODE has a characteristic description field, located in the facet pane of the browser, where terminologists can enter a description or explanation of the characteristic in cases where the characteristic name itself is not very transparent. For example, the characteristic description for the characteristic **degree of writability** might be: Degree of writability refers to whether the user can write to the media many times, only once, or not at all.

Inheritance

Within the CODE system, CDs are normally arranged in inheritance hierarchies (i.e. generic–specific hierarchies) so that more specific concepts may automatically inherit characteristics from more general ones. Inheritance allows any characteristic of a given concept to be implicitly true for all subordinate concepts of this concept, for all the subordinate concepts of the subordinate concepts, etc. For example, if the concept ALCOHOLIC BEVERAGE, shown in figure 5–4, has a characteristic **alcohol content**, then its subordinate concepts (FERMENTED ALCOHOLIC BEVERAGE and DISTILLED ALCOHOLIC BEVERAGE) will inherit this characteristic, as will all their subordinate concepts (e.g. BEER, WINE, VODKA, RUM, WHISKEY).[68] Thus, CODE's inheritance mechanisms eliminate the need for manual repetition of information from one level of a concept system to another.

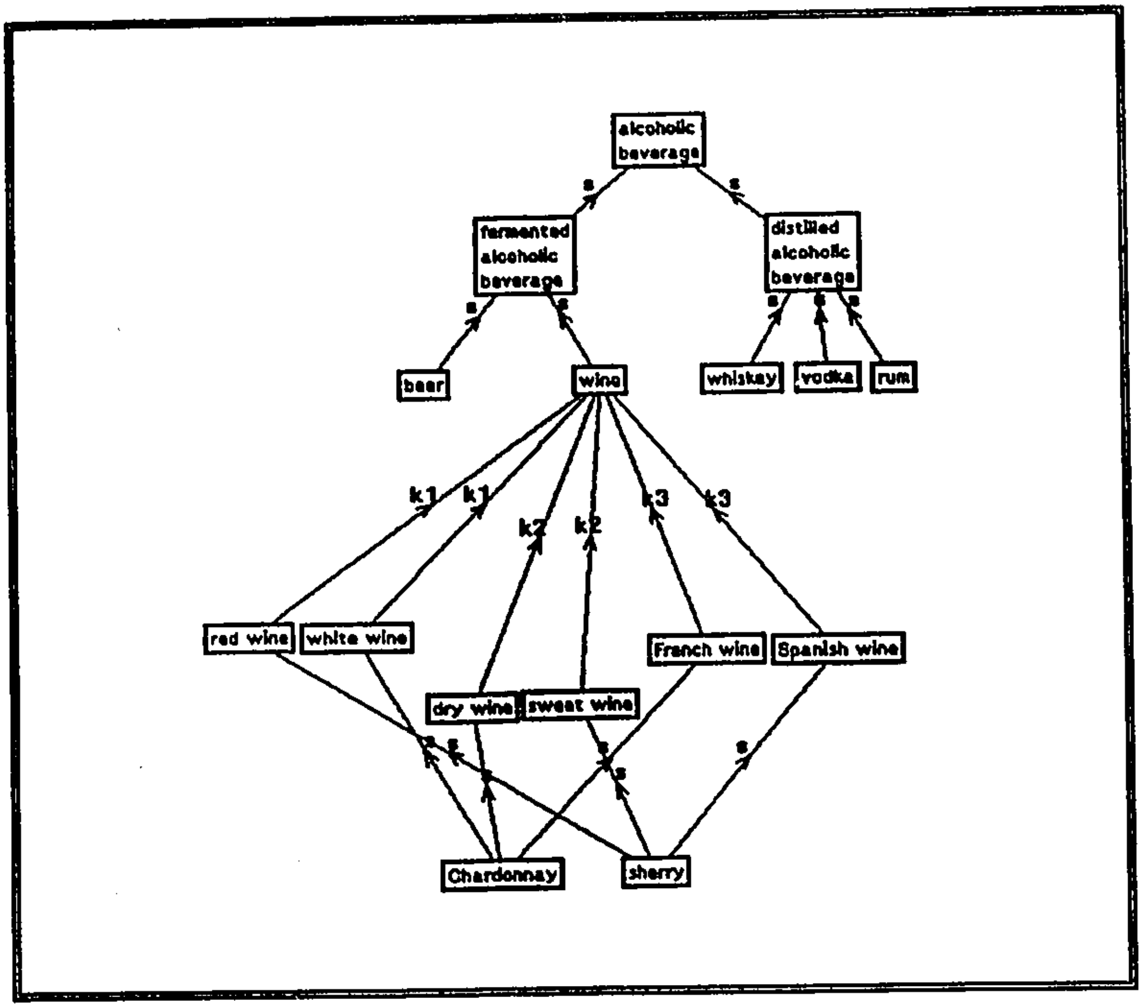


Figure 5-4. A CODE graph showing inheritance paths.

Detection of inconsistencies

CODE's inheritance mechanisms are associated with mechanisms for detecting inconsistencies in the classification. When a change is made at one hierarchical level, it will

[68] The *value* of the characteristic may change, but the subordinate concepts will still have a characteristic with the name **alcohol content**. In exceptional cases where a certain characteristic should not inherit at all, the inheritance can be blocked.

normally filter down to the lower levels. CODE has mechanisms for detecting undesirable repercussions (e.g. inconsistencies), and it signals these potential problems to the terminologist when they occur. For example, suppose a terminologist decided to introduce the characteristic **sugar content** in the CD for the concept DISTILLED ALCOHOLIC BEVERAGE. CODE's detection mechanisms could discern that this characteristic had already been introduced in the CD for FERMENTED ALCOHOLIC BEVERAGE, a co-ordinate concept of DISTILLED ALCOHOLIC BEVERAGE. If this were the case, a dialog box would automatically pop up on screen informing the terminologist that this characteristic had already been introduced in the concept DISTILLED ALCOHOLIC BEVERAGE, and suggesting that it therefore be introduced in the concept ALCOHOLIC BEVERAGE, which is the superordinate concept of both DISTILLED ALCOHOLIC BEVERAGE and FERMENTED ALCOHOLIC BEVERAGE. Detection mechanisms such as this help terminologists maintain clarity and consistency in their work.

5.1.3 CODE features that are particularly useful for handling multidimensionality

In addition to having features which assist in good classification, CODE also has several features which make it particularly suitable for handling multidimensionality. These features include multiple inheritance; kinds; document characteristics; a graphing facility which permits numbered k-links to be shown on graphs; and a masking facility.

Multiple inheritance

As discussed in section 5.1.2 (*Inheritance*), CODE supports inheritance mechanisms, which allow characteristics to be inherited from superordinate concepts to subordinate concepts. What is especially interesting for multidimensional representations is the possibility of multiple inheritance: a concept can have any number of superordinate concepts, including superordinate concepts from different dimensions.

For example, referring back to the graph shown in figure 5-4, we can see that the concept CHARDONNAY has three superordinate concepts: WHITE WINE, DRY WINE, and FRENCH WINE. In this case, CHARDONNAY would inherit the characteristics (and values[69]) of each of these three superordinate concepts; it would inherit the characteristic **colour** with a value of ***white*** from WHITE WINE, the characteristic **sugar content** with a value of ***dry*** from DRY WINE, and the characteristic **country of origin** with a value of ***French*** from FRENCH WINE. In addition, CHARDONNAY would inherit all the characteristics of the concepts WINE, FERMENTED ALCOHOLIC BEVERAGE, and ALCOHOLIC BEVERAGE from preceding levels of the classification.

Kinds

Another CODE feature that is crucial for handling multidimensionality is the possibility of indicating *kinds*. In CODE, a *kind* can be seen as the equivalent of a *dimension*; it can be interpreted as "a kind of way of subclassifying a concept."

If a classification is unidimensional, the user simply lists the subordinate concepts of a given concept in the *subConcepts* field of the browser.[70] For example, as shown in figure 5-5, if terminologists want to classify WINE solely on the basis of the characteristic **colour** into the concepts RED WINE and WHITE WINE, they would enter RED WINE and WHITE WINE in the subConcepts field of the browser, leaving the kinds field blank.

---

[69] The values can be specialized where appropriate.

[70] One might argue that a single dimension could still be entered as a kind; however, conventionally, a dimension is only indicated when it is found in opposition to one or more other dimensions.

cdName: wine
super: fermented alcoholic beverage
kinds:
subConcepts: red wine, white wine

Figure 5-5. The subConcepts field of the browser.

However, if terminologists wish to create a multidimensional representation of the classification of WINE, they could use the kinds field to indicate that a concept may be classified in more than one way. For example, as shown in figure 5-6, to classify the concept WINE into three dimensions based on the differentiating characteristics **colour, sugar content**, and **country of origin**, terminologists would enter RED WINE and WHITE WINE as one kind (based on **colour**), DRY WINE and SWEET WINE as a second kind (based on **sugar content**), and FRENCH WINE and SPANISH WINE as a third kind (based on **country of origin**). In this case, the subConcepts field would be left blank. As always, the graph would be automatically updated to reflect these changes by naming the links between WINE and its subordinate concepts k1, k2, etc. The graph's k-links are explained in more detail in the upcoming section entitled *Graph with k-links*.

cdName: wine
super: fermented alcoholic beverage
kinds: (1) red wine, white wine
(2) dry wine, sweet wine
(3) French wine, Spanish wine
subConcepts:

Figure 5-6. The kinds field of the browser.

Document characteristics

Another CODE feature which terminologists can make use of when handling multidimensionality is the Document characteristics pane (cf. figure 5-2). This pane is a type of catch-all place for any miscellaneous information that the terminologist feels is useful for documenting a given concept. It is in the Document characteristics pane that terminologists can record information regarding the dimension(s) of a concept.

Dimension information is recorded in the format "D-<superordinate concept> by <differentiating characteristic>," where the *D* stands for dimension. Continuing with the WINE example, in the CD for RED WINE, the terminologist would record the document characteristic as "D-wine by colour", and in the CD for DRY WINE, "D-wine by sugar content," etc. Like other characteristics, document characteristics are inherited to subordinate concepts. In cases where a concept belongs to more than one dimension, it will inherit document characteristics from all its superordinate concepts. For example, the document characteristics of the concept CHARDONNAY (cf. figure 5-4) would be:

D-wine by colour
D-wine by sugar content
D-wine by country of origin

By explicitly recording the dimension information, terminologists make it easy for users to find out what differentiating characteristic is underlying each dimension. This information is displayed in both browsers and CD views. Although the characteristic name does not appear directly on any graphs, users working with a graph can quickly open a CD view by clicking on a concept on the graph.

Graph with k-links

In section 4.2, it was suggested that terminologists would be wise to follow the lead of biologists and library scientists and systematically sketch out, in the form of a tree diagram, the concept system representation of a subject field. CODE's sophisticated graphing feature (cf. section 5.1.1 (*Graph*)) allows terminologists to easily sketch out the concept system representation; when terminologists enter a new concept into the system, they indicate its superordinate concept, and CODE automatically adds the new concept to the graph with a line linking the new concept to its superordinate concept. Skuce and Meyer (1990a: 190) stress the value of a graph, stating:

> We cannot overemphasize the importance of good graphical representations for terminology: in our experience the terminologists referred to the graph constantly and wanted the graph window open at all times, no matter what other components of the knowledge base were being focused on.

On a graph, concepts belonging to different dimensions can be distinguished by their link notations. Link notations are determined by the way relations between concepts are established in the browser. If a concept is not part of a multidimensional classification, it will be entered strictly as a subordinate concept of its superordinate concept in the subConcepts field of the browser (cf. figure 5-5). In CODE, this type of straightforward link between a subordinate concept and its superordinate concept is known as an *s-link* (where *s* stands for subordinate concept). S-links are indicated on a graph by a letter *s* with an arrow on the line pointing from a subordinate concept to its superordinate concept. S-links can be seen on the graph shown in figure 5-7.

In multidimensional representations, however, CODE allows terminologists to distinguish the different dimensions on a graph using special links called *k-links* (where *k* stands for *kind*, which can be seen as the equivalent of a dimension). K-links are created by entering the subordinate concepts into the browser as *kinds* of the superordinate concept (cf.

figure 5-6). The k-links on a graph are numbered in accordance with the entries in the kinds field. For example, all the concepts entered as kind (1) in the browser's kinds field will be labelled with a k1 on the graph, and those concepts entered as kind (2) will be labelled k2, etc. The numbered k-links, illustrated on the graph shown in figure 5-7, enable users to see at a glance which concepts on the graph belong to the same dimension. As mentioned in the previous section, if users wish to know the characteristic underlying each dimension, they merely need to open a CD view by clicking on a concept, and then refer to the dimension information recorded in the Document characteristics pane.

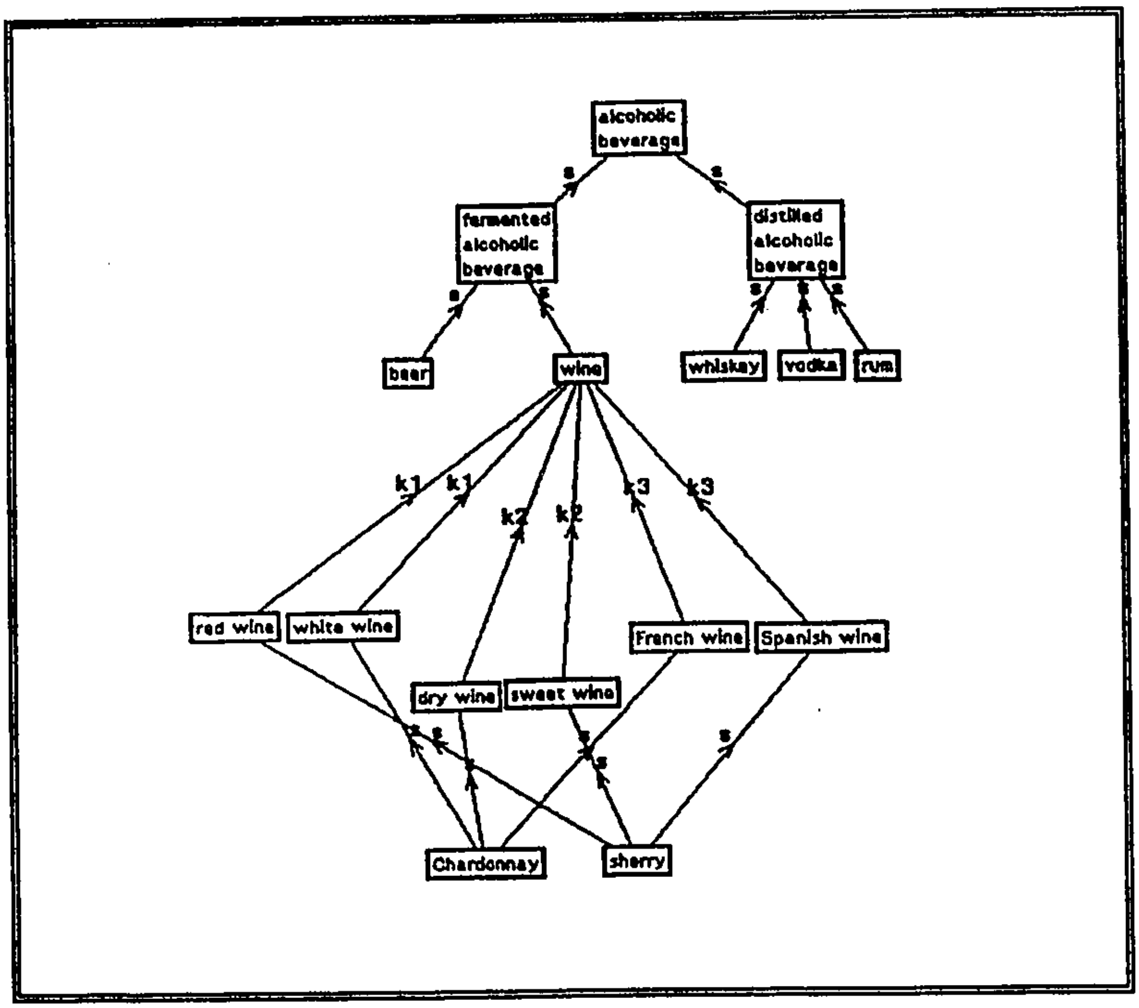


Figure 5-7. A graph showing s-links and k-links.

Mask

As established in section 4.2, multidimensionality can be quite complex when all the dimensions are viewed at once; therefore, terminologists might wish to work with only one dimension at a time. To help prevent information overload, CODE has a masking facility which permits users to restrict the number of concepts and characteristics viewed on screen, either in a browser or on a graph. One way in which the mask is commonly used is to specify that one wants to see only those concepts with the characteristic X. Thus, the mask can be used to temporarily hide unwanted dimensions by specifying the dimensional document characteristics (cf. section 5.1.3 (*Document characteristics*)). For example, using the mask, a terminologist could say to the system "Show me only those concepts having the document characteristic ***D-wine by colour***." The mask would then filter out those concepts which do not contain this document characteristic and temporarily hide them from view. Such a mask and the resulting graph are shown in figure 5-8. The mask can be removed at any time by saying "Show me all the concepts in the knowledge base." When a mask is removed, all previously hidden concepts reappear on the screen.

As we argued in section 2.2.3, it is better for terminologists to work with one dimension at a time. End-users may also be interested in viewing only one dimension and may not want the screen to be cluttered with unnecessary information. Even end-users who do want to view multiple dimensions may find it clearer to view them consecutively with the help of the mask, rather than simultaneously.

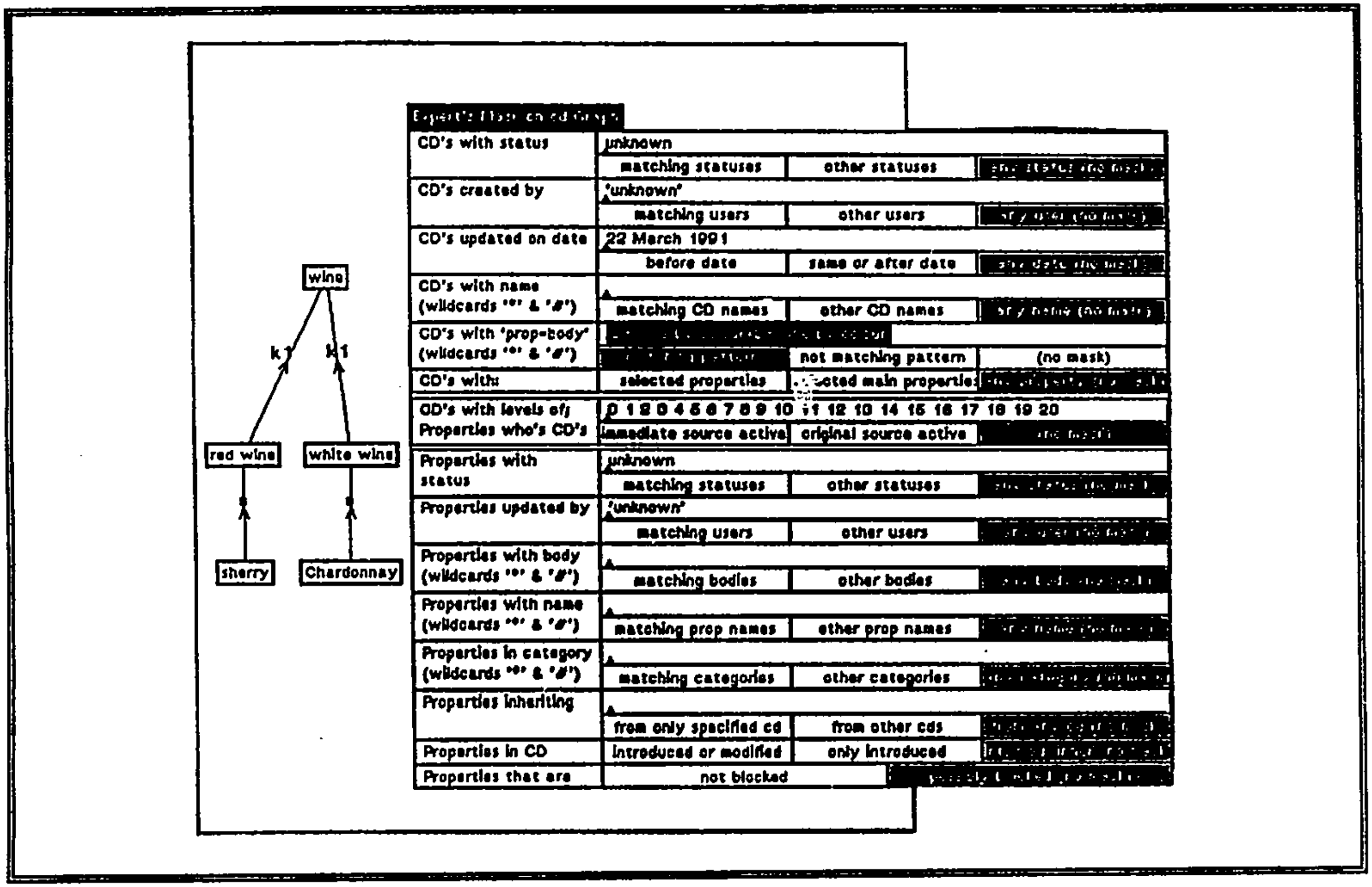


Figure 5–8. A mask and resulting graph. The mask is set to show only those concepts having the document characteristic ***D–wine by colour***; accordingly, only those concepts having this characteristic appear on the graph.

The CODE system clearly has many features that can help terminologists to handle multidimensionality more easily than they could using conventional paper-and-pencil methods. In the following chapter, we will see precisely how these features can be applied to a real example.

CHAPTER SIX

A STEP-BY-STEP EXAMPLE OF HOW CODE CAN FACILITATE THE HANDLING OF MULTIDIMENSIONALITY

6.0 SOURCE FOR THE EXAMPLE

In Chapter 3, we presented an extended example in which we identified occurrences of multidimensionality in the subfield optical storage media. In Chapter 5, we introduced the CODE system, and detailed those features which are most useful for handling multidimensionality. We will now continue with the second part of our extended example, which will show how CODE can facilitate the creation of a multidimensional concept system for the subfield optical storage media.

6.1 SKETCHING OUT THE PRELIMINARY REPRESENTATION OF THE CONCEPT SYSTEM

In section 3.1, we established that the concept OPTICAL STORAGE MEDIA could be subclassified in two different ways. In the first dimension, OPTICAL STORAGE MEDIA was classified into READ-ONLY MEDIA, WRITE-ONCE MEDIA, and ERASABLE MEDIA on the basis of the characteristic **degree of writability**. In the second dimension, OPTICAL STORAGE MEDIA was classified into OPTICAL DISC, OPTICAL CARD, OPTICAL FILM, OPTICAL TAPE, and DIGITAL PAPER on the basis of the characteristic **physical form**.

Using CODE, terminologists would begin drawing a preliminary sketch of the concept system representation from the moment they commenced their introductory reading. Terminologists would add concepts to the KB, either directly on a graph or in a browser, as soon as they were encountered in the documentation.

For example, in introductory readings on optical storage technology, when terminologists came across the concept OPTICAL STORAGE MEDIA, they would add it to the graph by choosing the command **add concept** from a pop-up menu, and typing OPTICAL STORAGE MEDIA at the prompt. The system would ask the terminologists to identify the superordinate concept of OPTICAL STORAGE MEDIA, and they would click on the concept OPTICAL STORAGE.[71] The graph would be automatically updated with a line (marked with an s-link) leading from the new concept, OPTICAL STORAGE MEDIA, to its superordinate concept, OPTICAL STORAGE. Initially, the concepts are positioned on the graph by the CODE system; however, the users can easily rearrange the layout by simply dragging the concept nodes to new positions and saving their locations.

Concepts that belong to a specific dimension would be entered as different kinds in the kinds field of the browser (cf. section 5.1.3 (*Kinds*)). For instance, as illustrated in figure 6-1a, the terminologists would enter the concepts READ-ONLY MEDIA, WRITE-ONCE MEDIA, and ERASABLE MEDIA as kind (1), and OPTICAL DISC, OPTICAL CARD, OPTICAL FILM, OPTICAL TAPE, and DIGITAL PAPER as kind (2). The concepts entered as kind (1) would have links marked k1 on the graph, and those entered as kind (2) would have links marked k2, as shown in figure 6-1b.

At this early stage of the project, it is likely that the concept system representation would be very rough, containing only the major, higher-level concepts. The terminologists would constantly refine and update the representation throughout the various stages of the project as their knowledge of the subject matter increases. The initial rough sketch of the concept system for the subfield optical storage media produced using the CODE graph would look as illustrated in figure 6-1b.

---

[71] This is assuming that the superordinate concept OPTICAL STORAGE is already part of the knowledge base.

a)

cdName: optical storage media
super: optical storage
kinds: (1) read-only media, write-once media, erasable media
(2) optical disc, optical card, optical tape, optical film, digital paper
subConcepts:

b)

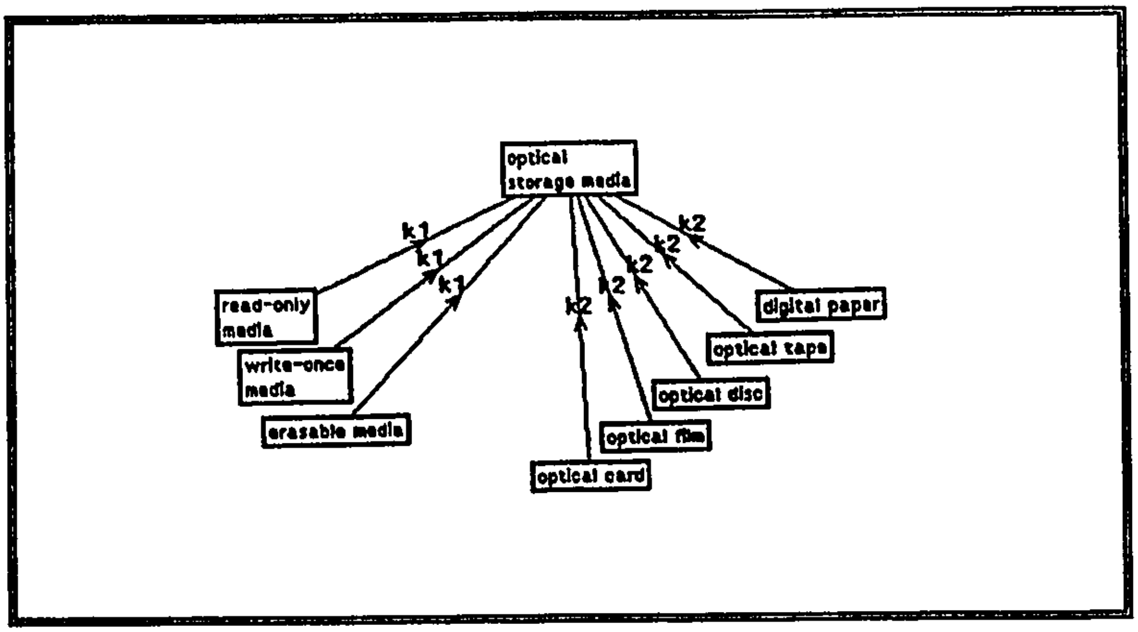


Figure 6–1. The two dimensions of OPTICAL STORAGE MEDIA. a) The subordinate concepts belonging to the different dimensions of OPTICAL STORAGE MEDIA are entered as two different kinds in the kinds field of the browser. b) Based on the information entered in the kinds field, a graph is *automatically* generated. On the graph, the links connecting the subordinate concepts in the different dimensions to their superordinate concept OPTICAL STORAGE MEDIA are marked k1 or k2 accordingly.

6.2 RANKING THE DIMENSIONS

After mapping out the initial sketch of the subject field, and indicating multidimensionality using CODE's kinds feature, terminologists would rank the existing dimensions (cf. section 4.2). CODE cannot be used to help terminologists actually *determine* a ranking order (this would be done as discussed in section 2.3.3); however, CODE's numerical notations, both in the browser's kinds field and on the graph (cf. figure 6–1a and b), make it easy for terminologists to remember the order once it has been determined.

Based on the introductory readings about optical storage media, terminologists would decide that the classification of OPTICAL STORAGE MEDIA into READ-ONLY MEDIA, WRITE-ONCE MEDIA and ERASABLE MEDIA was the most common and best-explained classification. Hence, it would be ranked as the first dimension, and terminologists would ensure that READ-ONLY MEDIA, WRITE-ONCE MEDIA, and ERASABLE MEDIA were entered in the kinds field as kind (1). The classification of OPTICAL STORAGE MEDIA into OPTICAL DISC, OPTICAL CARD, OPTICAL FILM, OPTICAL TAPE, and DIGITAL PAPER would therefore be ranked as the second dimension and its concepts would be entered as kind (2) (cf. figure 6–1a).

If the terminologists decided at any time to change the ranking order (e.g. if, after further reading, they discovered that the classification based on the characteristic **physical form** is actually the more common classification), they would simply rearrange the order of the kinds in the browser. That is, the concepts entered as kind (1) would be changed to kind (2), and vice versa.[72] These changes would be automatically reflected on the graph and in the CD views.

6.3 RECORDING THE CHARACTERISTICS UNDERLYING THE DIMENSIONS

Once the dimensions had been ranked, terminologists would next record the differentiating characteristic underlying each dimension because, as pointed out in section 3.2.2 (*Problem 2*), these characteristics are not always transparent.

[72] CODE has a cut-and-paste function that allows such changes to be made very easily.

In the first dimension of the optical storage media classification, terminologists would be able to determine that OPTICAL STORAGE MEDIA had been classified into READ-ONLY MEDIA, WRITE-ONCE MEDIA, and ERASABLE MEDIA on the basis of the characteristic **degree of writability** of the media (cf. section 3.2.2 (*Problem 2*)). Therefore, in the CD for each of these three subordinate concepts, terminologists would record ***D-media by degree of writability*** as a characteristic in the document characteristic pane (cf. section 5.1.3 (*Document characteristics*)).

They would then determine that in the second dimension, OPTICAL STORAGE MEDIA had been classified into OPTICAL DISC, OPTICAL CARD, OPTICAL FILM, OPTICAL TAPE, and DIGITAL PAPER on the basis of the characteristic **physical form**. The dimension information ***D-media by physical form*** would be duly recorded as a document characteristic in the CD for each of the subordinate concepts of the second dimension. Figure 6-2 shows the document characteristic for the concept OPTICAL DISC.

category > property > cd
----
creation date
D-media by physical form
done by
last update date
level
status

Figure 6-2. The document characteristic *D-media by physical form* in the Document characteristic pane of the CD for the concept OPTICAL DISC.

6.4 WORKING ON THE FIRST DIMENSION

In section 2.2.3, we argued that terminologists would likely find it easier to work with one dimension at a time. Once the dimension information has been recorded in the form of document characteristics, the mask can be used to filter out unwanted dimensions (cf. section 5.1.3 (*Mask*)). As illustrated in figure 6–3, terminologists would initially use the mask to show only those concepts in the first dimension, ***D-media by degree of writability***.

6.4.1 Selection of documentation

Next, terminologists would select documentation to use as their primary source of knowledge for the dimension in question. Although it may not be possible to find documents dealing exclusively with a given dimension, the selected documents must at least treat that dimension. Terminologists would consult the rough sketch of the concept system drawn with the CODE graph during the earlier stages of the project because the names of key concepts and characteristics already identified can provide entry points into the documentation (cf. section 2.1.1 (*Selection of documentation*).

Some of the documents that terminologists would select for scanning while working on the dimension based on the characteristic **degree of writability** would have titles or chapters containing key words which very transparently reflected this characteristic. For instance, "***WORMs*** for Mass Storage," and "The Reference ***ROM***" [italics mine] are two titles containing key words that are clearly related to the characteristic **degree of writability**.

When terminologists finished working on the first dimension and began working on the second, they would choose documents that focused on the **physical form** of the media, such as "An Optical Memory Using Suboxide Thin ***Film***," "Applications for the Drexxon ***LaserCard***," and "Optical ***Disks***: Systems and Applications" [italics mine].

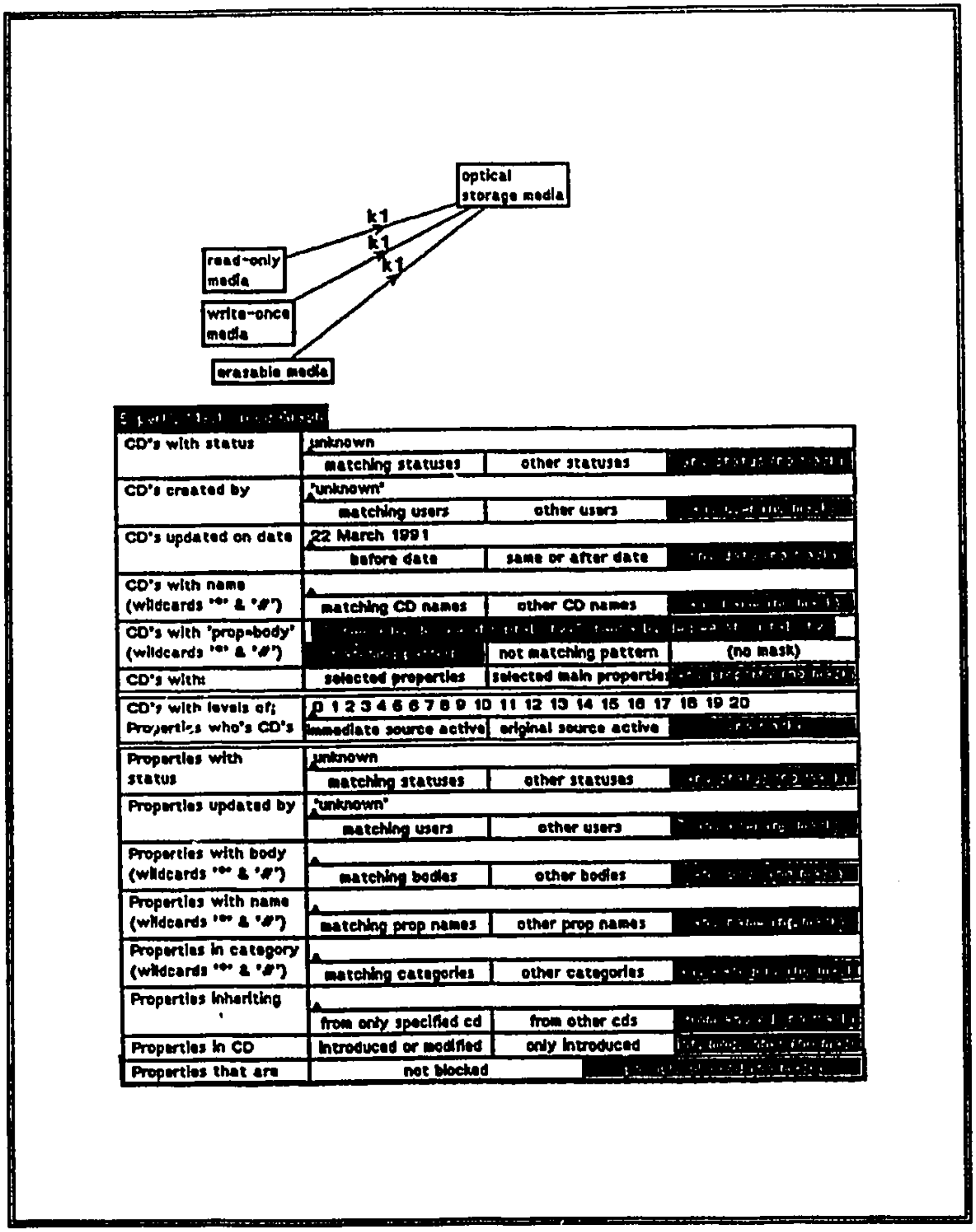


Figure 6-3. Dimension 1: based on the characteristic **degree of writability**. The mask has been used to filter out those concepts not containing the document characteristic ***D–media by degree of writability***.

6.4.2 Scanning

Having selected their documentation, terminologists would proceed to scan it, extracting potential terms and contexts. Each time terminologists encountered a potential term, they would create a CD for the term (cf. section 5.1), filling in the appropriate superordinate concept(s) when known[73] so that the pertinent characteristics would be inherited, and the graph correctly updated. They would also record new characteristics or specialize values of existing characteristics as they came across relevant information. They would continue scanning, creating CDs, and explicitly recording characteristics until they felt that they had covered the terminology in the first dimension as exhaustively as possible. Figure 6–4 shows a CD view for the concept OPTICAL DISC; the characteristics have been explicitly recorded in CODE's easy-to-read format.

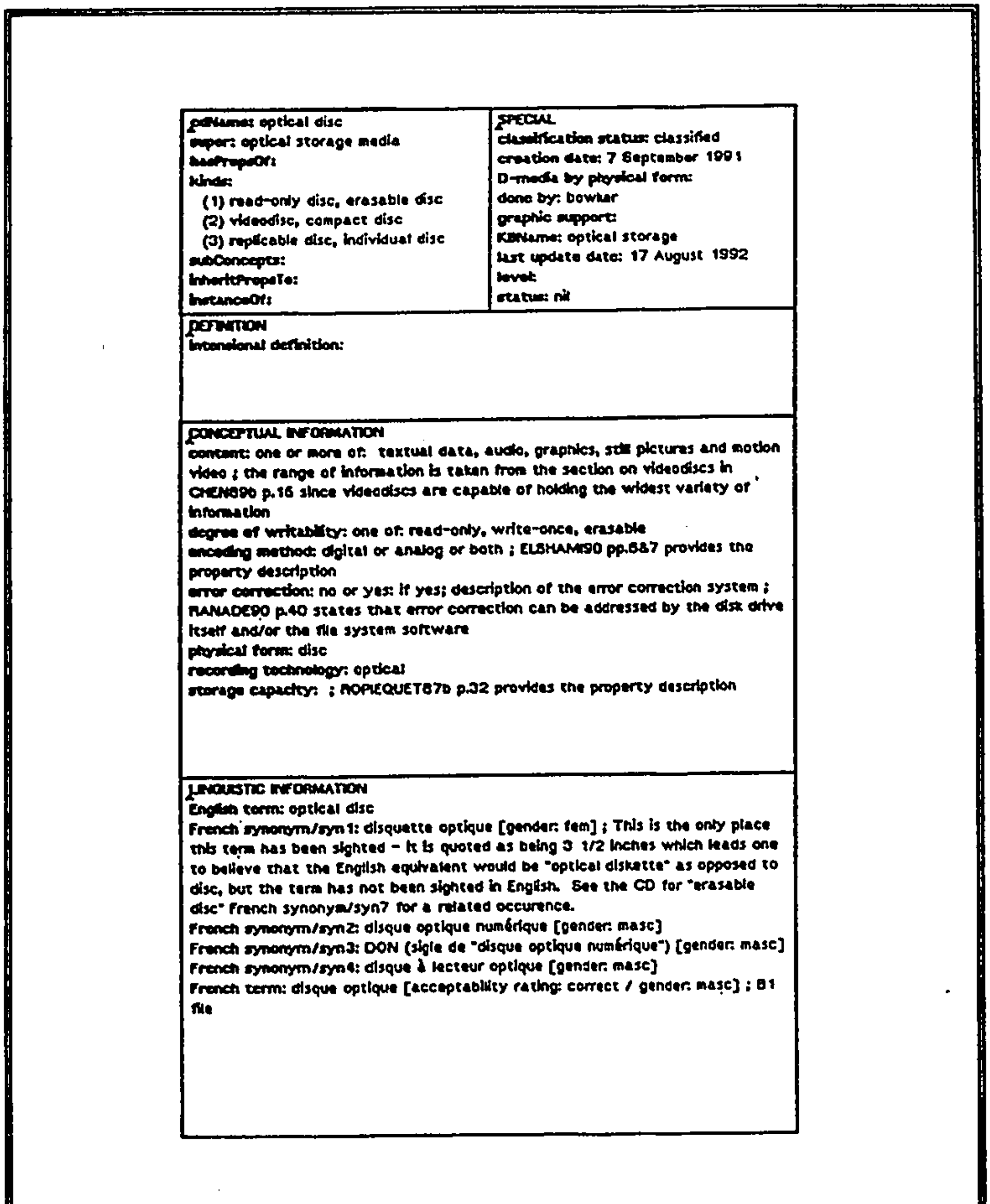
6.5 WORKING ON SUBSEQUENT DIMENSIONS

When work on a given dimension has been completed as fully as possible, terminologists would then begin working on the next-most essential dimension (cf. figure 4–1). In this case, when they had finished working on the dimension ***D-media by degree of writability***, the terminologists would mask out this dimension and began working on the dimension ***D-media by physical form***, as illustrated in figure 6–5. They would select new documents (cf. section 6.4.1), scan for relevant terms, and create and fill in CDs accordingly.

Terminologists would discover that the concept OPTICAL DISC can be subclassified on the basis of more than one characteristic, thus creating a second level of multidimensionality within the classification. When a second (or third, etc.) level of multidimensionality is discovered, terminologists would follow the same procedure, working through one complete dimension before moving on to the next, until finally, all dimensions have been completed (cf. figure 4–1 in section 4.2).

---

[73] When the superordinate concept is not known, the concepts can be put into a sort of *miscellaneous* category, represented by a dummy concept, until more knowledge is acquired. For more details on this miscellaneous category, please refer to Meyer, Bowker and Eck (1992b: 161–2).



pdName: optical disc
super: optical storage media
hasPropsOf:
kinds:
(1) read-only disc, erasable disc
(2) videodisc, compact disc
(3) replicable disc, individual disc
subConcepts:
inheritPropsTo:
instanceOf:

SPECIAL
classification status: classified
creation date: 7 September 1991
D-media by physical form:
done by: bowker
graphic support:
KBName: optical storage
last update date: 17 August 1992
level:
status: nil

DEFINITION
intensional definition:

CONCEPTUAL INFORMATION
content: one or more of: textual data, audio, graphics, still pictures and motion video ; the range of information is taken from the section on videodiscs in CHEN090 p.16 since videodiscs are capable of holding the widest variety of information
degree of writability: one of: read-only, write-once, erasable
encoding method: digital or analog or both ; ELSHAMI90 pp.6&7 provides the property description
error correction: no or yes: if yes; description of the error correction system ; RANADE90 p.40 states that error correction can be addressed by the disk drive itself and/or the file system software
physical form: disc
recording technology: optical
storage capacity: ; ROPIEQUET87b p.32 provides the property description

LINGUISTIC INFORMATION
English term: optical disc
French synonym/syn1: disquette optique [gender: fem] ; This is the only place this term has been sighted - it is quoted as being 3 1/2 inches which leads one to believe that the English equivalent would be "optical diskette" as opposed to disc, but the term has not been sighted in English. See the CD for "erasable disc" French synonym/syn7 for a related occurence.
French synonym/syn2: disque optique numérique [gender: masc]
French synonym/syn3: DON (sigle de "disque optique numérique") [gender: masc]
French synonym/syn4: disque à lecteur optique [gender: masc]
French term: disque optique [acceptability rating: correct / gender: masc] ; B1 file

Figure 6–4. A CODE CD view for the concept OPTICAL DISC.

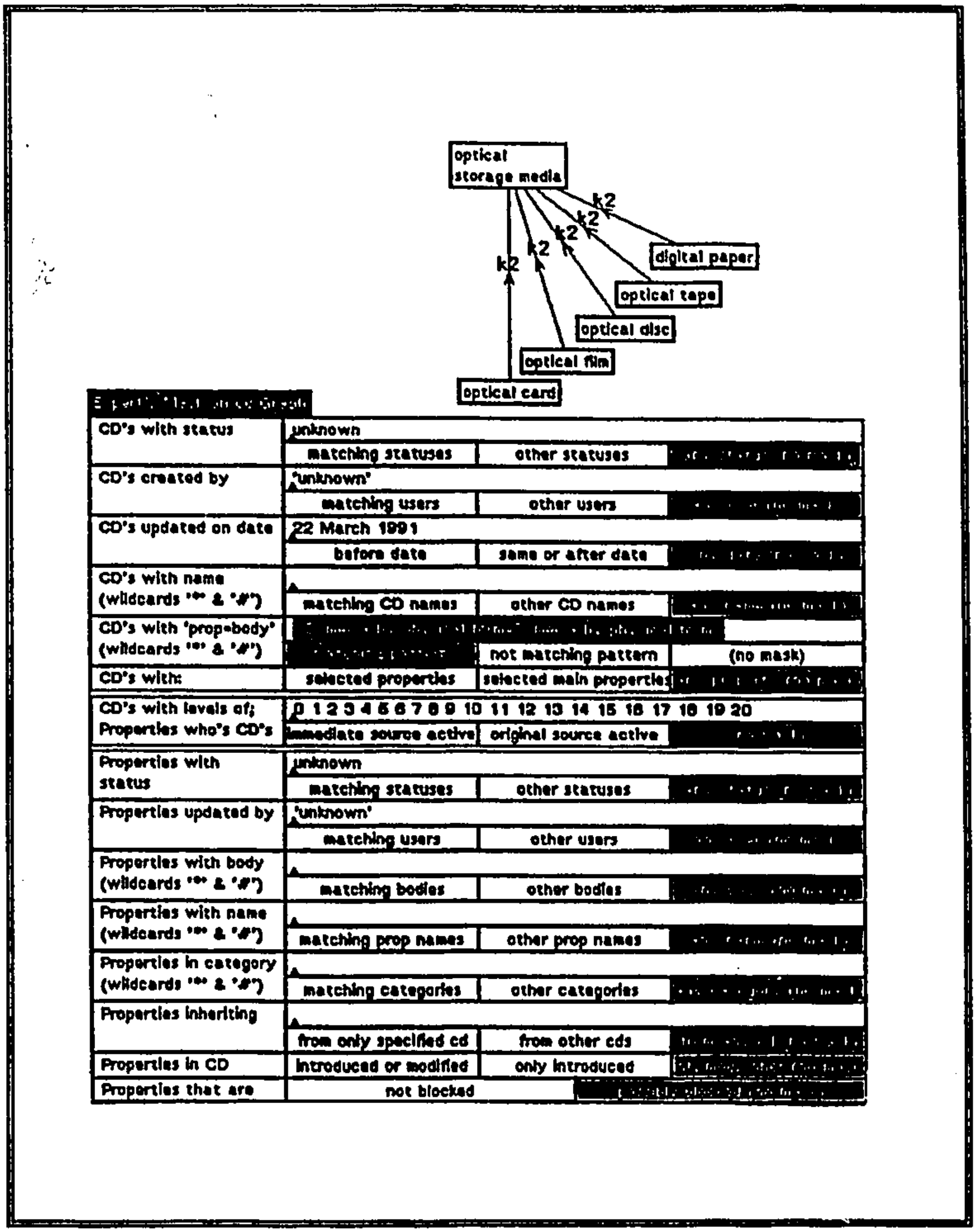


Figure 6–5. Dimension 2: based on the characteristic **physical form**. The mask has been used to filter out those concepts not containing the document characteristic ***D-media by physical form***.

Figure 6-6 shows an unmasked, completed multidimensional representation of the subject field optical storage media. Multidimensionality can be seen at two levels of the classification: OPTICAL STORAGE MEDIA has two dimensions (based on the characteristics **degree of writability** and **physical form**), and OPTICAL DISC has three (based on **degree of writability**, **size**, and **replicability**).

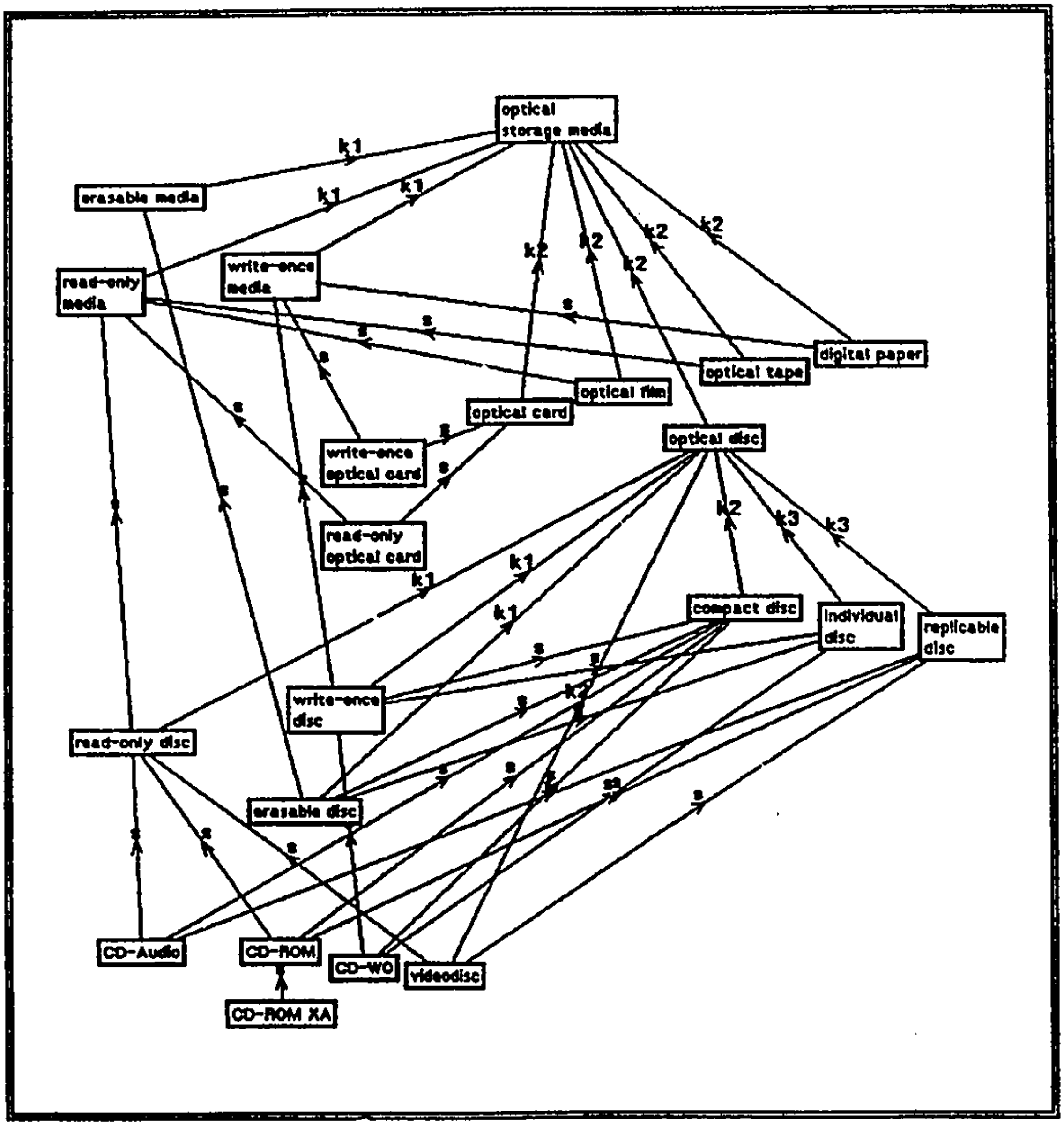


Figure 6-6. An unmasked, completed multidimensional representation of the subject field optical storage media.

CHAPTER SEVEN

TECHNICAL COMPLEMENT TO THE INITIAL GUIDELINES

7.0 SOURCES FOR THE TECHNICAL COMPLEMENT

In Chapter 5, we described the CODE system, detailing those features that were particularly useful for handling multidimensionality. In Chapter 6, using an extended example, we illustrated how CODE could assist terminologists in creating a multidimensional representation of a concept system. Based on the research carried out for these two chapters, we would now like to add a technical complement to the initial guidelines presented in Chapter 4. This technical complement will summarize how the CODE system *facilitates* the performance of some of the tasks enumerated in the initial guidelines for recognizing and representing multidimensionality.

As mentioned in section 4.3, although the guidelines for *recognizing* and *representing* multidimensionality have been documented separately, the procedure for handling multidimensionality is not strictly sequential.

7.1 HOW CODE HELPS TERMINOLOGISTS RECOGNIZE MULTIDIMENSIONALITY

The CODE system can be used to *aid* terminologists, but it cannot do their work for them. The responsibility for recognizing multidimensionality in the documentation lies with the terminologist; however, CODE does afford some help for this recognition by providing graphical support, and by forcing terminologists to organize their thoughts and to work systematically.

In Chapter 4, we established the following four guidelines for helping terminologists to recognize multidimensionality.

1) Work with a graphic representation, ideally a tree diagram, of the concept system. (If such a representation is not found in the documentation, create one.)

2) Consider all possible ways of classifying a subject field at all levels of the classification.

3) Recognize indications that multidimensionality exists (e.g. repetition of concepts, use of different contrasting lexical pairs).

4) Identify and correct false multidimensionality and other classification errors.

Guideline 1) stresses that terminologists should work with a graphic representation of the concept system (cf. section 4.2). Sometimes, adequate tree diagrams can be found in the documentation; however, in cases where graphic representations cannot be found, CODE's graph permits terminologists to sketch out their own quite easily. They simply type the name of the concept to be added, and indicate its superordinate concept(s). CODE automatically creates a box containing the name of the new concept (i.e. a node), and draws the line(s) (i.e. link(s)) to the appropriate superordinate concept(s). In cases where a number of different tree diagrams are found in the documentation, CODE can be used to amalgamate them into a single diagram.

Guideline 2) suggests that terminologists must consider all possible ways of classifying a subject field. Once terminologists have a graph to work with, it is easier for them to see whether they have considered all possible ways of classifying a subject field.

Similarly, the graph makes it easier for terminologists to see whether there is repetition of concepts within the classification, which, as stated in guideline 3), is one possible indication of the presence of multidimensionality. Such repetition will be more apparent on a visual representation than it will if it is embedded in free text.

Finally, guideline 4) recommends that terminologists identify and correct classification errors such as false multidimensionality. The best way to identify and correct such errors is by using good classification techniques, such as those promoted by CODE's features. For instance, CODE's inheritance mechanisms, coupled with its mechanisms for detecting and signalling inconsistencies (cf. section 5.1.2 (*Inheritance* and *Detection of inconsistencies*) help to ensure that information is correctly and consistently recorded in relevant CDs. Moreover, the explicit recording of the characteristics in an easy-to-read format (cf. section 5.1.2 (*Explicit recording of characteristics*) means that information is readily accessible to users.

7.2 HOW CODE HELPS TERMINOLOGISTS REPRESENT MULTIDIMENSIONALITY

CODE provides some assistance for recognizing multidimensionality, but it is even more useful for facilitating the representation of multidimensionality.

The following guidelines for helping terminologists to represent multidimensionality were established in Chapter 4.

5) Rank the dimensions in order of importance.

6) Clearly distinguish the different dimensions on the graphic representation.

7) Clearly indicate the characteristic underlying each dimension in a place that is easily accessible to the user.

8) Revise the graphic representation of the concept system as more is learned about the subject field.

9) Work on one dimension at a time, completing it as exhaustively as possible before moving on to the next dimension.

Guideline 5) is concerned with the ranking of dimensions. As mentioned in section 6.2, CODE cannot actually help terminologists to rank the dimensions, but its numerical notations make it easy for them to remember the ranking order once it has been established. If the terminologists wish to alter the ranking order as they learn more about the subject field, the changes can easily be reflected in the CODE system by rearranging the order of the kinds in the browser (cf. section 6.2); these changes would be automatically updated on the graph and in the CD views.

CODE's notations, which show up as numbered kinds in the browser and numbered k-links on the graph, are also a tidy way of clearly differentiating the dimensions, as suggested in guideline 6). If the subordinate concepts are correctly entered in the kinds field of the browser (cf. figure 5–6), then the lines connecting the subordinate concepts to their superordinate concept on the graph are *automatically* marked with a numbered k–link. All concepts belonging to the first (and most important) dimension will be marked k1, those belonging to the second dimension are marked k2, etc. (cf. figure 5–7).

Guideline 7) stated that it would be helpful to indicate the characteristic underlying each dimension, especially in cases where the characteristic is not transparent (e.g. the classification of OPTICAL DISC into COMPACT DISC and VIDEODISC based on the characteristic **size**). In CODE, the characteristic underlying each dimension can be added as a document characteristic (cf. section 5.1.3 (*Document characteristics*). The dimension information is then displayed in the Document characteristic pane in the browser and CD view, and can be easily accessed from the graph by clicking on a concept and opening its CD view. In this way, the user can see at a glance, without having to wade through free natural language text, which characteristic the dimension has been based on. These document characteristics are also inherited to subordinate concepts, which helps terminologists remain consistent in their work and eliminates the need for manual repetition of information from generic to specific concepts.

Guideline 8) recommends that terminologists revise and update the graphic representation as their knowledge of the subject field expands. Modifying the CODE concept system representation is extremely simple. For example, if the terminologists decide that a concept should have a different (or additional) superordinate concept, they can reparent this concept by merely changing (or adding) the name of the superordinate concept in the browser. Furthermore, information entered directly in the browser is automatically updated on the graph and in the CD views, and vice versa.

Finally, guideline 9) suggested that terminologists may find it less confusing to work with one dimension at a time, completing it as fully as possible before going on to work on the next dimension. CODE's masking facility (cf. section 5.1.3 (*Mask*)) can be used in conjunction with the dimension information found in the Document characteristics pane (cf. section 5.1.3 (*Document characteristics*)) to temporarily hide the concepts belonging to unwanted dimensions (cf. figure 5-8). The terminologist simply has to say, for example, "Show me only those concepts with the characteristic ***D-media by writability***," and all concepts not having this characteristic (i.e. all concepts not belonging to the dimension based on this characteristic) will temporarily disappear from view. Masks can be applied to both graphs and browsers.

7.3 SUMMARY OF THE TECHNICAL COMPLEMENT TO THE INITIAL GUIDELINES

The following two tables summarize the ways in which the CODE system can facilitate the 1) recognition and 2) representation of multidimensionality. The left-hand column of each table lists the initial guidelines (cf. section 4.3), while the right-hand column outlines the technical complement to these initial guidelines.

Initial Guidelines for Recognizing Multidimensionality	*Technical Support in CODE for Recognizing Multidimensionality*
1) Work with a graphic representation, ideally a tree diagram, of the concept system. (If such a representation is not found in the documentation, create one.)	· The CODE graphing facility can be used to easily sketch out a rough concept system representation. The user merely has to enter the name of the concept to be added, and then click on its superordinate concept(s). The nodes and links are automatically generated by the system.
2) Consider all possible ways of classifying a subject field at all levels of the classification.	· The CODE graph gives users an overview of the entire subject field, making it easier for them to determine if all possible ways of classifying the subject field have been considered.
3) Recognize indications that multidimensionality exists (e.g. repetition of concepts, use of different contrasting lexical pairs).	· CODE automatically signals the user if the same concept or characteristic is entered into the knowledge base in two different places.
4) Identify and correct false multidimensionality.	· CODE forces the user to employ good classification techniques by 1) requiring that characteristics be explicitly recorded in an easy-to-read format, and 2) providing powerful mechanisms for inheritance and for detection of inconsistencies.

Table 7.1. A summary of the technical complement to the initial guidelines for *recognizing* multidimensionality. The left-hand column lists the four initial guidelines for recognizing multidimensionality (cf. section 4.3). The right-hand column lists the ways in which the CODE system can facilitate this recognition.

Initial Guidelines for Representing Multidimensionality	*Technical Support in CODE for Representing Multidimensionality*
5) Rank the dimensions in order of importance.	· CODE's kinds field allows users to enter concepts in different dimensions in the order of ranking. This order is automatically indicated on the graph (with numbered k-links). Changes in the ranking order can easily be reflected by using CODE's cut-and-paste function to re-order the information in the kinds field. These changes will be automatically reflected on the graph.
6) Clearly distinguish the different dimensions on the graphic representation.	· k-links on CODE graphs are automatically numbered (one number per dimension) by the system in accordance with the information entered in the kinds field.
7) Clearly indicate the characteristic underlying each dimension in a place that is easily accessible to the user.	· CODE allows the characteristic underlying each dimension to be recorded as a document characteristic. These characteristics are inherited to subconcepts. Document characteristics can be viewed in browsers and CD views; users working with a graph can easily open a CD view for any given concept by clicking on it.
8) Revise the graphic representation of the concept system as more is learned about the subject field.	· CODE has many features to facilitate revision: · automatic updating between the graphs, browsers and CD views; · inheritance of conceptual and document characteristics; · detection of inconsistencies; · possibility of rearranging concept positions on the graph; · possibility of reparenting concepts.
9) Work on one dimension at a time, completing it as exhaustively as possible before moving on to the next dimension.	· CODE's masking facility allows users to focus on one dimension at a time. When the first dimension has been completed, the mask can be removed, and then used to focus on the second dimension, etc.

Table 7.2. A summary of the technical complement to the initial guidelines for *representing* multidimensionality. The left-hand column lists the five initial guidelines for representing multidimensionality (cf. section 4.3). The right-hand column lists the ways in which the CODE system can facilitate this representation.

PART IV

CHAPTER EIGHT

APPLICATION OF THE PROPOSED GUIDELINES

8.0 SUBJECT MATTER FOR THE APPLICATION OF THE GUIDELINES

In this chapter, we will subject our proposed guidelines (cf. tables 7.1 and 7.2) to a modest test by applying them to the construction of a small knowledge base on a subfield of hypertext, namely hypertext links. When looking for a subject field, we had the following six criteria in mind.

1) The subject field should be a modest size (i.e., containing 15–30 concepts). This obviously meant that it would have to be a fairly restricted subfield of a broader field.
2) The subject field should deal with concepts that are different in nature from those used in the examples in previous chapters. Since those examples focused primarily on concrete objects (e.g. optical storage media, vehicles, wine), we decided to concentrate on a field that contained more abstract concepts to see if the guidelines could be applied in this situation.
3) Even though we wanted to focus on a subject field whose concepts were abstract in nature (cf. 2 above), we felt that the subject field should still be within our general area of expertise and interest (i.e. computers), and still somewhat related to the original field that was used to develop the guidelines (i.e. computer hardware/media). We felt it would be best to test the guidelines gradually, choosing subject fields that were increasingly different, instead of jumping directly into a field that was dramatically different from the original field, such as law or fine arts.

4) The subject field should be dominated by generic–specific relations. As we decided at the outset of the thesis to restrict our investigations to multidimensionality within generic–specific relations, we did not want a subject field that was dominated by other types of relations (e.g. the subfield of optical disc drives is dominated by part–whole relations).

5) As terminologists traditionally consult with subject field experts when conducting a terminological research project, we wanted to be sure that we could locate a subject-field expert who was willing to help us.

6) Finally, we wanted to ensure that the subject field was interesting from a terminological point of view. In other words, we did not wish to "reinvent the wheel" by working on a subject field that has already been explored in some depth; rather, we wanted to work on an emerging field.

With these criteria in mind, we decided to work on a subfield of hypertext, namely hypertext links. This seemed to be an ideal choice because it met all of the specified criteria. Firstly, the subfield of links contains a limited number of concepts.[74] Secondly, hypertext links are not concrete objects; rather, they are *abstract* in nature. Thirdly, since the subfield is concerned with computer software, it falls within our general area of expertise and interest (i.e. computers), and it is still somewhat related to the original field that was used to develop the guidelines (i.e. computer hardware/media). Fourthly, the type of relations found between concepts in the subfield of hypertext links are generic–specific. Fifthly, we were able to locate a subject-field expert in the Artificial Intelligence Laboratory of the Department of Computer Science at the University of Ottawa, who was willing to help us. Lastly, hypertext is an

---

[74] Our knowledge base contains 20 concepts; however, as explained in section 8.2.10, the dimension based on the characteristic **product** is open-ended, hence more concepts can be added here.

emerging field for which little terminological research has been carried out. Currently, very few terms in the field of hypertext, and no terms in the subfield of hypertext links, are included in either TERMIUM III or the *Banque de terminologie du Québec (BTQ)*.[75]

8.1 A BRIEF INTRODUCTION TO THE SUBJECT MATTER

Hypertext is a polysemous term. In one sense, it can be seen as a tool: a type of computer software that mimics the human brain's ability to access information quickly and intuitively by association (Chen 1989a: 3). In another sense, it can be considered as the product created with this tool: a type of *non-sequential*, computerized document created using hypertext software (Nielsen 1990: 1).

For our present purpose, which is to understand links, we will consider *hypertext* in the second sense, that of a non-sequential document. Traditional documents generally contain ordered units of information, such as chapters in a book. These documents are intended to be read in an ordered fashion (i.e. from front to back).[76] In hypertext documents, units of information are called nodes, and they may be of varying "sizes" (e.g. a short text string, a larger region of text, or a complete file). Unlike traditional documents, hypertext documents do not impose a predetermined order in which the text is to be read. Rather, the nodes are electronically connected to one another by *links*. The number of links depends on the content of each node; some nodes are related to many others and will have many links, while other nodes serve only as destinations for links, but have no outgoing links of their own. At various points in the hypertext document, the reader is presented with several different options (i.e.

---

[75] TERMIUM III is the federal government terminology bank maintained by the Translation Bureau of the Department of the Secretary of State of Canada. The BTQ is the provincial government terminology bank maintained by the *Office de la langue française* of the Quebec government.

[76] Of course, reference books, such as dictionaries and encyclopedias, are an exception to this rule.

links to other nodes), and he or she determines which one to follow. Nodes and links are the foundation of hypertext; they are illustrated in figure 8-1.

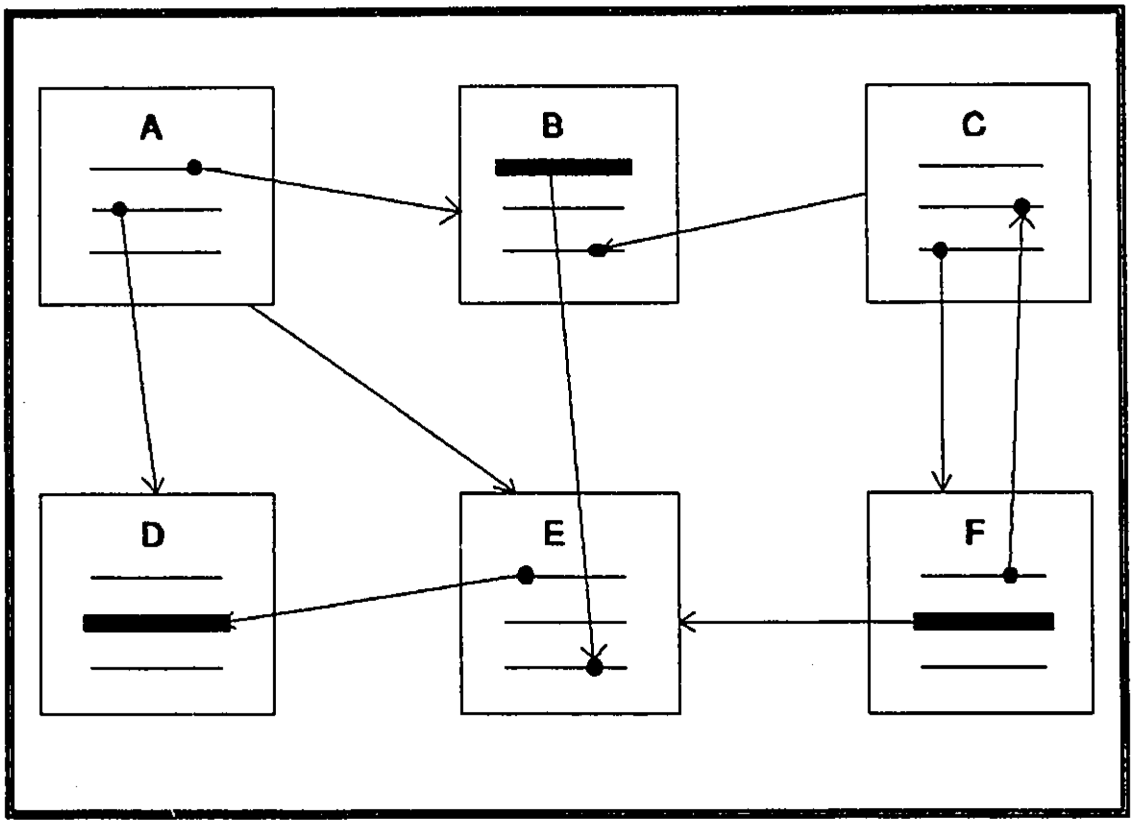


Figure 8-1. A simplified view of a small hypertext document. The nodes are of varying sizes: the dots represent text strings, the thick lines represent larger regions of text, and the boxes represent complete files. The arrows represent links. *Source: adapted from Nielsen (1990: 1).*

8.1.1 Hypertext links

A link is basically an electronic connection between two (or more) nodes. When users activate a link, they will be taken from the link's starting point (called the *anchor node*) to its ending point (called the *destination node*). As shown in figure 8-1, sometimes links have very specific anchors and destinations; they lead from a single lexical item or string within a node to another specific lexical item or string in another node. Meanwhile, other links have less specific anchors and destinations; they may merely point from a region or even a complete file to another region or file.

Some links have a predefined destination that can never be changed; they will always take the user from point A to point B. Other links do not have predetermined destinations as such; rather, their destinations are computed by the system based on information provided in the current or previous nodes. Moreover, some links can be traversed in only one direction, while others allow users to travel back and forth between nodes.

Generally, links appear on the screen in the form of a button or icon; however, some links are invisible to the user. The appearance of a link is related to its method of activation. To activate links that appear as buttons or icons, users are generally required to click on the button or icon. Invisible links are activated through a keystroke or a menu selection.

Some links are labelled in such a way that users know what their function will be before they activate the link. Other links have no distinctive label, and users must activate the link in order to see what it will do.

These main characteristics of links are summarized in table 8.1.

8.2 THE APPLICATION

In this section, we will describe in detail the step-by-step process that was used to apply the proposed guidelines. As mentioned in section 4.3, the nine guidelines, which are summarized in tables 7.1 and 7.2, cannot be followed in a strictly sequential fashion. A certain amount of backtracking or repetition of steps is necessary. For example, guideline 2) recommends that terminologists consider all possible ways of classifying a subject field *at each level of the classification*. Similarly, guideline 3) points out that terminologists should be on the lookout for indications that multidimensionality exists. These guidelines cannot be followed just once; terminologists must be prepared to repeat these steps as necessary.

Characteristic name[77]	Characteristic description
number of nodes connected	· whether the link connects two or more nodes.
specificity of anchor	· whether the link is anchored in a specific text string, a larger region of text, or a complete file.
specificity of destination	· whether the destination of the link is a specific text string, a larger region of text, or a complete file.
determination of destination	· whether the destination of the link is predefined by the system author, or computed by the system based on information provided in the current or previous nodes.
directionality	· whether the link can be traversed in one or both directions.
appearance	· whether the link appears on the screen as a button, an icon, or not at all.
method of activation	· whether the link is activated by a click, a keystroke, or a menu selection.
transparency of function	· whether the link is labelled so that the user is aware of its function before activating it.

Table 8.1. The main characteristics of hypertext links.

Because the guidelines cannot be followed in a strictly sequential order, the following description of the application is very generally organized according to the stages of a terminological research project as outlined in section 2.1.1, and references are made to specific guidelines (cf. tables 7.1 and 7.2) when appropriate.

[77] As discussed in section 3.2.2, assigning names to characteristics is not an easy task. Our reasons for choosing the names for many of these characteristics will be discussed in more detail in sections 8.2.4 and 8.2.8.

We should also mention that when creating the knowledge base, we followed the general methodological framework adopted within the COGNITERM Project,[78] and as such, we prepared full CDs containing both linguistic and conceptual information. However, since the focus of this thesis is multidimensionality, we will only describe those aspects of our work that are directly related to multidimensionality. The completed CDs for all the concepts in the knowledge base are found in the appendix.

8.2.1 Contacting the subject-field expert

Traditional terminology practice involves consultation with subject-field experts, and since our proposed guidelines are in keeping with traditional terminology practice, one of our first tasks before finalizing our choice of a subject field was to contact a subject-field expert. We contacted Mr. Timothy Lethbridge, a Ph.D. student in the Department of Computer Science at the University of Ottawa who has conducted hypertext-related research at both the masters and doctoral level, and he agreed to act as our expert. Mr. Lethbridge was extremely co-operative and assisted us at various stages of the project: we had four formal meetings with Mr. Lethbridge, and we contacted him informally on several other occasions. The first meeting took place before we actually began doing any research. At this meeting, Mr. Lethbridge gave us a general overview of the subject field and helped to direct our documentation search. We met with Mr. Lethbridge a second time after we had done some introductory reading and made a rough sketch of the concept system. At this point, Mr. Lethbridge helped us to identify, understand, and name the characteristics underlying the different dimensions of LINK that we had encountered. Our third meeting was held once we had conducted further research and established a potential ranking order, which Mr. Lethbridge

[78] For more details about the general methodological framework of the COGNITERM Project, please refer to Meyer, Skuce, Bowker and Eck 1992b.

confirmed. The final formal meeting was held during the latter stages of our research. At this meeting, Mr. Lethbridge helped us to identify and rank a new dimension that we had encountered.

8.2.2 Introductory reading and rough sketch of concept system

We began to familiarize ourselves with the subject field by doing some introductory reading. At our first meeting, Mr. Lethbridge recommended an introductory article by Jeffrey Conklin (1987), and a book by Jakob Nielsen[79] (1990). We also consulted a hypertext resource book by Franklin and Kinnell (1990).

None of these three documents contained a graphic representation (i.e. tree diagram) of the subfield. Therefore, in keeping with guideline 1), which recommends that terminologists work with a graphic representation of the concept system, we immediately began using CODE's graphing facility to sketch out the main divisions of the subject field that we encountered during our introductory reading.

We did not find the tables of contents of the introductory documents to be helpful in identifying different dimensions of the subject field (cf. section 4.1.1 (*Tables of contents*)); however, we did discover that each of the three authors contrasted different pairs of lexical items when referring to links (cf. section 4.1.1 (*Lexical items*)): Nielsen (1990: 3) contrasted *unidirectional* with *bidirectional links*, Franklin and Kinnell (1990: 258) *typed* with *untyped links*, and Conklin (1987: 33–34) *referential* with *organizational links*. Conklin (1987: 35) also made a separate reference to *cluster links*. As suggested in guideline 3), we took this variety of contrasting lexical pairs to be a possible indication that multidimensionality existed. In other words, we hypothesized that links could be classified in at least three different ways,

---

[79] In Mr. Lethbridge's opinion, Jakob Nielsen is "the most prolific author in the field of hypertext."

and we started to construct our representation of the concept system accordingly. First, on the graph, we created the concept LINK as this was to be the top concept of the subfield. Next, in the browser, we opened the CD for LINK and entered UNIDIRECTIONAL LINK and BIDIRECTIONAL LINK as kind (1), TYPED LINK and UNTYPED LINK as kind (2), and REFERENTIAL LINK and ORGANIZATIONAL LINK as kind (3). The order in which the kinds were entered was purely arbitrary, as it was not yet clear what the order of importance would be. Meanwhile, CLUSTER LINK was simply entered as a subconcept of LINK, as there was no evidence at this point that it belonged to a particular dimension. This is illustrated in figure 8–2a. Based on this information, CODE automatically generated a graph showing our initial representation of the concept system; this graph is illustrated in figure 8–2b. Guideline 6) suggests that all the dimensions on the graphic representation be clearly distinguished. The CODE system automatically distinguishes the different dimensions by numbering the k-links (one number per dimension) in accordance with the information entered in the kinds field of the browser. The numbered k-links are also shown in figure 8–2b.

8.2.3 Ranking the dimensions

Following guideline 5), we next considered how we should rank the dimensions (cf. 2.3.3). After brief consideration, we decided that we could not legitimately rank them at this point in the research since each had appeared in only one document and each contained the same number of terms. Therefore, we decided to leave the dimensions in their current arbitrary order, and to alter this order, if necessary, as we learned more about the subject field.

a)

cdName: link
super:
kinds: *(1) unidirectional link, bidirectional link*
(2) typed link, untyped link
(3) referential link, organizational link
subConcepts: cluster link

b)

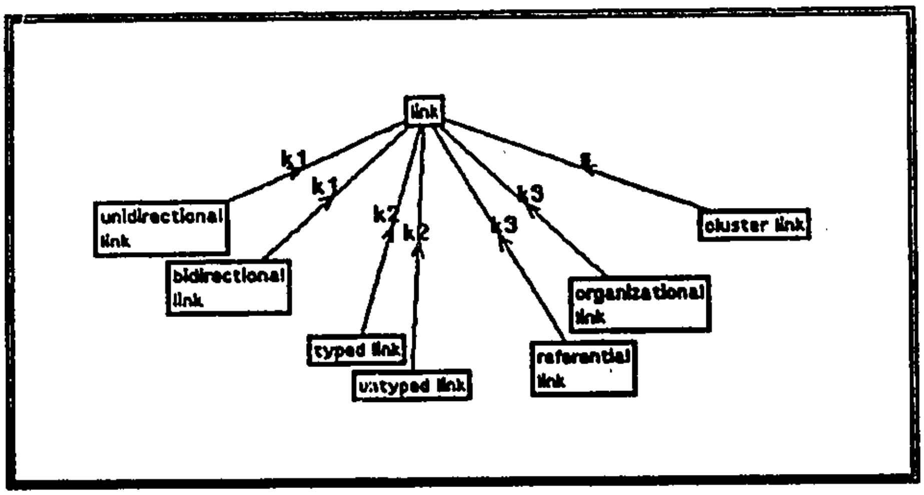


Figure 8-2. Initial classification of hypertext links produced using CODE during the introductory reading stage of the project. a) shows the entry of concepts in the kinds and subConcepts fields of the CD for LINK. b) shows the initial graphical representation of the subject field that was automatically generated by CODE based on the information entered in the browser. Note how the dimensions are distinguished by numbered k-links.

8.2.4 Identifying the characteristic underlying each dimension

Our next task was to identify and record the characteristic underlying each dimension, as recommended in guideline 7). It was pointed out in section 3.2.2 (*Problem 2*) that identifying and assigning names to characteristics is not a simple matter, and Mr. Lethbridge assisted us a great deal with this task.

We determined that UNIDIRECTIONAL LINK and BIDIRECTIONAL LINK had been classified on the basis of whether the link could be traversed in both directions, or only one. In the introductory literature, we did not find any established term to describe this characteristic, so following the COGNITERM rules of thumb for naming characteristics (cf. section 3.2.2 (*Problem 2*)), we decided that **directionality** would be a good name for this characteristic since it was both short and transparent. In order to be sure that users would understand what we meant by **directionality**, we included the following description in the characteristic description field of the facet pane (cf. section 5.1.2 (*Explicit recording of characteristics*)): The directionality of a link refers to whether the link can be traversed in only one direction or in both directions. Mr. Lethbridge approved of this characteristic name, and at a later point in our research, we actually found this name used as a term (Woodhead 1991: 118). We then entered ***D-link by directionality*** as a document characteristic in the CDs for UNIDIRECTIONAL LINK and BIDIRECTIONAL LINK (cf. section 5.1.3 (*Document characteristics*)).

In the next dimension, the characteristic underlying the classification of TYPED LINK and UNTYPED LINK was slightly harder to identify. Eventually, we determined that these two concepts are related to the question of whether users are aware of a link's function *before* they activate that link.[80] Once again, we could find no established term in the documentation to describe this characteristic; however, after discussing the characteristic with Mr. Lethbridge, and considering the COGNITERM rules of thumb for naming characteristics (cf. 3.2.2 (*Problem 2*)), we decided that **transparency of function** was a reasonable name, and we recorded it as the document characteristic ***D-link by transparency of function*** in the CDs for both TYPED LINK and UNTYPED LINK. We described the characteristic in the

[80] The difference between TYPED LINK and UNTYPED LINK is somewhat complicated; however, Mr. Lethbridge provided us with an explanation which we have summarized as follows. A typed link is marked in some way to indicate its function to the user. For example, links which always take the user from the current node to the reference for this node may always be labelled "Ref," or they may always appear as red buttons, etc. All links that perform the same function (i.e. taking the user to the reference for the current node) are labelled in exactly the same way; therefore, users are aware of what the function of the link is *before* they actually activate it. In contrast, users following an untyped link will not know the function of this link until *after* it has been activated; in other words, they cannot tell just by looking at the link what function it will perform.

characteristic description pane as follows: The transparency of function of a link refers to whether the link is labelled in such a way that the user is aware of its function *before* activating it.

Finally, we determined that in the third dimension, REFERENTIAL LINK and ORGANIZATIONAL LINK had been classified on the basis of whether or not the nodes were connected in a hierarchical fashion. Although we could find no established term for this characteristic, we felt that the name **type of node relation** was a good choice because it was relatively short and as transparent as possible. Accordingly, we recorded it as the document characteristic ***D-link by type of node relation*** in the CDs for REFERENTIAL LINK and ORGANIZATIONAL LINK. The characteristic description was: The type of node relation of a link refers to whether the link connects nodes in a hierarchical or non-hierarchical fashion.

8.2.5 Working on the concept LINK

Before we began gathering detailed information on the concepts belonging to the various dimensions of LINK, we worked on completing the CD for the concept LINK itself, following the standard COGNITERM methodology. Using the extensive bibliography found in Nielsen's book (1990), and another bibliography that we located on multimedia and hypermedia technologies (Chen 1989a), we managed to gather a reasonable selection of documentation. We scanned this documentation, and filled in the appropriate linguistic and conceptual characteristics of LINK as we learned more about the concept (cf. table 8.1 for a list of conceptual characteristics). Figure 8-3 shows the completed CD view for the concept LINK; note how the characteristics and values have been explicitly recorded in the easy-to-read format prescribed by CODE. These characteristics are automatically inherited to all the subordinate concepts of LINK, and all the subordinate concepts of these subordinate concepts, etc. (cf. section 5.1.2 (*Inheritance*)).

cdName: link
super: hypertext document
hasPropsOf:
kinds:
(1) unidirectional link, bidirectional link
(2) explicit link, implicit link
(3) typed link, untyped link
(4) cluster link, one-to-one link
(5) Guide link, Justus link
subConcepts:

SPECIAL
subject field: hypertext
creation date: 26 April 1992
last update date: 17 September 1992
done by: Lynne Bowker
figure:
status: nil
level:

DEFINITION
intensional definition: an electronic path or connection between two or more associated nodes. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: one of: two, more than two
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped

LINGUISTIC INFORMATION
English term: link ; COOK91: 10, FIDERIO88: 239, FRANKLIN90: 5, NIELSEN90: 2, BYERS87: 247, CONKLIN87: 33, RAYMOND88: 837, JONASSEN89A: 8, JONASSEN89B: 6, WATTERS92: 129, HAAVIND90: 42, CARR88: 280
English part of speech: noun
English textual support/cont1: "Various applications of the hypertext idea may differ in many ways. One feature they share is the ability to form "links" to connect objects on one screen with something else, usually on another screen. These "links" may not be visible to the user, but they are what holds a hypertext system together." ; COOK91: 10
English textual support/cont2: "In general, "links" are used to connect the nodes. A "link" is like an electronic footnote, endonte, or a parenthetical phrase. That is, just as footnotes and parenthetical phrases direct readers of printed material to related points or further topics for research, "links" connect you to associated text or ancillary information." ; FIDERIO88: 239
English textual support/cont3: "Button and "link" are complementary concepts: a button is a visible indication that two nodes intersect, and a "link" is the electronic connection between them." ; FRANKLIN90: 5
English collocation: to forge a link ; BYERS87: 250, CARR88: 280, HAAVIND90: 42
English collocation/textual support/cont1: "The key to implementing hypertext is the ability to constantly add new data and "forge new links"..." ; BYERS87: 250
English collocation/textual support/cont2: "In some hypertext systems, the linking process is so simple that any reader can cut a path through the text and "forge new links"." ; CARR88: 280

Figure 8–3. The CD view for the concept LINK. The CD name is in the upper left pane. The concept is defined in the Definition pane. Conceptual characteristics and their values are found in the Conceptual Information pane, while linguistic characteristics and their values are located in the Linguistic Information pane. References for characteristic values are given in short form (i.e. the author's last name and the year of publication) in each CD; full references are provided in a separate bibliographic CD called *References* (see appendix).

8.2.6 Working on the first dimension

Once we had completed the CD for the concept LINK as fully as possible, we moved on to consider its subordinate concepts. Adhering to guideline 9) which recommends that terminologists work on one dimension at a time, we used the mask to show only the dimension based on the characteristic **directionality**. The mask and resulting graph are shown in figure 8-4.

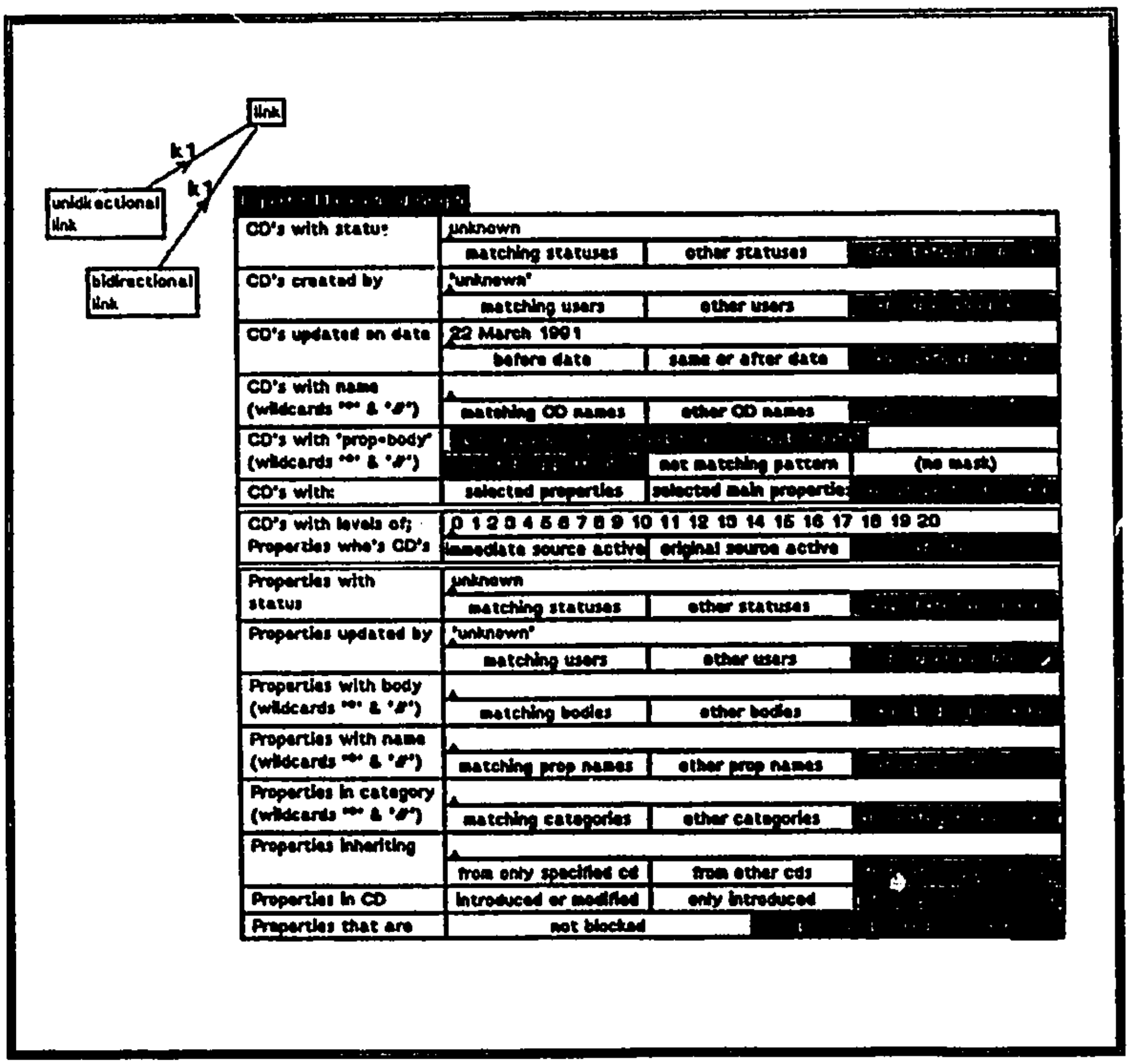


Figure 8-4. A CODE mask and the resulting graph showing only those concepts having the document characteristic *D-link by directionality.*

Next we chose the documentation to be scanned. We did not find any documents dealing exclusively with unidirectional and bidirectional links; however, this is not surprising since links themselves are already a very specialized area of the subject field with very few documents devoted exclusively to them. Nonetheless, the terms *unidirectional link* and *bidirectional link* did appear in some indexes.

We scanned the documentation and, based on our findings, we specialized the values of the characteristics[81] that had been inherited. We did not discover any completely new characteristics or subordinate concepts to be added to the concepts UNIDIRECTIONAL LINK or BIDIRECTIONAL LINK (cf. appendix for completed CDs).

8.2.7 Working on the second dimension

When we had completed the first dimension as exhaustively as possible, we began working on the second dimension.[82] We used the mask to show only those concepts having the characteristic ***D-link by transparency of function***, as illustrated in figure 8-5.

Not surprisingly, we did not find any documentation devoted to these specific kinds of links, but the terms *typed link* and *untyped link* did appear in some indexes. Once again, we scanned the documentation and specialized the values of inherited characteristics. We did not find any completely new characteristics or subordinate concepts to be added to the concepts TYPED LINK or UNTYPED LINK (cf. appendix for competed CDs).

[81] As previously mentioned, we will only detail the work that is directly related to multidimensionality; therefore, all references to characteristics are understood to be *conceptual* and not *linguistic* characteristics.

[82] Note that the second dimension is only arbitrarily second; a ranking order has not yet been established.

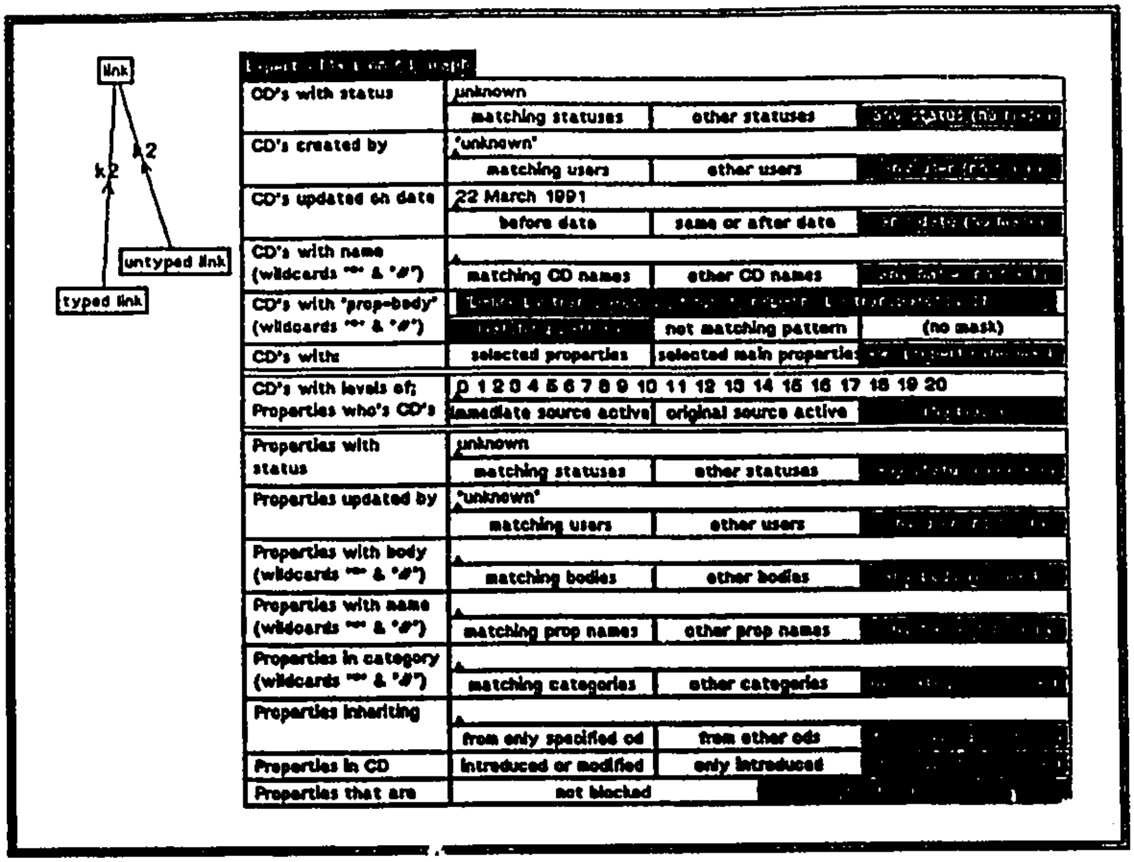


Figure 8–5. A CODE mask and the resulting graph showing only those concepts having the document characteristic *D-link by transparency of function*.

8.2.8 Working on the third dimension and discovering a new dimension

After completing the second dimension as fully as possible, we used the mask once again, this time to show only those concepts having the characteristic ***D-link by type of node relation***, as shown in figure 8–6.

As we researched this dimension, we observed that Conklin (1987: 33) states "There are two methods for *explicitly* [italics mine] linking two points in hypertext – by reference and by organization." Further reading (e.g. Nielsen 1990: 107; Irler 1990: 266; Carr 1988: 280) revealed the existence of another pair of contrasting lexical items: EXPLICIT LINK and

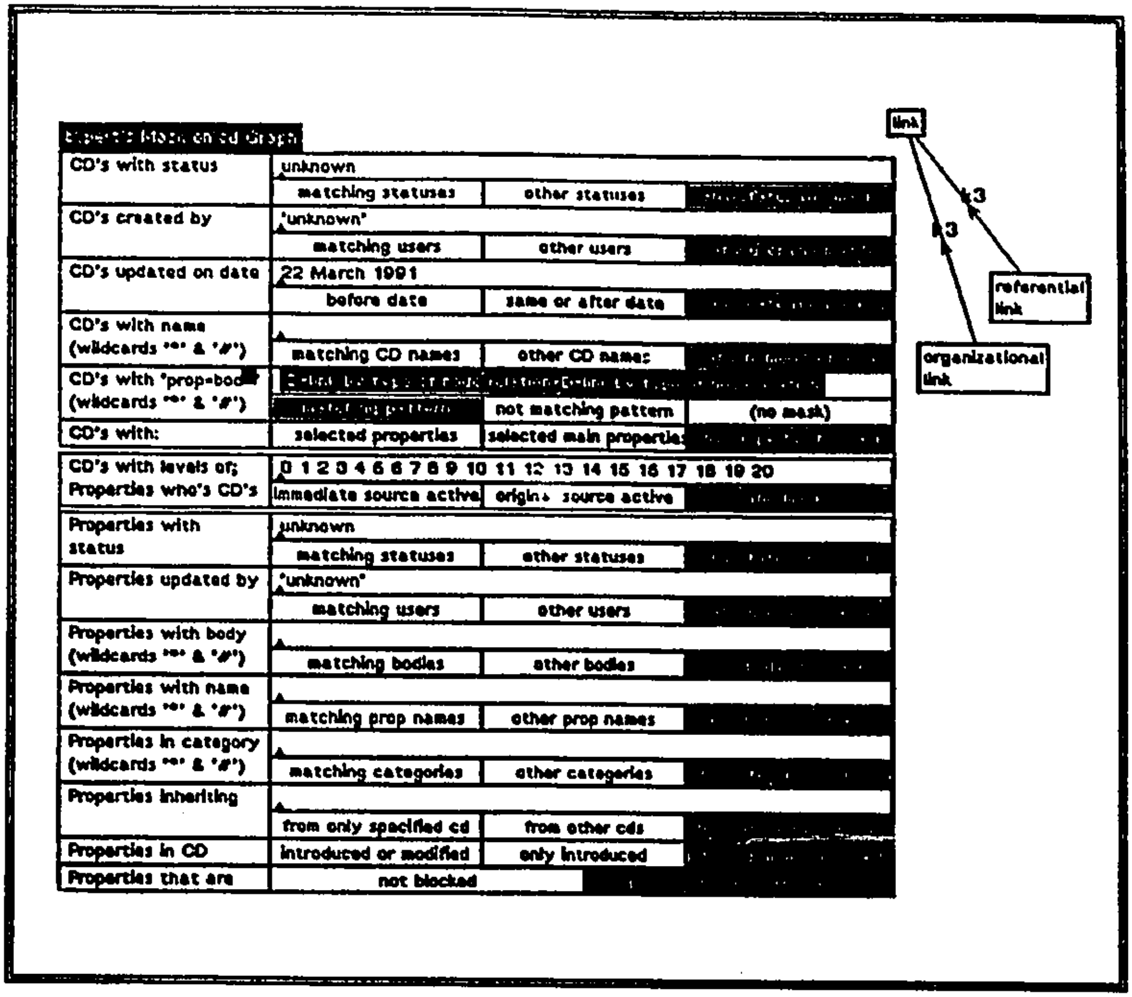


Figure 8–6. A CODE mask and the resulting graph showing only those concepts having the document characteristic ***D–link by type of node relation.***

IMPLICIT LINK, which are both subordinate concepts of LINK. In researching EXPLICIT LINK and IMPLICIT LINK, we determined that they are classified on the basis of whether the destination of the link is predefined by the system author and is thus invariable, or whether it is computed by the system based on information provided in the current or previous nodes, and thus can vary from usage to usage. This was another difficult characteristic to name, and we considered several options[83] before finally deciding on the name **determination of destination**. We entered **D-link by determination of destination** as a Document characteristic in the CDs for EXPLICIT LINK and IMPLICIT LINK with the following characteristic description: The determination of destination of a link refers to whether the destination of a link is predefined by the system author, or computed by the system based on information provided in the current or previous nodes. We then returned to the CD for LINK and there entered **determination of destination** as a conceptual characteristic so that it would inherit to all the subordinate concepts of LINK.

We next turned our attention back to Conklin's classification. Conklin's statement "There are two methods for *explicitly* [italics mine] linking two points in hypertext...." (1987: 33) made us suspect that perhaps Conklin's classification was not complete. It initially appeared that Conklin might have made a classification error: he did not specifically reflect the classification of LINK into EXPLICIT LINK and IMPLICIT LINK (according to the characteristic **determination of destination**), but rather, he immediately classified LINK into REFERENTIAL LINK and ORGANIZATIONAL LINK (which are both types of EXPLICIT LINKS) on the basis of the characteristic **type of node relation**. In an informal discussion, Mr. Lethbridge said that Conklin may not have made an error, but rather a deliberate oversight when he neglected to reflect the classification of LINK into EXPLICIT LINK and IMPLICIT LINK. Mr. Lethbridge explained that although hypertext technology is rapidly evolving, even today explicit links are more common than implicit links; therefore, in 1987, when Conklin wrote his article, it is possible that he considered implicit links to be so rare that he did not feel it necessary to take them into consideration at that time. This is

[83] One characteristic name that we considered was "variability of destination"; however, this choice was rejected by Mr. Lethbridge because, in his opinion, the characteristic is not really whether the link can have different destinations, but how the destination is determined. A second name that we considered was "computedness." This characteristic name was Mr. Lethbridge's initial suggestion; however, we felt that it was not transparent enough for non-experts, and we were slightly concerned by the fact that it was not an accepted English word.

entirely conceivable since, as discussed in section 1.2.1, changes in the reality of a phenomenon can result in multidimensionality. Mr. Lethbridge also pointed out that implicit links are more complex than explicit links, and since Conklin's article is intended to provide an introduction to hypertext, it is possible that he chose to focus on the more common and tangible explicit links, rather than overload learners with information about the rarer and more complicated implicit links. As noted in section 1.2.1, purpose is another possible cause of multidimensionality.

Given the new information that we had uncovered, we revised our representation by adding EXPLICIT LINK and IMPLICIT LINK as kind (3) in the CD for LINK, and making REFERENTIAL LINK and ORGANIZATIONAL LINK subconcepts of EXPLICIT LINK. These changes were automatically reflected on the graph and in the CD views. Figures 8–7a and b show the changes made in the CDs, while figure 8–7c shows the modified concept system representation.

As we researched explicit and implicit links, we discovered that this dimension came up quite frequently in recent documentation (e.g. Nielsen 1990: 107; Irler 1990: 266; Watters 1992: 70). We decided that although this classification was less common than the dimension based on **directionality** (i.e. UNIDIRECTIONAL LINK vs. BIDIRECTIONAL LINK), it was more common than the dimension based on **transparency of function** (i.e. TYPED LINK vs. UNTYPED LINK). At our third formal meeting with Mr. Lethbridge, he confirmed our new ranking order. Consequently, we rearranged the order of the concepts in the kinds field so that the dimension based on the characteristic **determination of destination** (i.e. EXPLICIT LINK vs. IMPLICIT LINK became kind (2). This change was automatically reflected on the graph (the k–links were renumbered), and in the CD views. Figure 8–8 shows the new ranking order of the concepts in the browser and on the graph.

a)

cdName: link
super:
kinds: (1) unidirectional link, bidirectional link
(2) typed link, untyped link
(3) explicit link, implicit link
subConcepts: cluster link

b)

cdName: explicit link
super: link
kinds:
subConcepts: referential link, organizational link

c)

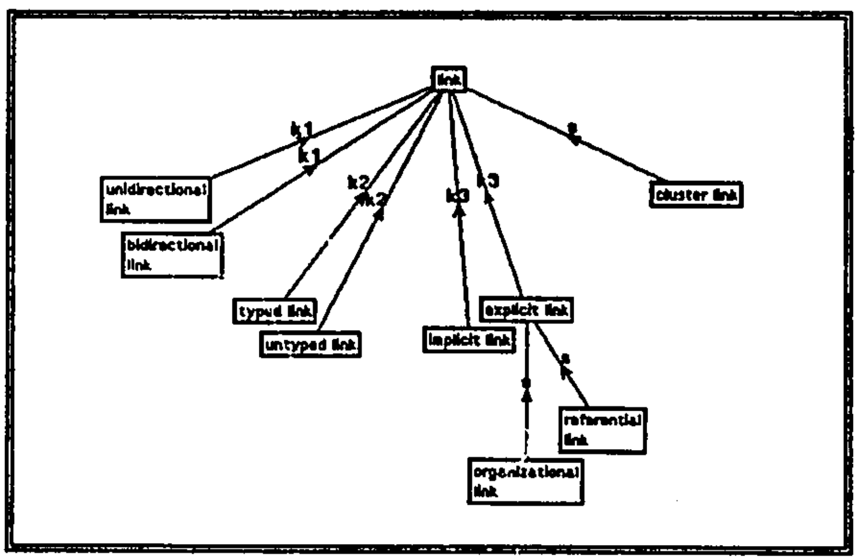


Figure 8-7. A modified classification of LINK. a) shows how the concepts EXPLICIT LINK and IMPLICIT LINK were added to the kinds field of the CD for LINK. b) shows how the concepts REFERENTIAL LINK and ORGANIZATIONAL LINK were added to the subConcepts field of the CD for EXPLICIT LINK. c) shows the corresponding modifications that were automatically reflected on the graph.

a)

cdName: link
super:
kinds: ***(1) unidirectional link, bidirectional link***
(2) explicit link, implicit link
(3) typed link, untyped link
subConcepts: cluster link

b)

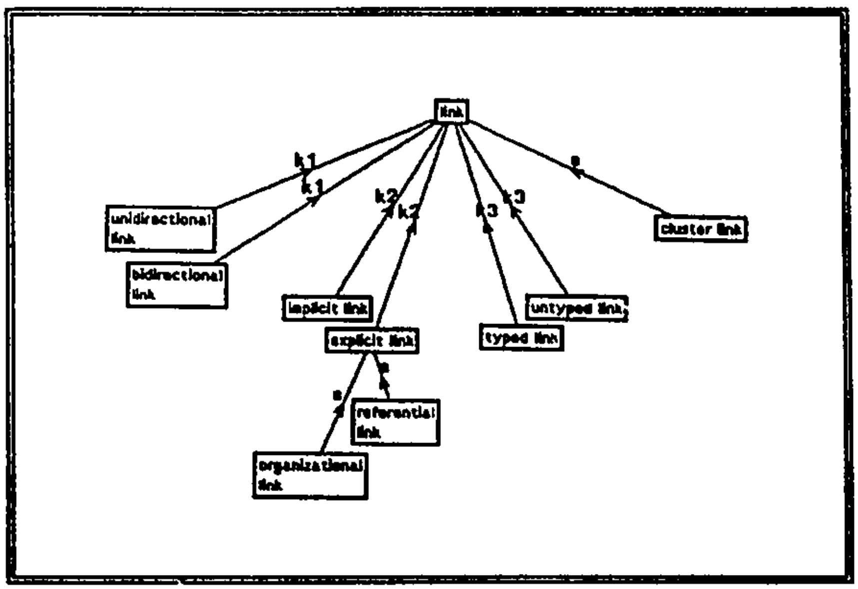


Figure 8-8. The modified ranking order for the dimensions of LINK. a) shows how the concepts were rearranged in the kinds field of the browser, while b) shows how these changes were automatically reflected on the graph.

As we scanned the documentation, we specialized the values of inherited characteristics for the concepts EXPLICIT LINK and IMPLICIT LINK. We also specialized the values of the inherited characteristics for the subordinate concepts of EXPLICIT LINK, namely REFERENTIAL LINK and ORGANIZATIONAL LINK. We did not find any completely new characteristics or subordinate concepts to be added to the concepts IMPLICIT LINK, REFERENTIAL LINK, or ORGANIZATIONAL LINK, nor any additional subordinate concepts to be added to the concept EXPLICIT LINK (cf. appendix for completed CDs).

8.2.9 Working on the subordinate concept CLUSTER LINK

Once we had completed the three dimensions to the best of our ability, we turned our attention to the subordinate concept, CLUSTER LINK, which had been somewhat neglected to this point. We were suspicious of the fact that it did not appear in opposition to another term, as had all the other links thus far encountered, and so we raised this concern at our fourth meeting with our subject-field expert. Mr. Lethbridge said that the overwhelming majority of links connect only two nodes. Therefore, in his opinion, because this is the normal situation, reference is only made to this fact when these links are discussed in opposition to the much rarer links connecting multiple nodes (i.e. cluster links). Mr. Lethbridge said that he was not aware of an established term used to refer specifically to links that connect only two nodes; rather, he said that this is generally understood in the term *link*. Mr. Lethbridge suggested that if we wished to assign a specific term to this concept, he would recommend the term *one-to-one link*.

As a general rule, term banks only contain true terms; however, within the general framework of the COGNITERM methodology, provisions are made for creating *dummy concepts*[84] to represent concepts that are not *lexicalized* (i.e. not designated by a specific term) when the inclusion of such concepts in the concept system is linguistically useful (i.e. if these concepts greatly assist in understanding the subject field, or if they have subordinate concepts that are lexicalized). At this point, we were unsure whether the creation of a dummy concept for ONE-TO-ONE LINK would be linguistically useful. We therefore decided to create the concept, and to mark the node with a dashed line to indicate our uncertainty. The decision as to whether the concept would remain in the term bank could be made after more research had been carried out.

[84] For more details regarding the creation of dummy concepts within the framework of the COGNITERM Project, please refer to Meyer, Skuce, Bowker and Eck 1992b.

Mr. Lethbridge then pointed out that by adding the concept ONE-TO-ONE LINK to the classification, we were actually creating a fourth dimension based on the characteristic **number of nodes connected**. He recommended that this dimension be placed last in the ranking order since, of the four dimensions we had identified, it was the least common way of distinguishing links. Therefore, we returned to the CD for LINK, where we first added the conceptual characteristic **number of nodes connected** so that it would be inherited to all the subordinate concepts of LINK. Next, we removed CLUSTER LINK from the subConcepts field, and added CLUSTER LINK and ONE-TO-ONE LINK as kind (4). As always, the change was automatically reflected on the graph and in the CD views. Figure 8-9 shows the modifications in the kinds field of the CD for LINK, and the updated graph.

a)

cdName: link
super:
kinds: ***(1) unidirectional link, bidirectional link***
(2) explicit link, implicit link
(3) typed link, untyped link
(4) cluster link, one-to-one link
subConcepts:

b)

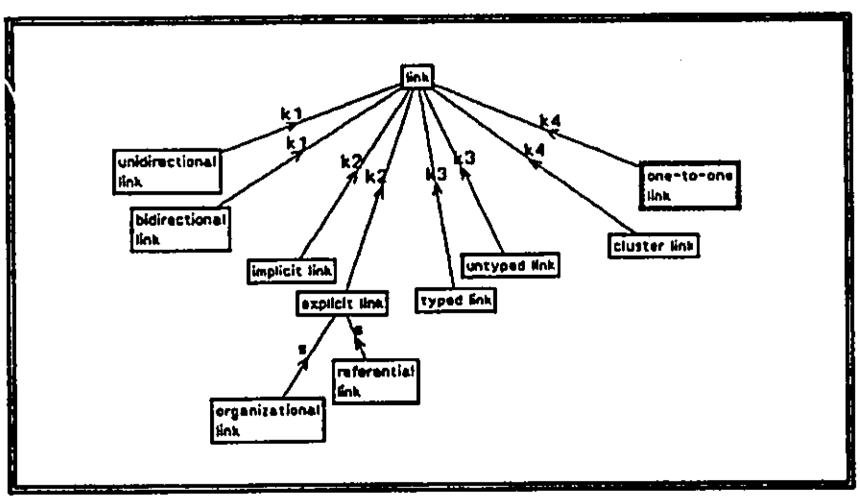


Figure 8-9. CLUSTER LINK and ONE-TO-ONE LINK added as a fourth dimension of the concept LINK. a) shows how CLUSTER LINK was removed from the subConcepts field and added to the kinds field as kind (4) along with ONE-TO-ONE LINK. b) shows how the fourth dimension is automatically added to the graph and the concepts marked with a k-link numbered 4. Note that the node for the concept ONE-TO-ONE LINK is marked with a dashed line to indicate our uncertainty.

As we researched this dimension, we specialized the appropriate characteristic values, and we also found two subconcepts for CLUSTER LINK: SEQUENTIAL LINK and TAXONOMIC LINK (DeRose 1989: 252–3). As illustrated in figure 8–10, we added these concepts to the subConcepts field in the CD for CLUSTER LINK, and the changes were reflected on the graph. We determined that these two concepts differed based on the order in which they connected a series of nodes: sequential links connect an ordered series of nodes, while taxonomic links connect multiple nodes in a non-sequential order. Therefore, we added the conceptual characteristic **node connection order** to the CD for CLUSTER LINK, and it inherited to its subordinate concepts, where we specialized the value for this characteristic (cf. appendix for completed CDs).

a)

cdName: cluster link
super: link
kinds:
subConcepts: sequential link, taxonomic link

b)

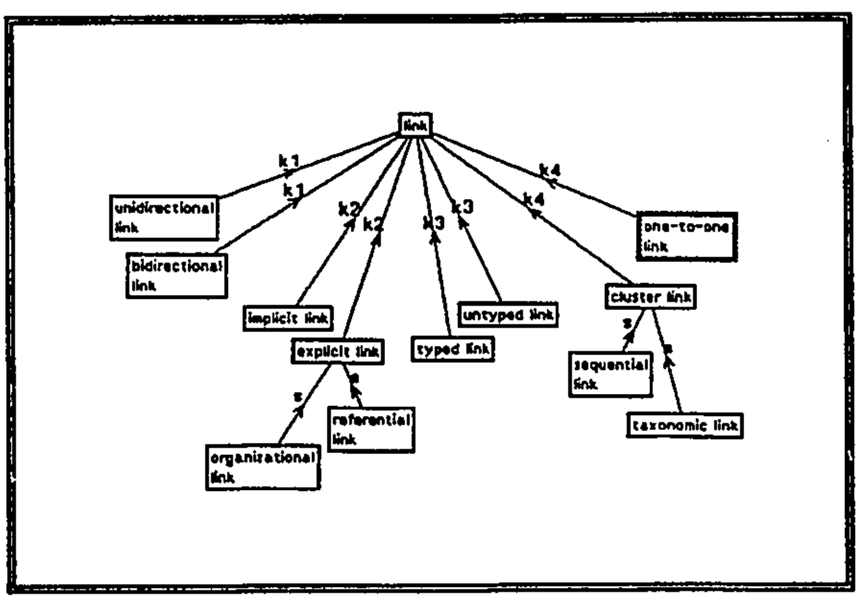


Figure 8-10. SEQUENTIAL LINK and TAXONOMIC LINK added as subconcepts of CLUSTER LINK. a) shows how these concepts were added to the subConcepts field of the browser, while b) shows how the changes are automatically reflected on the graph.

8.2.10 New problems encountered

As we scanned the documentation during our research in the various dimensions, we became aware that many off-the-shelf hypertext systems (e.g. Guide™, Justus™) had their own specific links designated by specific terms (e.g. *replacement link*, *concept link*). We felt it was important to include these terms in the TKB because they appear in the documentation that accompanies the off-the-shelf systems (i.e. user manuals), and they also appear in documentation about other products that are built using the off-the-shelf hypertext systems.[85] However, we were not yet able to determine the place of these specific links in the classification of the subject field.

Faced with this significant problem, we asked Mr. Lethbridge, during an informal discussion, how these terms could be included in the concept system representation. Mr. Lethbridge suggested that they could be introduced as subordinate concepts in another dimension of LINK based on the characteristic **product**. In order to create this new dimension, we were required to create dummy concepts (cf. section 8.2.9) representing all the links belonging to particular hypertext systems. These dummy concepts, which appear on the graph as rounded nodes,[86] are not actual terms themselves, but they make it possible to include specialized terms in a logical place in the classification. Therefore, in the CD for LINK, we added the dummy concepts GUIDE LINK (representing links from the Guide hypertext system) and JUSTUS LINK (representing links from the Justus hypertext system)[87] as kind (5). These additions were automatically updated on the graph, as illustrated in figure 8-11.

---

[85] Hypertext systems are often used to build other applications, and the names of the links used in the original hypertext system are likely to be retained in the new applications.

[86] Rounded nodes are used to indicate dummy concepts that are considered to be linguistically useful.

[87] Readers may wonder why dummy concepts have not been created for other more popular hypertext systems such as HyperCard or NoteCards. The reason is that many hypertext systems do not come with predefined links, but rather, users can create and name their own links.

a)

cdName: link
super:
kinds: (1) unidirectional link, bidirectional link
(2) explicit link, implicit link
(3) typed link, untyped link
(4) cluster link, one-to-one link
(5) Guide link, Justus link
subConcepts:

b)

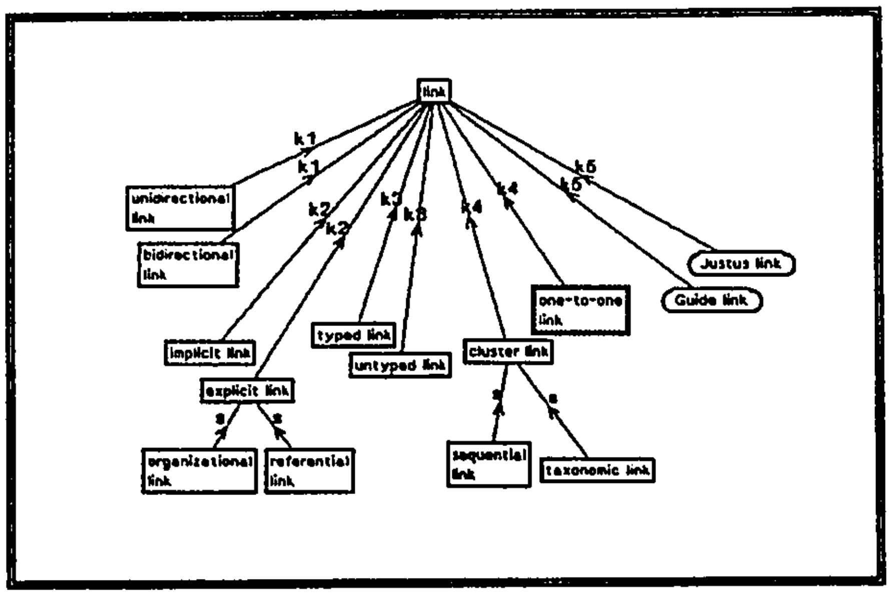


Figure 8-11. GUIDE LINK and JUSTUS LINK added as a fifth dimension of the concept LINK. a) shows how GUIDE LINK and JUSTUS LINK were added to the kinds field as kind (5). b) shows how the fifth dimension is automatically added to the graph and the concepts marked with a k-link numbered 5. Note how the dummy concepts are indicated with rounded nodes.

We were then able to add the specialized terms REPLACEMENT LINK, REFERENCE LINK, NOTE LINK, ASSOCIATION LINK and CONCEPT LINK as subordinate concepts of the appropriate dummy concepts, as shown in figure 8-12.

a)

cdName: Guide link
super: link
kinds:
subConcepts: replacement link, note link, reference link

b)

cdName: Justus link
super: link
kinds:
subConcepts: association link, concept link

c)

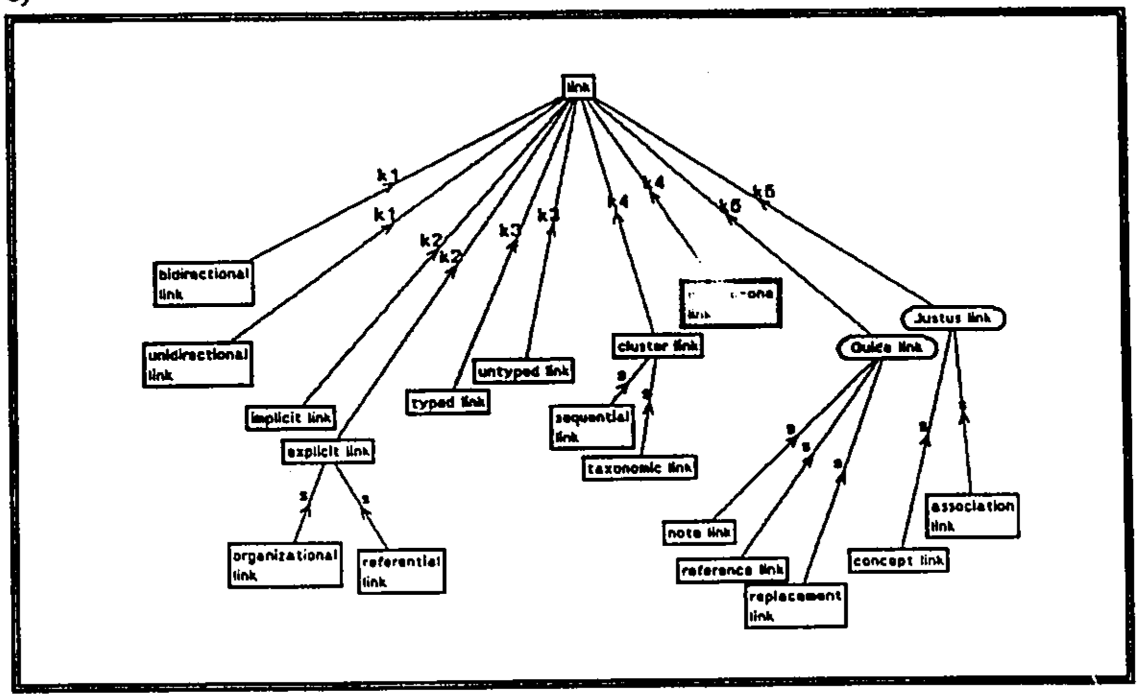


Figure 8–12. Specialized links added as subordinate concepts of dummy concepts. a) and b) show how the concepts are added as subConcepts in the browsers of the appropriate CDs, while c) shows the updated graph.

As we researched the above specialized links, we discovered that they also inherited characteristics from superordinate concepts in other dimensions (cf. section 5.1.3 (*Multiple inheritance*)). For example, we discovered that NOTE LINK, in addition to being a subordinate concept of GUIDE LINK in the dimension based on the characteristic **product**, also has the following characteristic values:

directionality: ***unidirectional*** (LETHBRIDGE 1992)
determination of destination: ***predefined*** (LETHBRIDGE 1992)
type of node relation: ***hierarchical*** (FIDERIO 1988: 242)
transparency of function: ***typed*** (JONASSEN 1989a: 79)
number of nodes connected: ***two*** (CONKLIN 1987: 32)

In examining these characteristic values, we determined that the concept NOTE LINK also had other superordinate concepts in other dimensions: UNIDIRECTIONAL LINK (in the dimension based on the characteristic **directionality**), REFERENTIAL LINK (one of the subordinate concepts of EXPLICIT LINK in the dimension based on the characteristic **determination of destination**), TYPED LINK (in the dimension based on the characteristic **transparency of function**), and ONE-TO-ONE LINK (in the dimension based on the characteristic **number of nodes connected**).

We made similar discoveries for the concepts REFERENCE LINK, REPLACEMENT LINK, ASSOCIATION LINK, and CONCEPT LINK, and we added the specialized links as subordinate concepts of the appropriate superordinate concepts. Figure 8–13 shows a graph with the specialized links inheriting from superordinate concepts in several dimensions.

We should also point out that the discovery that the concept ONE-TO-ONE LINK does indeed have lexicalized subordinate concepts (i.e. NOTE LINK, REFERENCE LINK, REPLACEMENT LINK and ASSOCIATION LINK) prompted us to decide that it was a linguistically useful concept (cf. section 8.2.9). Therefore, we decided to keep the concept ONE-TO-ONE LINK in the knowledge base, and we changed its node shape from a dashed line (which had indicated uncertainty) to a rounded node (which indicates that it is a linguistically useful dummy concept).

This final[88] representation of the concept system for the subfield of hypertext links, illustrated in figure 8–13, was approved by our subject field expert, Mr. Lethbridge.

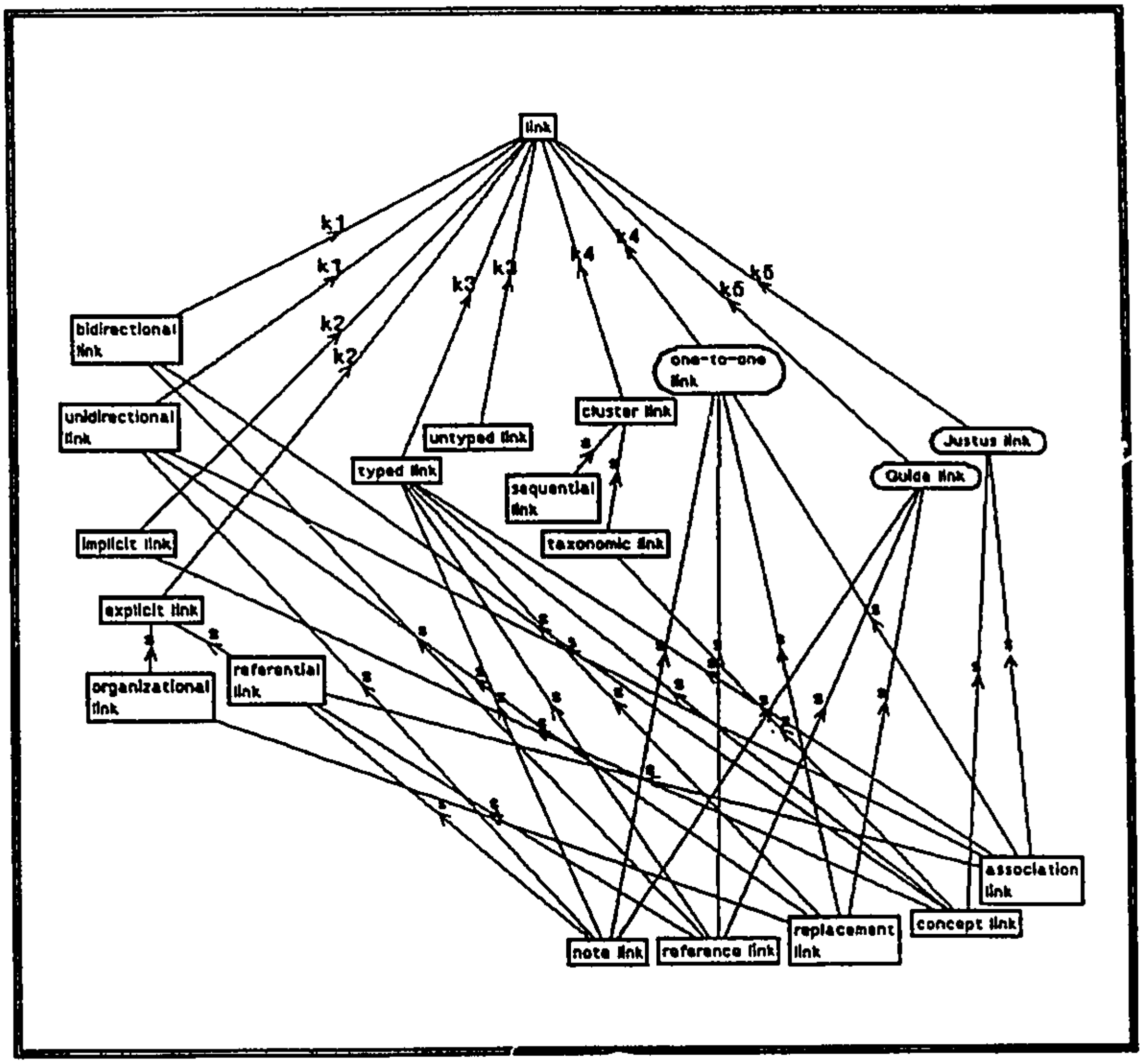


Figure 8–13. A final representation of the concept system for the subfield hypertext links. The specialized links have multiple superordinate concepts from different dimensions.

[88] We use the term *final* here to mean 'our final representation.' In actual fact, the representation of the subfield is not complete because the dimension based on the characteristic **product** (represented by k5 on the graph) is actually an open-ended dimension: dummy concepts can be created for all existing hypertext systems, which may in turn have specialized links.

CHAPTER NINE

CONCLUDING REMARKS

9.1 GENERAL REMARKS

The objectives of this thesis have been to 1) *develop* and 2) *apply* a set of guidelines for handling multidimensionality in a TKB. We began by gaining a basic understanding of the fundamental principles of classification, and the basic concept of multidimensionality (Chapter 1). We then proceeded to examine the classification techniques used in the fields of terminology, biology, and library science, to see if any of these techniques could be incorporated into our set of guidelines (Chapter 2). Next, we conducted some practical experimentation using an extended example on optical storage media, and identified some of the classification problems that terminologists are faced with in the course of their work (Chapter 3). Using the information gleaned from this research and experimentation, we developed an initial (non-technical) set of guidelines for handling (i.e. recognizing and representing) multidimensionality. Our next step was to explore the potential of the CODE system, a knowledge engineering tool, for helping terminologists handle multidimensionality (Chapter 5). We continued our experimentation with the extended example that was begun in Chapter 3, to see how the CODE system could be used to facilitate the handling of multidimensionality (Chapter 6). With the information culled from our technical research and experimentation, we developed a technical complement to the initial guidelines (Chapter 7). Finally, we applied the proposed guidelines to the construction of a modest TKB on the subfield hypertext links (Chapter 8).

In this final chapter, we will first discuss some of the issues raised during the application of our proposed guidelines. We will also make some general recommendations for technical modifications to the CODE system that will enable it to better handle multidimensionality in future versions. Finally, we will consider some possible areas for future research in multidimensionality.

9.2 ISSUES RAISED DURING THE APPLICATION

The application of the proposed guidelines (cf. tables 7.1 and 7.2) raised some interesting points of discussion, including the non-sequential nature of the guidelines, the dynamic and somewhat arbitrary nature of multidimensionality, and the role of subject-field experts in multidimensional classification.

9.2.1 The non-sequential nature of the guidelines

It was pointed out in section 8.2 that the proposed set of nine guidelines cannot be followed in a strictly sequential fashion. Some of the guidelines should be adhered to continually, such as guideline 1), which advises terminologists to work with a graphic representation of the concept system. Other guidelines may only need to be followed once, such as guideline 6), which suggests that terminologists distinguish the different dimensions on the graphic representation. Still other guidelines may need to be repeated on different occasions throughout the project, such as guideline 2), which recommends that terminologists consider all possible ways of classifying a subject field at *all* levels of the classifications. We should also point out that there may be cases (i.e. certain subject fields) where some of the guidelines do not apply at all. For example, in our documentation on the subfield of hypertext links, we did not encounter any instances where the repetition of concepts indicated the existence of multidimensionality (guideline 3)).

Basically, terminologists should regard our proposed set of guidelines as something to be kept in mind at all times, and to be applied whenever appropriate. Terminologists can follow through the traditional stages for conducting a thematic research project (cf. section 2.1.1), applying the appropriate guidelines for recognizing and representing multidimensionality as the need arises.

9.2.2 The dynamic and arbitrary nature of multidimensionality

Throughout the application of our proposed guidelines, it became evident that the phenomenon of multidimensionality was both dynamic and arbitrary in nature. With regard to dynamics, Conklin (1987: 32) describes the thinking process in the following words:

> The thinking process does not build new ideas one at a time, starting with nothing and turning out each idea as a finished pearl. Thinking rather seems to proceed on several fronts at once, developing and rejecting ideas at different levels and on different points in parallel, each idea depending on and contributing to the others.

This description also applies very well to the way in which terminologists learn about a new subject field. As pointed out by Sager (1990: 13), "Knowledge structures are not absolute entities but reflect the current state of knowledge of an individual or group of specialists." Sager (1990: 14) also states that "...conceptual systems are relatively fluid entities constantly undergoing change...." In other words, as terminologists do more and more reading and confer with experts, their understanding of the subject field grows and changes, and their representation of the concept system should be revised accordingly (cf. guideline 8)).

The terminologists' understanding of a multidimensional representation is part of their understanding of the subject field, hence the dimensions will change over the course of a project. For example, in our investigation of multidimensionality in the subfield of hypertext links, we found that we changed the ranking order of the dimensions (cf. section 8.2.8), we added new dimensions (cf. sections 8.2.8, 8.2.9, and 8.2.10), and we removed what we initially mistook for a dimension (cf. section 8.2.8).

In addition to changing as terminologists' understanding of the field increases, the dimensions will change according to the purpose of the classification (cf. section 1.2.1 (*Purpose*)) or as phenomena evolve (cf. section 1.2.1 (*Changes in the reality of a phenomenon*)). Conklin's classification of links reflected both these factors (cf. section 8.2.8).

As well as being dynamic, the phenomenon of multidimensionality is also somewhat arbitrary in nature; one could almost say that multidimensionality is itself multidimensional. There is no absolutely correct way of structuring knowledge; there are merely ways that are helpful or useful. For example, referring back to our classification of hypertext links, some people may not agree with our inclusion of the dummy concepts GUIDE LINK and JUSTUS LINK (and other kinds of system links) as a fifth dimension of LINK based on the characteristic **product** (cf. section 8.2.10). However, we felt that the creation of these dummy concepts clarified the placement of some more specialized links (i.e. REFERENCE LINK, NOTE LINK, REPLACEMENT LINK, ASSOCIATION LINK, and CONCEPT LINK) in the concept system.

The very fact that dimensions can be ranked is another indication of the arbitrariness of multidimensionality. While some dimensions, such as that based on the characteristic **directionality** (cf. section 8.2.6), were very obvious and were noted in a variety of documents, other dimensions, such as that based on **number of nodes connected** (cf. section 8.2.9), were less obvious. In fact, one of the concepts in this dimension (i.e. ONE–TO–ONE LINK) was not even lexicalized.

9.2.3 The role of subject–field experts in multidimensional classification

A third interesting issue that was raised during the application of our guidelines was the role of subject–field experts when dealing with the phenomenon of multidimensionality. We found that our hypertext expert, Mr. Lethbridge, was very helpful in clarifying the basic concepts of the subject field (cf. section 8.2.2), directing our documentation search (cf. 8.2.2), and helping us to name characteristics (cf. section 8.2.4). With regard to multidimensionality, we found that although Mr. Lethbridge did not find it very easy to spontaneously name the various dimensions of the concept LINK, he was able to confidently verify or reject our findings when presented with the results of our research. This would seem to indicate that experts may sometimes find it easier to critique dimension information, than to identify or create dimensions themselves. This may be because experts often work with a specific purpose in mind, and hence tend to be focused on a single dimension. Therefore, terminologists would probably be wise to turn to experts for verification of dimension information, but not necessarily depend on them to initially suggest different ways of classifying a subject field. As far as possible, terminologists should also try to consult a number of subject-field experts. In retrospect, we would have been wise to do this also.

9.3 RECOMMENDATIONS FOR TECHNICAL MODIFICATIONS TO CODE

As mentioned in section 5.1, the CODE system is generic by design. As pointed out by Skuce (1992a: 2), "We have preferred to develop a general purpose system, which is very useful as such, but which could easily be further specialized for some particular application if necessary." There are several modifications that could be made in order to tailor CODE for handling multidimensionality in a TKB.

A simple improvement to the CODE system would be to change the name of the *kinds* field to the *dimensions* field, and have the corresponding links on the graph show up as *d-links* instead of *k-links*; in this way, the terminology would be more transparent, and more consistent with the terminology used in the field of terminology.

Another possible improvement to CODE that would make it easier for users to quickly distinguish the different dimensions on the visual representation of the concept system would be to use different coloured lines to indicate the different dimensions. For example, as shown in figure 9-1, the concepts in the first dimension of LINK are joined to their superordinate concept with black lines, while those in the second dimension are represented with blue lines, those in the third with red lines, those in the fourth with green lines, and those in the fifth with purple lines. Of course, this would require a colour monitor, which not all people have.

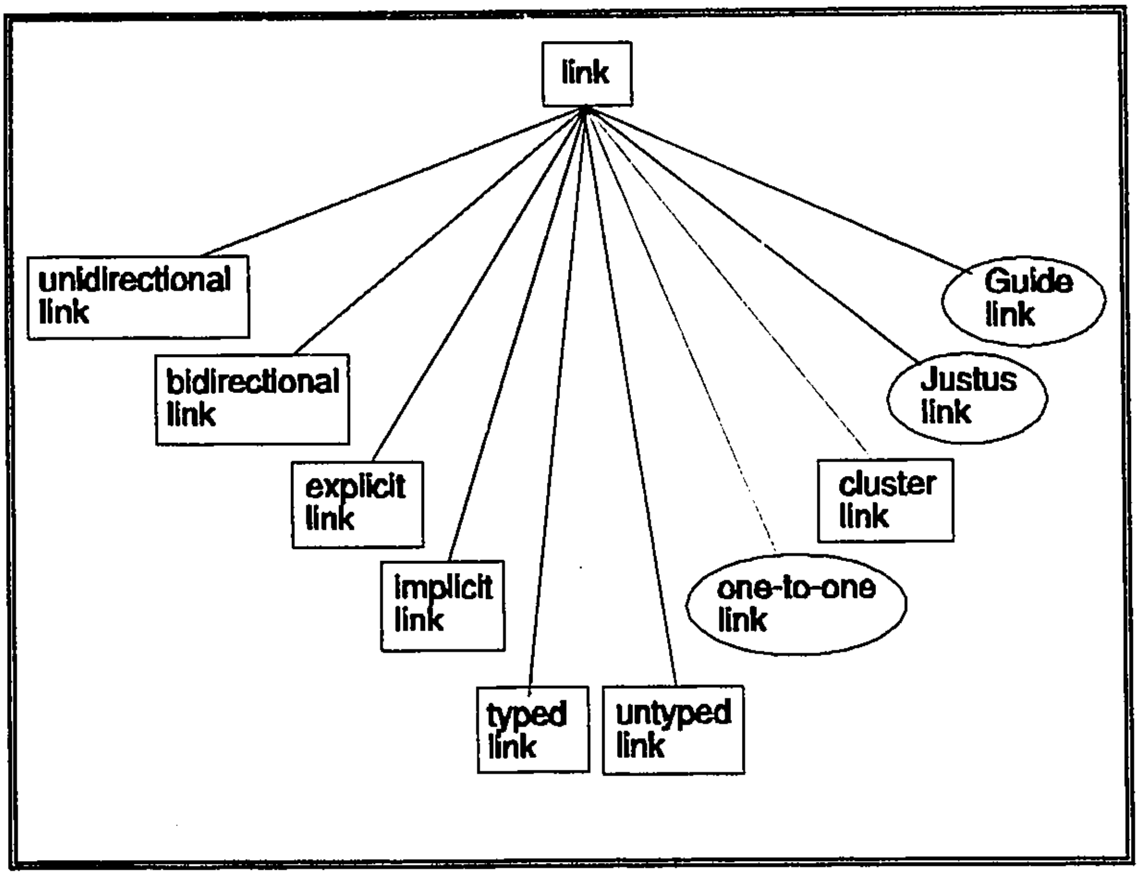


Figure 9-1. Different coloured lines used to indicate different dimensions.

Finally, a third helpful modification would be to make the characteristic underlying each dimension more easily accessible to the user. One way to do this would be to have the characteristic automatically appear beside the appropriate d-links on the graph. As this can sometimes be messy, especially if the graphical representation is complex, it might be a good idea to incorporate a toggle that would allow users to switch this option on and off. An even better way to make the characteristic underlying each dimension more accessible to the users would be to incorporate more hypertext capabilities in the system. In this case, the characteristic underlying each dimension could be stored in a pop-up window that could be activated by clicking on the appropriate d-link. When users position the cursor on a d-link and click and hold a mouse button, the characteristic would appear in a pop-up window that would subsequently disappear from view when the mouse button was released.

9.4 POSSIBLE AREAS FOR FUTURE RESEARCH

As mentioned in the introduction, due to the potential vastness of the subject matter, we restricted the scope of our thesis in several ways: to develop our initial guidelines, we only examined the classification techniques used in terminology, biology, and library science; we only examined multidimensionality as it occurred in generic-specific relations; and we only investigated the *recognition* and *representation* of multidimensionality. Moreover, we only subjected our proposed guidelines to a very modest test. Obviously, there is the potential for much more research to be carried out in the area of multidimensionality.

For instance, it may prove fruitful to examine the role of classification in other disciplines. For example, the relationship between classification and cognitive psychology is raised by Datta and Farradane (1974: 321), who observe that knowledge is a phenomenon of the mind, adding "...since knowledge comprises concepts and their interrelations, any ultimately enduring scheme of classification must be based upon the manner in which the mind forms and manipulates its concepts." Similarly, classification also plays a central role in the field of philosophy, and specifically in the subfield of logic. According to *The Encyclopedia Britannica* (1990: Vol. 14, p. 940), "The first great generalizer in classification was Aristotle, who virtually invented the science of logic, of which for 2,000 years classification was a part." It is possible that the classification techniques practised in these disciplines could add to the proposed guidelines.

It would also be interesting to see if the proposed guidelines could be used to successfully handle occurrences of multidimensionality in different relation types, such as partitive relations (e.g. part–whole) or non-hierarchical relations (e.g. cause–effect). Cruse (1986: 169), for example, not only observes that multidimensionality exists in part–whole relations, but adds that it can be quite complex, occurring at several levels within one classification.

> A *house*, for instance, may be divided into *living-room, dining-room, kitchen, hall, bedrooms, cellar, loft*, etc. (segmental parts); or *brickwork, joinery, plasterwork, plumbing, wiring*, etc. (systemic parts [...]). The case of *house* (and other buildings) is further complicated by the fact that the segmental parts can be seen either in terms of spaces (e.g. rooms), or in terms of the structural elements which define those spaces (for instance the *walls, floors* and *ceilings* are also parts of a *house*).

Furthermore, investigations of the implications of multidimensionality in other areas of terminology, such as definition construction, could also raise some interesting issues. For example, as explained in section 2.1.1 (*Analysis of data*), when terminologists define a term using the classic intensional definition, they refer to 1) the generic class to which the concept belongs, and 2) the characteristic(s) which differentiates this concept from other members of the same class. In multidimensional representations, concepts can belong to more than one dimension, and thus may have more than one generic concept, and more than one set of co-ordinate concepts.

With regards to testing the guidelines, further steps could include applying the guidelines to the construction of TKBs in a wider variety of subject fields containing different types of concepts, and also in larger subject fields to see whether the guidelines will scale up successfully. Finally, it would be interesting to how well the guidelines could be adapted to work with other knowledge engineering tools.

APPENDIX

Appendix

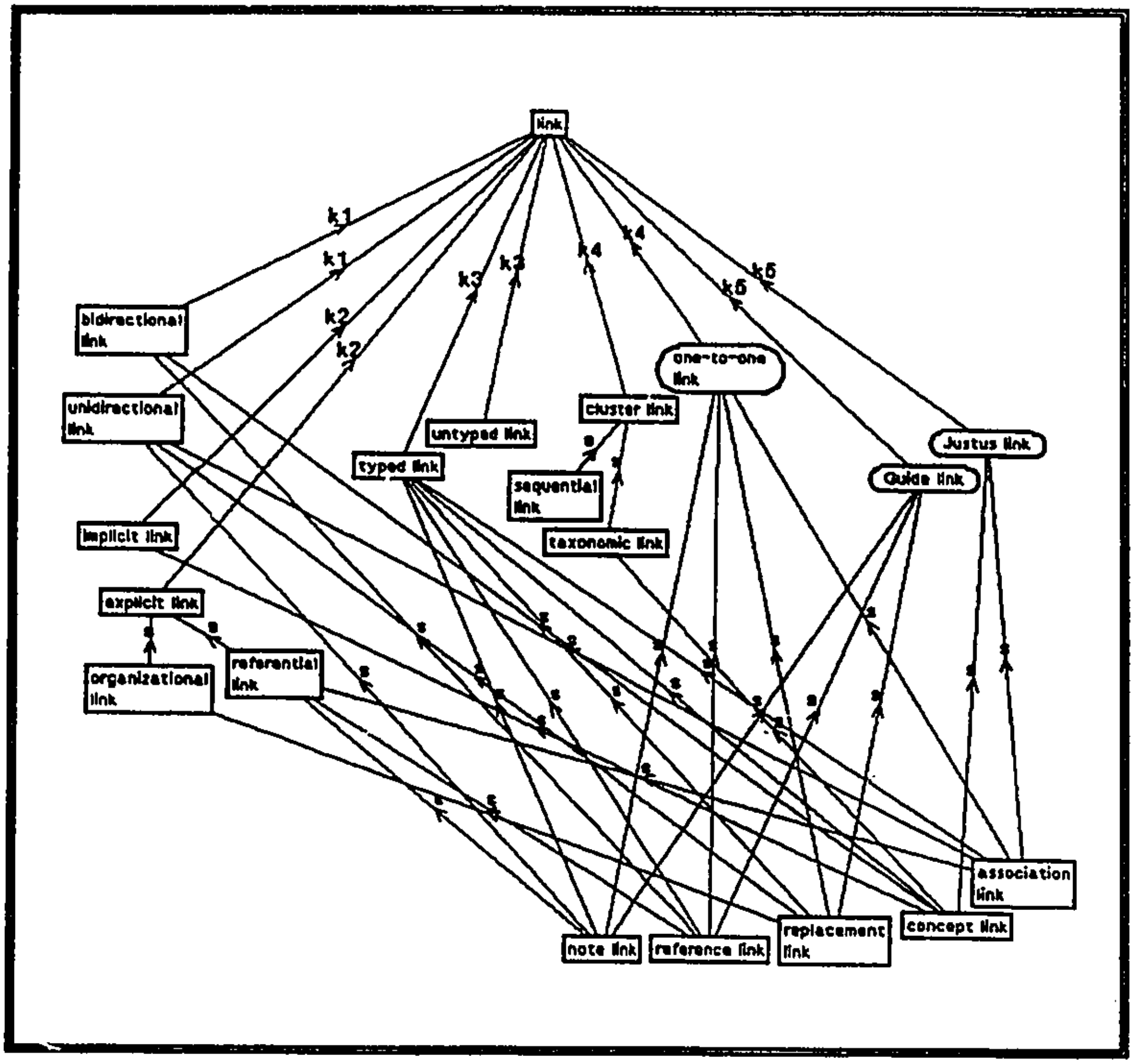
Completed CDs for the concepts in the TKB on the subfield of hypertext links

Organization of the CDs

A two-dimensional medium is not conducive to the systematic presentation of multidimensional information. Nonetheless, we have attempted to present the CDs in as logical an order as possible. We begin with the CD for LINK, which is the top concept in the classification, followed by the CDs for those concepts in dimension 1, followed by those in dimension 2, etc. Within the dimensions, the CDs for the concepts are presented in a top-down fashion whenever possible, and CDs of co-ordinate concepts are presented in alphabetical order. In cases where concepts belong to more than one dimension, we have placed their CDs with the CD of the superordinate concept to which we feel they are most strongly associated. The last CD presented is the bibliographic CD, which contains the full references for all the abbreviated references used in the other CDs.

The actual order of presentation of the CDs is as follows:

	1) link
(Dimension 1)	2) bidirectional link
	3) unidirectional link
(Dimension 2)	4) explicit link
	5) organizational link
	6) referential link
	7) implicit link
(Dimension 3)	8) typed link
	9) untyped link
(Dimension 4)	10) cluster link
	11) sequential link
	12) taxonomic link
	13) one-to-one link
(Dimension 5)	14) Guide link
	15) note link
	16) reference link
	17) replacement link
	18) Justus link
	19) association link
	20) concept link
(References)	21) bibliographic CD



A graph showing the relations between all the concepts in the TKB on the subfield hypertext links.

cdName: link
super: hypertext document
hasPropsOf:
kinds:
(1) unidirectional link, bidirectional link
(2) explicit link, implicit link
(3) typed link, untyped link
(4) cluster link, one-to-one link
(5) Guide link, Justus link
subConcepts:

SPECIAL
subject field: hypertext
creation date: 26 April 1992
last update date: 17 September 1992
done by: Lynne Bowker
figure:
status: nil
level:

DEFINITION
intensional definition: an electronic path or connection between two or more associated nodes. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: one of: two, more than two
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped

LINGUISTIC INFORMATION
English term: link ; COOK91: 10, FIDERIO88: 239, FRANKLIN90: 5, NIELSEN90: 2, BYERS87: 247, CONKLIN87: 33, RAYMOND88: 837, JONASSEN89A: 8, JONASSEN89B: 6, WATTERS92: 129, HAAVIND90: 42, CARR88: 280
English part of speech: noun
English textual support/cont1: "Various applications of the hypertext idea may differ in many ways. One feature they share is the ability to form *links* to connect objects on one screen with something else, usually on another screen. These *links* may not be visible to the user, but they are what holds a hypertext system together." ; COOK91: 10
English textual support/cont2: "In general, *links* are used to connect the nodes. A *link* is like an electronic footnote, endonte, or a parenthetical phrase. That is, just as footnotes and parenthetical phrases direct readers of printed material to related points or further topics for research, *links* connect you to associated text or ancillary information." ; FIDERIO88: 239
English textual support/cont3: "Button and *link* are complementary concepts: a button is a visible indication that two nodes intersect, and a *link* is the electronic connection between them." ; FRANKLIN90: 5
English collocation: to forge a link ; BYERS87: 250, CARR88: 280, HAAVIND90: 42
English collocation/textual support/cont1: "The key to implementing hypertext is the ability to constantly add new data and *forge new links*..." ; BYERS87: 250
English collocation/textual support/cont2: "In some hypertext systems, the linking process is so simple that any reader can cut a path through the text and *forge new links*." ; CARR88: 280

cdName: bidirectional link **super:** link **hasPropsOf:** **kinds:** **subConcepts:** reference link, concept link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **creation date:** 26 April 1992 **last update date:** 17 September 1992 **done by:** Lynne Bowker **figure:** **D-link by directionality:** **status:** nil **level:**

DEFINITION
intensional definition: a link that is anchored in two nodes, and can be traversed in either direction between the two nodes. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: one of: two, more than two
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: bidirectional ; NIELSEN90: 4
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped

LINGUISTIC INFORMATION
English term: bidirectional link ; NIELSEN90: 4, IRLER90: 264, HALASZ88: 840
English part of speech: noun
English textual support/cont1: "K. Eric Drexler has advocated the use of *bidirectional links*. [...] One example comes from Intermedia, which supports *bidirectional links*. A hypertext structure on Chinese poetry has links from each poem to the references of those anthologies where it has been reprinted and/or translated. This set-up automatically ensures that each listing for an anthology or a translator has a complete set of links pointing to occurrences of the relevant poems in the Intermedia hypertext." ; NIELSEN90: 4
English textual support/cont2: "*Bidirectional links* permit one to step back to the departure anchor." ; IRLER90: 264
English textual support/cont3: This system includes typed nodes connected by labeled, *bidirectional links*. ; HALASZ88: 840
English synonym/syn1: two-way link ; WILSON90: 204
English synonym/syn1/cont1: They could be converted into *two-way links* but the result of such a conversion would be to create an index of dictionary headwords. ; WILSON90: 204
English textual support/obs1: Authors often refer to the directionality of a link without specifically using the term bidirectional link. For example, "The links are bidirectional..." NIELSEN90: 98.

cdName: unidirectional link **super:** link **hasPropsOf:** **kinds:** **subConcepts:** replacement link, note link, association link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **creation date:** 26 April 1992 **last update date:** 17 September 1992 **done by:** Lynne Bowker **figure:** Figure 1.1 ; NIELSEN90: 1 **D-link by directionality:** **status:** nil **level:**

DEFINITION
intensional definition: a link that is anchored in only one node, and can therefore be traversed in only one direction, from its anchor node to its destination node. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: one of: two, more than two
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: unidirectional ; NIELSEN90: 3
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped

LINGUISTIC INFORMATION
English term: unidirectional link ; NIELSEN90: 3
English part of speech: noun
English textual support/cont1: "Almost all current hypertext systems are limited to providing "unidirectional links".... This means that the system can show the user the links that have the current node as their departure point, but not the ones that have it as their arrival point. In other words, the system will tell you where you can go next, but not in what alternative ways you might have arrived at where you are now." ; NIELSEN90: 3
English synonym/syn1: one-way link ; WILSON90: 204
English synonym/syn1/cont1: "I see association links, however they are implemented, as "one-way links"." ; WILSON90: 204
English textual support/obs1: Authors often refer to the directionality of a link without specifically using the term "unidirectional link". For example, "The links are also unidirectional..." SHERMAN90: 17, "...intensional links are in principle unidirectional..." DEROSE89: 250.

cdName: explicit link **super:** link **hasPropsOf:** **kinds:** **subConcepts:** referential link, organizational link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **creation date:** 4 May 1992 **last update date:** 17 September 1992 **done by:** Lynne Bowker **figure:** **D-link by determination of destination:** **status:** nil **level:**

DEFINITION
intensional definition: a link which has been defined by the author as always connecting a particular departure node with a particular destination node. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: one of: two, more than two
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: predefined ; NIELSEN90: 107
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped
type of node relation: one of: hierarchical or non-hierarchical

LINGUISTIC INFORMATION
English term: explicit link ; NIELSEN90: 107, 178, KNOPIK90: 123
English part of speech: noun
English textual support/cont1: "Given that hypertext is based on "explicit links", the next issue is whether or not to make the anchors especially prominent on the screen compared with the rest of the node." ; NIELSEN90: 107
English textual support/cont2: "Since the OED contains definitions of almost every word in the English language, it also contains an astronomical number of implicit links since users may jump from every word in the complete text to that word's entry even if the editors have not included an "explicit link"." ; NIELSEN90:178
English textual support/cont3: "... it is quite straightforward to incorporate "explicit links" in StrathTutor, but for many domains there will be a large amount of information that will remain implicit, even after coding into the most complex of hypertexts." ; KNOPIK90: 123
English synonym/syn1: hardwired link ; CARR88: 280
English synonym/syn1/cont1: "With a "hardwired link", the author specifies the launch point (where you jump from) and the destination of the link (where you jump to). [...] "Hardwired links" ... are established by an author who painstakingly links launch points and destinations. ..."hardwired links" are products of a meticulous editing process, in which many trains of thought and sources of argument carefully intersect." ; CARR88: 280
English synonym/syn2: hard-wired link ; IRLER90: 266
English synonym/syn2/cont1: "Systems with dynamic links - as opposed to static, ""hard-wired" links" - cannot simply show buttons with a generic label." ; IRLER90: 266
English synonym/syn3: extensional link ; DEROSE89: 251
English synonym/syn3/cont1: "Extensional links" are idiosyncratic, tying various parts of the docuverse together in upredicatble ways. They must be stored individually. ; DEROSE89: 251

cdName: organizational link **super:** explicit link **hasPropsOf:** **kinds:** **subConcepts:** replacement link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **creation date:** 26 April 1992 **last update date:** 17 September 1992 **done by:** Lynne Bowker **figure:** **D-link by determination of destination:** **status:** nil **level:**

DEFINITION
intensional definition: an explicit link that joins two nodes in a hierarchical fashion, showing the superordinate-subordinate relationship between them. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: one of: two, more than two
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: predefined ; NIELSEN90: 107
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped
type of node relation: hierarchical ; NIELSEN90: 91

LINGUISTIC INFORMATION
English term: organizational link ; CONKLIN87: 34, JONASSEN89A: 8, JONASSEN89B: 6
English part of speech: noun
English textual support/cont1: "There are two methods for explicitly linking two points in hypertext - by reference and by organization. [...] Like referential links, "organizational links" establish explicit links between points in hypertext. "Organizational links" differ from referential links in that they implement hierarchical information. "Organizational links" connect a parent node with its children and thus form a strict tree subgraph within the hypertext. They correspond to the ISA (or superconcept) links of semantic net theory, and thus operate quite differently than referential links." ; CONKLIN87: 34
English textual support/cont2: ""Organizational links" impart hierarchical information. "Organizational links" connect a parent node to its children in a tree fashion. Users are presented special links to parent, child or sibling nodes. For instance, in a hypertext that the author is developing, the parent node is a definition node. Users may connect to example nodes or implication or application nodes. Hierarchical systems are simpler with fewer navigational problems, but they constrain the user to pre-conceptualized organiztion." ; JONASSEN89A: 8
English textual support/cont3: "Links may be referential or organizational. [...] "Organizational links" are typically organized in a network of related nodes." ; JONASSEN89B: 6
English textual support/obs1: Organizational links often form cluster links since they frequently connect a single superordinate node to multiple subordinate nodes. ; BOWKER

cdName: referential link **super:** explicit link **hasPropsOf:** **kinds:** **subConcepts:** reference link, note link, association link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **creation date:** 26 April 1992 **last update date:** 17 September 1992 **done by:** Lynne Bowker **figure:** Figure 10 ; CONKLIN87: 34 **D-link by determination of destination:** **status:** nil **level:**

DEFINITION
intensional definition: an explicit link that connects two nodes, one of which contains a reference to the other, in a non-hierarchical fashion. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: one of: two, more than two
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: predefined ; NIELSEN90: 107
directionality: generally bidirectional (see obs1) ; JONASSEN89A: 8, JONASSEN89B: 6, LETHBRIDGE92
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped
type of node relation: non-hierarchical ; JONASSEN89B: 13

LINGUISTIC INFORMATION
English term: referential link ; CONKLIN87: 33, JONASSEN89A: 8, JONASSEN89B: 6, 13
English part of speech: noun
English textual support/cont1: "There are two methods for explicitly linking two points in hypertext - by reference and by organization. The reference method is a non-hierarchical method. It uses *referential links* that connect points or regions in the text. *Referential links* are the kind of link that most clearly distinguishes hypertext. They generally have two ends, and are usually directed, although most systems support "backward" movement along the link." (see obs1) ; CONKLIN87: 33
English textual support/cont2: "*Referential links* connect a source in the current node to a referent in the destination node. The user is typically allowed to return via the same link." ; JONASSEN89A: 8
English textual support/cont3: "Links may be referential or organizational. *Referential links* refer to information in another node and then permit the user to return via the same link." ; JONASSEN89B: 6
English textual support/cont4: "Unstructured hypermedia is random node-link hypermedia in which only *referential links* are used. This type of hypermedia provides random access directly from any node to any other node that is linked to it." ; JONASSEN89B: 13
English textual support/obs1: CONKLIN87: 33 states that referential links are usually unidirectional, but may be bidirectional, while JONASSEN89A: 8 and JONASSEN89B: 6 state the opposite. According to LETHBRIDGE92, the discrepancy may be due to the difference in publication dates: in 1987, most links were unidirectional; however, since then, technology has advanced and bidirectional links have become more popular.

cdName: implicit link **super:** link **hasPropsOf:** **kinds:** **subConcepts:** concept link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **creation date:** 4 May 1992 **last update date:** 17 September 1992 **done by:** Lynne Bowker **figure:** **D-link by determination of destination:** **status:** nil **level:**

DEFINITION

intensional definition: a link that is not predefined as being connected to a specific destination node; rather, its destination node is determined by the hypertext system based on the information given in the current and/or previous nodes. ; BOWKER

CONCEPTUAL INFORMATION

number of nodes connected: one of: two, more than two (DEROSE89: 254 notes "They [intensional links] may also have multiple ends.") ; DEROSE89: 254
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: computed ; NIELSEN90: 106, DEROSE89: 254
directionality: generally unidirectional ; DEROSE89: 254
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped

LINGUISTIC INFORMATION

English term: implicit link ; NIELSEN90: 106, DEROSE89: 254
English part of speech: noun
English textual support/cont1: "Some systems also provide *implicit links*, which are not defined as such, but follow from various properties of the information. A classic example of *implicit links* is the automatic glossary lookup possible in Intermedia. The InterLex server provides a link from any word in the dictionary, but it would obviously be ridiculous to have to store all these links explicitly. Only when the user requests the definition of a word does the system need to find the destination for the link." ; NIELSEN90: 106
English textual support/cont2: "*implicit links* may be used to access many other kinds of documents and subdocuments in addition to dictionaries. For example, online maps of all the world's countries ought to be accessible from any instance of a country name in any text." ; DEROSE89: 254
English textual support/cont3: "Since the OED contains definitions of almost every word in the English language, it also contains an astronomical number of *implicit links* since users may jump from every word in the complete text to that word's entry even if the editors have not included an explicit link." ; NIELSEN90:178
English synonym/syn1: dynamic link ; IRLER90: 266, WATTERS92: 70
English synonym/syn1/cont1: "Systems with *dynamic links* – as opposed to static, "hard-wired links" – cannot simply show buttons with a generic label. *Dynamic links* depend on the context of certain information rather than on its content or name. These links describe a destination anchor in generic terms, and its definitive instantiation is calculated from case to case. Simple *dynamic links* are text search features, as in AskSam and in Superbook: the occurrences of a selected word are searched and linked in sequence after activation." ; IRLER90: 266
English synonym/syn1/cont2: "...*dynamic link* is a hypertext link that was not explicitly declared during the authoring of the hypertext." ; WATTERS92: 70
English synonym/syn2: computed link ; NIELSEN90: 107, CARR88: 280
English synonym/syn2/cont1: "...the hypertext author is not expected to provide links between nodes, but is instead asked to define a set of relevant attributes for each node and for areas of interest in the node. These attributes are keywords taken from a predefined and restricted vocabulary. The areas of interest are called "hotspots" and serve a purpose similar to anchors in other hypertext systems. When the user activates a hotspot, the system will view the user's interests as being defined by the combination of attributes (keywords) from the current node and the selected hotspot. StrathTutor therefore links to a new node having the highest overlap between its own attributes and this set of attributes. [...] The StrathTutor links are an example of *computed links* determined by the system while the reader is reading, instead of being statically determined in advance by the author." ; NIELSEN90: 107
English synonym/syn2/cont2: "A *computed link* is much more like a word search command: Once you specify a search term, the hypertext system finds recurring instances and takes you there." ; CARR88: 280
English synonym/syn3: implicit computed link ; CONKLIN87: 36, MCCLELLAND89: 23
English synonym/syn3/cont1: "From a functional standpoint, link following and search are similar: Each is a way to access destination nodes that are of possible interest. Link following usually yields a single node, whereas search can yield many; hence, a keyword is a kind of *implicit computed link*." ; CONKLIN87: 36
English synonym/syn3/cont2: "At first, links will be based upon analysis of fairly structured fields. It is not, after all, a difficult task to construct a search from, say, an Author paragraph in order to link to documents in the same or another database by the same author. This merely requires translating the structure of the information in the original document to a suitable form for retrieving documents containing the same information in another database. What is required is an analysis of the structure of the corresponding fields so that the appropriate *"implicit computed link"* can be generated." ; MCCLELLAND89: 23
English synonym/syn4: computable link ; BRUILLARD89: 400
English synonym/syn4/cont1: ""True" dynamics in hypertext/hypermedia means that links and nodes are "computable." The possible links to be followed from a starting node are computed at the time the learner "wanders" through the network, and nodes may be programs instead of static information elements. In addition to *computable links* and nodes, computable networks should also be possible." ; BRUILLARD89: 400
English synonym/syn5: intensional link ; DEROSE89: 254
English synonym/syn5/cont1: ...*intensional link* types. These have in common that they follow strictly from the structure and content of the documents they link, and so need not be stored one by one in the hypertext system. In other words, the destination of an *intensional link* is defined by some function that finds the desired ends, rather than being a list of known ends. ; DEROSE89: 254

cdName: typed link **super:** link **hasPropsOf:** **kinds:** **subConcepts:** reference link, replacement link, note link, association link, concept link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **creation date:** 26 April 1992 **last update date:** 17 September 1992 **done by:** Lynne Bowker **figure:** Figure 7.1 ; WOODHEAD91: 118 **D-link by transparency of function:** **status:** nil **level:**

DEFINITION
intensional definition: a link that is a labelled in some way to indicate the nature of the predefined relationship between the nodes that it connects. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: one of: two, more than two
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: typed ; FRANKLIN90: 257

LINGUISTIC INFORMATION
English term: typed link ; FRANKLIN90: 257, DEYOUNG90: 240, FIDERIO88: 239, WOODHEAD91: 118
English part of speech: noun
English textual support/cont1: "A link is typed when it has a prespecified meaning. Guide, for example, has four "typed links": Command, Note, Reference, and Replacement." ; FRANKLIN90: 257
English textual support/cont2: ""Typed links" give users an indication of what lies at the destination..." ; DEYOUNG90: 240
English textual support/cont3: "A "typed link" specifies a particular relationship between two nodes..." ; FIDERIO88: 239
English textual support/cont4: "Providing "typed links" may add to the investment in development, but especially in larger knowledge bases, it is likely to result in the most manageable of structures - particularly if the links are presented consistenly." ; WOODHEAD91: 118
English textual support/obs1: According to LETHBRIDGE92, the vast majority of links are typed links. ; LETHBRIDGE92
English textual support/ex1: An example of a typed link would be a link that is defined so that it would allow a user who is reading a document to click on a button or link icon labelled "Reference" and see the reference for that document. ; LETHBRIDGE92

cdName: untyped link **super:** link **hasPropsOf:** **kinds:** **subConcepts:** **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **creation date:** 26 April 1992 **last update date:** 17 September 1992 **done by:** Lynne Bowker **figure:** **D-link by transparency of function:** **status:** nil **level:**

DEFINITION
intensional definition: a link which has no label to indicate its function. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: one of: two, more than two
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: untyped ; FRANKLIN90: 258

LINGUISTIC INFORMATION
English term: untyped link ; FRANKLIN90: 258
English part of speech: noun
English textual support/cont1: "The meaning of an "untyped link" varies depending on context. An "untyped link" can be used to do a wide range of things, from sending a command to a CD-ROM player to showing the definition of a word." ; FRANKLIN90: 258
English textual support/obs1: According to LETHBRIDGE92, untyped links are much rarer than typed links, and he was not able to provide a good definition of the term "untyped link". His statement would seem to be supported by the fact that the term "typed link" was found in four documents whereas the term "untyped link" was found in only one. ; LETHBRIDGE92

cdName: cluster link **super:** link **hasPropsOf:** **kinds:** **subConcepts:** sequential link, taxonomic link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **creation date:** 26 April 1992 **last update date:** 17 September 1992 **done by:** Lynne Bowker **figure:** **status:** nil **level:** **D-link by number of nodes connected:**

DEFINITION
intensional definition: a link that connects more than two nodes. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: more than two ; CONKLIN87: 35, DEROSE89: 252, NIELSEN90: 109
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped
order of node connection: one of: sequential, non-sequential

LINGUISTIC INFORMATION
English term: cluster link ; CONKLIN87: 35
English part of speech: noun
English textual support/cont1: "Links can connect more than two nodes to form "cluster links"." ; CONKLIN87: 35
English synonym/syn1: inclusion link ; DEROSE89: 252
English synonym/syn1/cont1: "...links which connect one originating location to many target locations, not just one, are called "inclusion links"..." ; DEROSE89: 252
English synonym/syn2: superlink ; NIELSEN90: 109, 127
English synonym/syn2/cont1: "In addition to the standard links connecting two nodes, some hypertext systems also have ""superlinks"" to connect a larger number of nodes." ; NIELSEN90: 109
English synonym/syn3: one-to-many link ; WILSON90: 204
English synonym/syn3/cont1: "One-to-many links" have an entirely different purpose from those considered so far. ; WILSON90: 204

cdName: sequential link super: cluster link hasPropsOf: kinds: subConcepts: inheritPropsTo: instanceOf: instances:	SPECIAL subject field: hypertext creation date: 6 May 1992 last update date: 17 September 1992 done by: Lynne Bowker figure: status: nil level: D-link by number of nodes connected:

DEFINITION

intensional definition: a cluster link that connects an ordered series of nodes. ; BOWKER

CONCEPTUAL INFORMATION

number of nodes connected: more than two ; CONKLIN87: 35, DEROSE89: 252, NIELSEN90: 109
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped
order of node connection: sequential ; DEROSE89: 252

LINGUISTIC INFORMATION

English term: sequential link ; DEROSE89: 252
English part of speech: noun
English textual support/cont1: "A "sequential link" has multiple ordered target locations. Paths are a simple example: it should be possible to associate a path with a given location, so the path is accessible from it as a matter of course, and this feature follows immediately from viewing paths as "sequential links". ; DEROSE89: 252

cdName: taxonomic link **super:** cluster link **hasPropsOf:** **kinds:** **subConcepts:** concept link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **creation date:** 6 May 1992 **last update date:** 17 September 1992 **done by:** Lynne Bowker **figure:** **status:** nil **level:** **D-link by number of nodes connected:**

DEFINITION
intensional definition: a cluster link that connects nodes in a non-sequential order. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: more than two ; CONKLIN87: 35, DEROSE89: 252, NIELSEN90: 109
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped
order of node connection: non-sequential ; DEROSE89: 253

LINGUISTIC INFORMATION
English term: taxonomic link ; DEROSE89: 253
English part of speech: noun
English textual support/cont1: "A "taxonomic link" leads to multiple target locations, but does not impose an order on them. Such links generally associate lists of properties with particular document elements." ; DEROSE89: 253

cdName: one-to-one link **super:** link **hasPropsOf:** **kinds:** **subConcepts:** reference link, replacement link, note link, association link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **done by:** bowker **status:** nil **last update date:** 17 September 1992 **figure:** **creation date:** 6 May 1992 **level:** **D-link by number of nodes connected:**

DEFINITION
intensional definition: a link that connects two nodes. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: two ; LETHBRIDGE92
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped

LINGUISTIC INFORMATION
English term: one-to-one link ; LETHBRIDGE92
English part of speech: noun
English textual support/cont1: NOTE: This concept does not appear to have been lexicalized, and the term "one-to-one link" was not found in any documentation. This term was suggested by a subject-field expert as an appropriate name for the concept.

cdName: Guide link **super:** link **hasPropsOf:** **kinds:** **subConcepts:** reference link, replacement link, note link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **done by:** bowker **status:** nil **last update date:** 17 September 1992 **figure:** **creation date:** 10 May 1992 **level:** **D-link by product:**

DEFINITION
intensional definition: a link used in the Guide hypertext system. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: one of: two, more than two
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped
Historical note: Peter J. Brown began developing the Guide hypertext system as a research project at the University of Kent (UK) in 1982. It was the first commercial hypertext system to become popular when a Macintosh version was released in 1986. A version has since been released for the IBM PC, and research is currently being conducted for a Unix version. ; NIELSEN90: 90

LINGUISTIC INFORMATION
English term: NOTE: "Guide link" is not an actual term. It is the name given by the terminologist to the dummy concept representing all links in the Guide hypertext system. ; BOWKER
English part of speech: noun
English textual support/cont1:

cdName: note link **super:** Guide link, unidirectional link, referential link, typed link, one-to-one link **hasPropsOf:** **kinds:** **subConcepts:** **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **done by:** bowker **status:** nil **last update date:** 17 September 1992 **figure:** **D-link by determination of destination:** **D-link by directionality:** **D-link by product:** **D-link by number of nodes connected:** **D-link by transparency of function:** **creation date:** 10 May 1992 **level:**

DEFINITION
intensional definition: a Guide link that causes the destination text to be displayed in a pop-up window, which disappears when the mouse button is released. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: two ; CONKLIN87: 32
specificity of anchor: text string ; NIELSEN90: 91, FIDERIO88: 242
specificty of destination: node ; CONKLIN87: 32, COOK91: 13
determination of destination: predefined ; LETHBRIDGE92
directionality: unidirectional ; LETHBRIDGE92
appearance: button ; COOK91: 13, NIELSEN90: 91, JONASSEN89A: 79, FIDERIO88: 242
method of activation: mouse click ; COOK91: 13, NIELSEN90: 91
transparency of function: typed ; JONASSEN89A: 79
type of node relation: hierarchical ; FIDERIO88: 242

LINGUISTIC INFORMATION
English term: note link ; CONKLIN87: 32, NIELSEN90: 91, JONASSEN89A: 79, FIDERIO88: 242
English part of speech: noun
English textual support/cont1: "Guide uses three kinds of links: ... *note links*, which display the destination text in a pop-up window...." ; CONKLIN87: 32

cdName: reference link **super:** Guide link, bidirectional link, referential link, typed link, one-to-one link **hasPropsOf:** **kinds:** **subConcepts:** **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **done by:** bowker **status:** nil **last update date:** 17 September 1992 **figure:** **D-link by determination of destination:** **D-link by directionality:** **D-link by product:** **D-link by number of nodes connected:** **D-link by transparency of function:** **creation date:** 10 May 1992 **level:**

DEFINITION
intensional definition: a Guide link which takes the user to another location in the hypertext document. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: two ; CONKLIN87: 32
specificity of anchor: text string or text region ; NIELSEN90: 91, COOK91: 13
specificty of destination: text region or node ; FIDERIO88: 242, JONASSEN89A: 79, COOK91: 13
determination of destination: predefined ; LETHBRIDGE92
directionality: bidirectional ; NIELSEN90: 91
appearance: button ; COOK91: 13, NIELSEN90: 91, JONASSEN89A: 79, FIDERIO88: 242
method of activation: mouse click ; COOK91: 13, NIELSEN90: 91
transparency of function: typed ; JONASSEN89A: 79
type of node relation: non-hierarchical ; JONASSEN89B: 13

LINGUISTIC INFORMATION
English term: reference link ; CONKLIN87: 32, NIELSEN90: 91, JONASSEN89A: 79, FIDERIO88: 242
English part of speech: noun
English textual support/cont1: "Guide uses three kinks of links: ... "reference links", which bring up a new window with the destination text as well." ; CONKLIN87: 32

cdName: replacement link **super:** Guide link, unidirectional link, organizational link, typed link, one-to-one link **hasPropsOf:** **kinds:** **subConcepts:** **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **done by:** bowker **status:** nil **last update date:** 17 September 1992 **figure:** **D-link by determination of destination:** **D-link by directionality:** **D-link by product:** **D-link by number of nodes connected:** **D-link by transparency of function:** **creation date:** 10 May 1992 **level:**

DEFINITION

intensional definition: a Guide link that causes any text in the current window to be completely replaced with the material at the end of the link. ; BOWKER

CONCEPTUAL INFORMATION

number of nodes connected: two ; CONKLIN87: 32
specificity of anchor: text region ; COOK91: 13
specificty of destination: text string, text region, complete file ; COOK91: 13, CONKLIN87: 32, JONASSEN89A: 79, NIELSEN90: 91
determination of destination: predefined ; LETHBRIDGE92
directionality: unidirectional ; LETHBRIDGE92
appearance: button ; COOK91: 13, NIELSEN90: 91, JONASSEN89A: 79, FIDERIO88: 242
method of activation: mouse click ; COOK91: 13, NIELSEN90: 91
transparency of function: typed ; JONASSEN89A: 79
type of node relation: hierarchical ; NIELSEN90: 91

LINGUISTIC INFORMATION

English term: replacement link ; CONKLIN87: 32, NIELSEN90: 91, JONASSEN89A: 79, FIDERIO88: 242
English part of speech: noun
English textual support/cont1: "Guide uses three kinds of links: *replacement links*, which cause the text in the current window to be completely replaced by the text pointed to by the link..." ; CONKLIN87: 32

cdName: Justus link **super:** link **hasPropsOf:** **kinds:** **subConcepts:** association link, concept link **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **done by:** bowker **status:** nil **last update date:** 17 September 1992 **figure:** **creation date:** 10 May 1992 **level:** **D-link by product:**

DEFINITION
intensional definition: a link in the Justus hypertext system. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: one of: two, more than two
specificity of anchor: one of: text string, text region, complete file
specificty of destination: one of: text string, text region, complete file
determination of destination: one of: predefined, computed
directionality: one of: unidirectional, bidirectional
appearance: one of: button, icon, invisible
method of activation: one of: mouse click, keystroke, menu selection
transparency of function: one of: typed, untyped
Historical note: Justus is a hypertext information retrieval system which runs under Unix and was developed at the University of Kent at Canterbury (UK). Justus was designed to provide an integrated document collection of primary and secondary sources in law with good browsing facilities and a range of access methods (including boolean query). ; WILSON90: 196

LINGUISTIC INFORMATION
English term: NOTE: "Justus link" is not an actual term. It is the name given by the terminologist to the dummy concept representing all links in the Justus hypertext system. ; BOWKER
English part of speech: noun
English textual support/cont 1:

cdName: association link **super:** Justus link, unidirectional link, referential link, typed link, one-to-one link **hasPropsOf:** **kinds:** **subConcepts:** **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **done by:** bowker **status:** nil **last update date:** 17 September 1992 **figure:** **D-link by determination of destination:** **D-link by directionality:** **D-link by product:** **D-link by number of nodes connected:** **D-link by transparency of function:** **creation date:** 10 May 1992 **level:**

DEFINITION
intensional definition: a Justus link that connect nodes containing supplementary information (e.g. definitions, references) to the main document. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: two ; WILSON90: 204
specificity of anchor: text string ; WILSON90: 203
specificty of destination: node ; WILSON90: 203
determination of destination: predefined ; WILSON90: 203
directionality: unidirectional ; WILSON90: 204
appearance: button ; WILSON90: 203
method of activation: mouse click ; WILSON90: 203
transparency of function: typed ; WILSON90: 203
type of node relation: non-hierarchical ; JONASSEN89B: 13

LINGUISTIC INFORMATION
English term: association link ; WILSON90: 204
English part of speech: noun
English textual support/cont1: "I see *association links*, however they are implemented, as one-way links. ...they are used to enhance and enlarge a view of the document collection from the perspective of a main document by linking in general reference texts such as dictionaries." ; WILSON90: 204

cdName: concept link **super:** Justus link, bidirectional link, implicit link, typed link, taxonomic link **hasPropsOf:** **kinds:** **subConcepts:** **inheritPropsTo:** **instanceOf:** **instances:**	**SPECIAL** **subject field:** hypertext **done by:** bowker **status:** nil **last update date:** 17 September 1992 **figure:** **D-link by directionality:** **D-link by transparency of function:** **D-link by product:** **D-link by number of nodes connected:** **D-link by determination of destination:** D-link by determination of destination **creation date:** 10 May 1992 **level:**

DEFINITION
intensional definition: a Justus link that connects areas of a document which have a similar meaning. ; BOWKER

CONCEPTUAL INFORMATION
number of nodes connected: more than two ; WILSON90: 204
specificity of anchor: text string ; WILSON90: 204
specificty of destination: text string ; WILSON90: 204
determination of destination: computed ; WILSON90: 204
directionality: bidirectional ; WILSON90: 204
appearance: button ; WILSON90: 204
method of activation: mouse click ; WILSON90: 204
transparency of function: typed ; WILSON90: 204
order of node connection: non-sequential ; DEROSE89: 253

LINGUISTIC INFORMATION
English term: concept link ; WILSON90: 204
English part of speech: noun
English textual support/cont 1: "*Concept links* help the reader to gain access to the interior of the document collection." ; WILSON90: 204

REFERENCES

BOWKER: These references are the opinion or work of the terminologist. Note: All definitions were written by the terminologist, but were based on the contexts referenced in each CD.
BRUILLARD89: Bruillard, Eric and Weidenfeld, Gérald. 1989. "Hyperinformation Requirements for an Integrated Authoring/Learning Environment," in <Designing Hypermedia for Learning>. Eds. David H. Jonassen and Heinz Mandl. Berlin: Springer Verlag. pp. 387-486.
BYERS87: Byers, T.J. "Built by Association," <PC World>. Vol. 5, No. 4. April 1987. pp. 244-251.
CARR88: Carr, Robert M. "The Hype of Hypertext," <PC World>. Vol. 6, No. 6. June 1988. pp. 277+.
CONKLIN87: Conklin, Jeffrey. "Hypertext: An Introduction and Survey," <IEEE Computer>. Vol. 20, No. 9. September 1987. pp. 17-41.
COOK91: Cook, Donald, A. "Lost (and Found) in Hyperspace," <CBT Directions>. July 1991. pp. 10-17.
DEROSE89: DeRose, Steven J. "Expanding the Notion of Links," <Hypertext '89 Proceedings>. November 1989. pp. 249-257.
DEYOUNG90: De Young, Laura. 1990. "Linking Considered Harmful," <Hypertext: Concepts, Systems and Applications>. Eds. N. Streitz, A. Rizk and J. André. Cambridge, UK: Cambridge University Press. pp. 238-249.
FIDERIO88: Fiderio, Janet. "A Grand Vision," <Byte>. October 1988. pp. 237-244.
FRANKLIN90: Franklin, Carl and Kinnell, Susan. 1990. <Hypertext/Hypermedia in Schools: A Resource Book>. Santa Barbara, CA: ABC-CLIO, Inc.
HAAVIND90: Haavind, Robert. "Hypertext: The Smart Tool for Information Overload," <Technology Review>. Vol. 93, No. 8. November/December 1990. p. 42.
HALASZ88: Halasz, Frank G. 1988. "Reflections on NoteCards: Seven Issues for the Next Generation of Hypermedia Systems," <Communications of the ACM>. Vol. 37, No. 7. pp. 836-852.
IRLER90: Irler, W.J. and Barbieri, G. 1990. "Non-intrusive Hypertext Anchors and Individual Colour Markings," in <Hypertext: Concepts, Systems and Applications>. Eds. N. Streitz, A. Rizk and J. André. Cambridge: Cambridge University Press. pp. 261-273.
JONASSEN89A: Jonassen, David H. 1989a. <Hypertext/Hypermedia>. Englewood Cliffs, NJ: Educational Technology Publications, Inc.
JONASSEN89B: Jonassen, David H. 1989b. "Problems and Issues in Designing Hypertext/Hypermedia for Learning," in <Designing Hypermedia for Learning>. Eds. David H. Jonassen and Heinz Mandl. Berlin: Springer Verlag. pp. 3-26.
KNOPIK90: Knopik, Thomas and Ryser, Sigrid. 1990. "AI Methods for Structuring Hypertext Information," in <Hypertext: State of the Art>. Eds. Ray McAleese and Catherine Green. Oxford: Intellect Limited. pp. 224-230.
LETHBRIDGE92: The terminologist consulted several times with subject-field expert Mr. Timothy C. Lethbridge, a PhD student in Computer Science at the University of Ottawa.
MCCLELLAND89: McClelland, Bruce. "Hypertext and Online... A Lot That's Familiar," <Online>. Vol. 13, No.1. January 1989. pp. 20-25.
NIELSEN90: Nielsen, Jakob. 1990. <Hypertext and Hypermedia>. San Diego, CA: Academic Press, Inc.
RADA91: Rada, Ray. 1991. <From Text to Expertext>. London, UK: McGraw-Hill.
RAYMOND88: Raymond, Darrell, R. and Tompa, Frank Wm. "Hypertext and the Oxford English Dictionary," <Communications of the ACM>. Vol. 31, No. 7. July 1988. pp. 871-879.
SHERMAN90: Sherman, Mark, Hansen, Wilfred, McInery, Michael and Neuendorffer, Tom. 1990. "Building Hypertext on a Multimedia Toolkit: An Overview of Andrew Toolkit Hypermedia Facilities," in <Hypertext: Concepts, Systems and Applications>. Eds. N. Streitz, A. Rizk and J. André. Cambridge: Cambridge University Press. pp. 13-24.
WATTERS92: Watters, Carolyn. 1992. <Dictionary of Information Science and Technology>. San Diego, CA: Academic Press, Inc.
WILSON90: Wilson, Eve. 1990. "Links and Structures in Hypertext Databases for Law," in <Hypertext: Concepts, Systems and Applications>. Eds. N. Streitz, A. Rizk and J. André. Cambridge: Cambridge University Press. pp. 194-211.
WOODHEAD91: Woodhead, Nigel. 1991. <Hypertext and Hypermedia>. Wokingham, UK: Addison-Wesley Publishing Co.

Bibliography

1. Terminology and Concept Systems

AUGER, Pierre and ROUSSEAU, Louis-Jean. 1978. *Méthodologie de la recherche terminologique*. Québec: Office de la langue française.

COLE, Wayne, D. 1987. "Terminology: Principles and Methods." *Computers and Translation*. Vol. 2, 77-87.

CRUSE, D.A. 1986. *Lexical Semantics*. Cambridge, UK: Cambridge University Press.

DUBUC, Robert. 1985. *Manuel pratique de terminologie*. Quebec: Linguatech.

ISO/TC 37, Terminology (principles and co-ordination). 1990. *ISO 1087: Terminology – Vocabulary*. Geneva: International Organization for Standardization.

ISO/TC 37, Terminology (principles and co-ordination). 1986. *ISO 704: Principles and Methods of Terminology*. Geneva: International Organization for Standardization.

MOUNIN, Georges. 1963. *Les problèmes théoriques de la traduction*. Paris: Éditions Gallimard.

PICHT, Heribert and DRASKAU, Jennifer. 1985. *Terminology: An Introduction*. Guildford: University of Surrey.

RONDEAU, Guy. 1984. *Introduction à la terminologie (Deuxième édition)*. Québec: Gaëtan Morin Éditeur.

SAGER, Juan C. 1990. *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

2. Knowledge-Based Terminology, CODE, and Artificial Intelligence

ECK, Karen. 1992 (working document). *Guidelines for Definition Construction in a Terminological Knowledge Base*. M.A. Thesis, University of Ottawa.

"KBMT – 89 Project Report." 1989. Pittsburgh, PA: Center for Machine Translation, Carnegie Mellon University.

MEYER, Ingrid. 1992a (in press). "Knowledge Management for Terminology-Intensive Applications: Needs and Tools." *Lexical Semantics and Knowledge Representation*, Eds. J. Pustejovsky and S. Bergler. Springer Verlag.

MEYER, Ingrid. 1992b (in press). "Concept Management for Terminology: A Knowledge Engineering Approach." *Standardizing Terminology for Better Communication: Practice, Applied Theory, and Results*. Eds. R.A. Strehlow and S.E Wright. Philadelphia: American Society for Testing Materials.

MEYER, Ingrid, BOWKER, Lynne and ECK, Karen. 1992a (in press). "Constructing a Knowledge-Based Term Bank: Fundamentals and Implications." *Proceedings of the International Symposium on Terminology and Documentation*. (Hull, Canada, October 7-8 1991). Department of the Secretary of State of Canada.

MEYER, Ingrid, BOWKER, Lynne and ECK, Karen. 1992b. "COGNITERM: An Experiment in Building a Knowledge-Based Term Bank." *Proceedings of the Fifth EURALEX International Congress (EURALEX '92)*. Tampere, Finland: Tampereen Yliopisto, 159-172.

MEYER, Ingrid, MILLER, David and MICHAUD, Diane. 1991. "Terminologie et analyse notionelle assistée par ordinateur." *Actes du colloque international sur les industries de la langue, Tome II*. Québec: Office de la langue française, 781-800.

MEYER, Ingrid and PARADIS, Line. 1991. "Applying Knowledge Engineering Technology to Terminology: A Pilot Project." *Terminology Update*, Vol. 24, No. 2, 3-8.

MEYER, Ingrid and SKUCE, Douglas. 1992. "Computer-Assisted Concept Analysis: A Knowledge-Based Approach to Terminology." *Proceedings of the Fourth EURALEX International Conference (EURALEX '90)*. Barcelona: Bibliograf, 129-138.

MEYER, Ingrid, SKUCE, Douglas, BOWKER, Lynne and ECK, Karen. 1992a. "Towards a New Generation of Terminological Resources: An Experiment in Building a Terminological Knowledge Base." *Proceedings of the 14th International Conference on Computational Linguistics (COLING '92), Vol. III*. Nantes: ICCL, 956-960.

MEYER, Ingrid, SKUCE, Douglas, BOWKER, Lynne and ECK, Karen. 1992b. "COGNITERM: Structure and Methodology for a Knowledge-Based Term Bank." Technical Report of the Artificial Intelligence Laboratory of the University of Ottawa.

PORTER, B., LESTER, J., MURRAY, K., PITTMAN, K., SOUTHER, A., ACKER, L., JONES, T. 1988. "AI Research in the Context of a Multifunctional Knowledge Base: The Botany Knowledge Base Project." Technical Report of the Department of Computer Science of the University of Texas at Austin.

SHAPIRO, Stuart C., ECKROTH, David, and VALLASI, George A. (Eds.) 1987. *Encyclopedia of Artificial Intelligence, Vol. 1 and 2*. New York: John Wiley & Sons, Inc.

SKUCE, Douglas. 1992a (in press). "A Wide Spectrum Knowledge Management System." *Knowledge Acquisition*.

SKUCE, Douglas. 1992b (submitted). "Managing Software Design Knowledge: A Tool and an Experiment." Submitted to *The IEEE Transactions on Knowledge and Data Engineering*.

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

SKUCE, Douglas and MEYER, Ingrid. 1990b. "Concept Analysis and Terminology: A Knowledge-Based Approach to Documentation." *Proceedings of the 13th International Conference on Computational Linguistics (COLING '90)*, 56-58.

SKUCE, Douglas and MEYER, Ingrid. 1991. "Terminology and Knowledge Engineering: Exploring a Symbiotic Relationship." *Proceedings of the 6th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop*. Vol. 2, 29-1 - 29-21.

3. Biological Classification

"Biological Science (Taxonomy)." *Encyclopedia Britannica*. 1990. Vol. 14. Chicago: Encyclopedia Britannica, Inc., 939-946.

DORIT, Robert L., WALKER, Warren F. Jr., and BARNES, Robert D. 1991. *Zoology*. Philadelphia: Saunders College Publishing.

GOTO, H.E. 1982. *Animal Taxonomy*. London: Edward Arnold Publishers Limited.

HICKMAN, Cleveland P. Jr., ROBERTS, Larry S., and HICKMAN, Frances M. 1986. *Biology of Animals (Fourth Edition)*. St. Louis: Times Mirror/Mosby College Publishing.

JEFFREY, Charles. 1982. *An Introduction to Plant Taxonomy (Second Edition)*. Cambridge: Cambridge University Press.

JEFFREY, Charles. 1989. *Biological Nomenclature (Third Edition)*. London: Edward Arnold.

MITCHELL, Lawrence G., MUTCHMOR, John A., and DOLPHIN, Warren D. 1988. *Zoology*. Menlo Park, CA: The Benjamin/Cummings Publishing Company, Inc.

4. Library Classification

BLISS, Henry Evelyn. 1929. *The Organization of Knowledge and the System of the Sciences*. New York: Henry Holt and Company.

BUCHANAN, Brian. 1979. *Theory of Library Classification*. London: Clive Bingley Ltd.

FOSKETT, A.C. 1977. *The Subject Approach to Information (Third edition).* London: Clive Bingley Ltd.

HUNTER, Eric J. 1988. *Classification Made Simple.* Aldershot, UK: Gower Publishing Company Limited.

MILLS, J. 1967. *A Modern Outline of Library Classification.* London: Figaro Press.

RANGANATHAN, S. R. 1967. *Prolegomena to Library Classification (Third edition).* London: Asia Publishing House.

SAYERS, Berwick W. C. 1967. *A Manual of Classification.* London: C. Tinling and Co. Ltd.

SHERA, Jesse H. 1976. *Introduction to Library Science.* Littleton, CO: Libraries Unlimited, Inc.

VICKERY, B.C. 1966. *Faceted Classification Schemes.* New Brunswick, NJ: Rutgers University Press.

5. Hypertext

BRUILLARD, Eric and WEIDENFELD, Gérald. 1989. "Hyperinformation Requirements for an Integrated Authoring/Learning Environment." *Designing Hypermedia for Learning.* Eds. David H. Jonassen and Heinz Mandl. Berlin: Springer Verlag, 387–486.

BYERS, T.J. 1987. "Built by Association." *PC World.* Vol. 5, No. 4, 244–251.

CARR, Robert M. 1988. "The Hype of Hypertext." *PC World.* Vol. 6, No. 6, 277+.

CHEN, Ching-chih. 1989a. *Hypersource on Multimedia/Hypermedia Technologies.* Chicago: American Library Association.

CONKLIN, Jeffrey. 1987. "Hypertext: An Introduction and Survey." *IEEE Computer.* Vol. 20, No. 9, 17–41.

COOK, Donald A. 1991. "Lost (and Found) in Hyperspace." *CBT Directions.* July 1991, 10–17.

DeROSE, Steven J. 1989. "Expanding the Notion of Links." *Hypertext '89 Proceedings*, 249–257.

De YOUNG, Laura. 1990. "Linking Considered Harmful." *Hypertext: Concepts, Systems and Applications.* Eds. N. Streitz, A. Rizk, and J. André. Cambridge, UK: Cambridge University Press, 238–249.

FIDERIO, Janet. 1988. "A Grand Vision." *Byte.* October, 237–244.

FRANKLIN, Carl and KINNELL, Susan. 1990. *Hypertext/Hypermedia in Schools: A Resource Book*. Santa Barbara, CA: ABC-CLIO, Inc.

HAAVIND, Robert. 1990. "Hypertext: The Smart Tool for Information Overload." *Technology Review*. Vol. 93, No. 8, 42.

HALASZ, Frank G. 1988. "Reflections of NoteCards: Seven Issues for the Next Generation of Hypermedia Systems." *Communications of the ACM*. Vol. 37, No. 7, 836-852.

IRLER, W.J. and BARBIERI, G. 1990. "Non-Intrusive Hypertext Anchors and Individual Colour Markings." *Hypertext: Concepts, Systems and Applications*. Eds. N. Streitz, A. Rizk, and J. André. Cambridge, UK: Cambridge University Press, 261-273.

JONASSEN, David H. 1989a. *Hypertext/Hypermedia*. Englewood Cliffs, NJ: Educational Technology Publications, Inc.

JONASSEN, David H. 1989b. "Problems and Issues in Designing Hypertext/Hypermedia for Learning." *Designing Hypermedia for Learning*. Eds. David H. Jonassen and Heinz Mandl. Berlin: Springer Verlag, 3-26.

KNOPIK, Thomas and RYSER, Sigrid. 1990. "AI Methods for Structuring Hypertext Information." *Hypertext: State of the Art*. Eds. Ray McAleese and Catherine Green. Oxford, UK: Intellect Limited, 224-230.

McCLELLAND, Bruce. 1989. "Hypertext and Online... A Lot That's Familiar." *Online*. Vol. 13, No. 1, 20-25.

NIELSEN, Jakob. 1990. *Hypertext and Hypermedia*. San Diego, CA: Academic Press, Inc.

RADA, Ray. 1991. *From Text to Expertext*. London: McGraw-Hill.

RAYMOND, Darrell R. and TOMPA, Frank Wm. 1988. "Hypertext and the Oxford English Dictionary." *Communications of the ACM*. Vol. 31, No. 7, 871-879.

SHERMAN, Mark, HANSEN, Wilfred, McINERY Michael and NEUENDORFFER, Tom. 1990. "Building Hypertext on a Multimedia Toolkit: An Overview of Andrew Toolkit Hypermedia Facilities." *Hypertext: Concepts, Systems and Applications*. Eds. N. Streitz, A. Rizk, and J. André. Cambridge, UK: Cambridge University Press, 13-24.

WATTERS, Carolyn. 1992. *Dictionary of Information Science and Technology*. San Diego, CA: Academic Press, Inc.

WILSON, Eve. 1990. "Links and Structures in Hypertext Databases for Law." *Hypertext: Concepts, Systems and Applications*. Eds. N. Streitz, A. Rizk, and J. André. Cambridge, UK: Cambridge University Press, 194-211.

WOODHEAD, Nigel. 1991. *Hypertext and Hypermedia*. Wokingham, UK: Addison-Wesley Publishing Co.

6. References Used for Examples

AVIS, Walter. S., et al. 1979. *Gage Canadian Dictionary*. Toronto: Gage Educational Publishing Company

Le Bescherelle 1 : L'Art de conjuguer. 1980. Paris: Hatier.

BRITISH COMPUTER SOCIETY GLOSSARY WORKING PARTY. 1989. *A Glossary of Computing Terms: An Introduction*. Cambridge: Cambridge University Press.

BUDDINE, Laura and YOUNG, Elizabeth. 1987. *The Brady Guide to CD-ROM*. New York: Prentice-Hall Press.

CHEN, Ching-chih. 1989b. *Hypersource on Optical Technologies*. Chicago: American Library Association.

COGNITERM. 1991. Optical Storage Technology Knowledge Base and Typesetting Knowledge Base. Artificial Intelligence Laboratory, Department of Computer Science, University of Ottawa.

DARGERY, Yvon. 1989. "Les disques optiques numériques." *Science & Vie Micro*. No. 65, 120–125.

DATTA, Suman and FARRADANE, Jason E.L. "A psychological basis for general classification." *Conceptual Basis of the Classification of Knowledge*. Ed. Jerzy A. Wojciechowski. Munich: Verlag Dokumnentation, 319–331.

ELSHAMI, Ahmed L. 1990. *CD-ROM Technology for Information Managers*. Chicago: American Library Association.

Good Housekeeping's Cookery Book. 1967. Edinburgh: Morrison and Gibb Ltd.

HEIMBURGER, Anneli. 1990. "A Guided Tour to Applications of Optical Disk Technology In the Nordic Countries." *CD-ROM Professional*. July, 31–35.

KENDRIS, Christopher. 1990. *French Verbs*. New York: Barron's Educational Series, Inc.

KRONENBERG, Shirley. 1967. *Cold Type Composition: Equipment and Techniques*. New York: John Wiley & Sons.

MACDONALD, Bonnie, and O'CONNELL, Marie. 1986. *Basics and Beyond: An Adventure in Cooking*. London: Pear Creative Ltd.

NEAL, Thomas and PARADIS, Line. 1986. *Terminology Series: Graphic Arts*. Ottawa: Department of The Secretary of State of Canada.

ROBERT, Paul. 1988. *Le Petit Robert 1: Dictionnaire de la langue française*. Paris: Dictionnaires Le Robert.

SAFFADY, William. 1988. *Optical Storage Technology: A State of the Art Review*. Westport, CT: Meckler Corporation.

Telesanast Teleordlista (Vocabulary of Telecommunications). 1991. Helsinki: Puhelinlaitosten Liitto R.Y. (A publication of Tekniikan Sanastokeskus, the centre for technical terminology in Finland).