A UDC WATER THESAURUS/CONCORDANCE: DEVELOPMENT AND USE. (LA CONCORDANCE DE CDU AVEC UN THESAURUS DE L'EAU: DEVELOPPEMENT ET UTILISATION)

Helen E. McCuaig
Marcel A. Mercier
Environment Canada
Ottawa, Ontario K1A 0E7

ABSTRACT

The efficient operation of an automated retrieval system requires that information in the system be effectively identified. Identification may be made through a controlled vocabulary, an uncontrolled vocabulary, or a classification or coding scheme. Each method has unique deficiencies which do not occur in others. With the advent of more sophisticated computer packages for information retrieval, it is possible to combine various identification methods and thus cancel individual disadvantages. In certain instances the system's vocabulary must be linked with the classification. A study to achieve such linking has been conducted at Environment Canada and has resulted in the concording of a water resources vocabulary with the Universal Decimal Classification scheme. Problems that arose, various tools that were used, basic methodology followed and results of the work are given. Plans for utilization of the concordance for indexing and information retrieval purposes within the WATDOC System are discussed and its potential use as an international switching language is outlined. (L'efficacité d'un système automatique de recherche documentaire requiert une identification claire de l'information. L'identification se fait par un vocabulaire contrôlé, un vocabulaire libre ou un système de classification ou de codage. Chacune de ces méthodes présente des déficiences uniques. Mais, avec l'avènement accru de programmes d'ordinateurs sophistiqués reliés à la recherche documentaire, on peut combiner diverses méthodes d'identification et éviter ainsi les désavantages particuliers de chaque méthode. Pour arriver à cette combinaison, il faut relier le vocabulaire utilisé au système de classification. Une étude afin de réaliser cette combinaison a été effectuée au ministère de l'environnement du Canada.

WATER THESAURUS/CONCORDANCE

Une correspondance a été établie entre un thésaurus de l'eau et le système de "classification décimale universelle". L'article suivant présente les problèmes rencontrés, les instruments de travail utilisés, la méthodologie de base suivie et les résultats obtenus. Les perspectives d'utilisation de la concordance "EAU/CDU" reliées à l'indexation et à la recherche documentaire à l'intérieur de WATDOC y sont discutées, ainsi que les possibilités d'utiliser la concordance comme langage de commutation.)

INTRODUCTION

During recent years information systems have increasingly dealt with matters of a multidisciplinary nature and increasingly have been directed towards a mixed clientele. WATDOC, Water Resources Document Reference System, Environment Canada, is an example of such a system, where references to all types of environmental information related to water resources are acquired from and used by the participants. The users are found at all levels of management and research in both the public and private sector.

For the efficient operation of such a system, it is frequently necessary to identify the information within the data base on a conceptual level and then within the general concept to separate documents through precise semantic coupling of keywords. Ohman (1971), Aitcheson (1970). WATDOC has attempted to provide this facility through the development and subsequent use of a concordance between a water resources thesaurus and a widely used classification scheme, the UDC.

With the tools that are available today, special editions of general classification schedules can be developed from the universal set and then the special edition may be combined with a structured vocabulary. In the case of an international classification it is possible to use such schedule modification for the development of an international switching language and to provide also an interdisciplinary switch between systems. Lloyd (1972).

It is interesting to note that the latest UNISIST recommendations indicate the need for a new broad system of ordering, that is a general classification, to identify information effectively and to allow its movement between various systems. However, despite such recommendations and despite extensive work on classification theory, Wilson (1972), people have not yet come to grips with the complex problems posed by classification systems. Some of these are the need for varying levels

of ordering, for linking relationships, and for a truly universal approach to classification. UDC, in spite of its deficiencies, is the only widely used scheme that attempts to provide any such facilities. Scribor (1971), Stueart (1971), Saveljeva (1971), Rigby (1970), Wright (1969).

CHOICE OF CLASSIFICATION AND INDEXING SYSTEMS

The Universal Decimal Classification was selected for the project of Environment Canada (WATDOC), on the basis of studies completed at the University of Alberta. These studies, even though preliminary, demonstrated that UDC could be regarded as a metalanguage and used effectively for automated retrieval purposes in water resources management. Mercier (1971). Through the computer manipulation of machine readable data bases, a significant correspondence between water resources terms and UDC notations was found.

To meet the requirements for controlled indexing within the WATDOC system it was recognized that a thesaurus was needed. The thesaurus would provide control of the vocabulary and would allow for consistency in selection of indexing terms; it would be a tool used by both indexer and searcher. Wellish (1972) The U.S. Water Resources Thesaurus was a natural choice.

COMBINATION OF SYSTEMS

The concordance between a Canadianized version of the U.S. Water Resources Thesaurus and a subset of the Universal Decimal Classification is to be used as an indexing, classifying and retrieval aid. The concordance and the subset of the UDC are being developed in parallel. Tapes used contained the U.S. Water Resources Thesaurus, sections of the UDC schedules, and the AIP/UDC magnetic tape. Mercier (1971, p.28).

Universal Decimal Classification Scheme

UDC is a system of classifying information by analysis of idea content so that related concepts are grouped and subordinated. It is a hierarchical, numerical classification based on the Dewey principle that all knowledge is a whole and can thus be designated by 0., with infinite capability of subdivision as a decimal fraction. It is very flexible in that it identifies certain recurring faceted features such as form, time and point of view. These, indicated by auxiliary numbers, can be appended to any concept. Other types of synthesis can also be indicated. British Standards Institution (1963), Mills (1964), Foskett (1969).

UDC is under the auspices of an international body, Fédération Internationale de Documentation (FID). Its tables -- though not all

parts -- are in 22 languages. Abridged editions have been published and are useful for small collections. Several Special Editions exist for the fields of Metallurgy, Nuclear Engineering, Education,and Polar Studies, where all relevant UDC numbers are brought together, those for the core subject in full detail and those on marginal subjects in abridged form. It is anticipated that a Special Edition for Water Management, in machine readable form, will grow out of this project. This water edition will exist as a concordance to the water resources thesaurus. Lloyd (1972). The Canadian National Committee (CNC/FID) has recently been enlarged and although the UDC has not been widely known or used in Canada, several interesting projects are underway.

U.S. Water Resources Thesaurus

The thesaurus used in this project as a basis for the core retrieval language was developed by Water Resources Scientific Information Center, U.S. Department of the Interior. It consists of 4,288 main terms and cross references interfiled with USE references. The thesaurus was "patriated" somewhat by deleting from it certain specific American terms, mostly geographic, and appending to it additional terms used extensively in Canada.

PROBLEMS IN DEVELOPMENT

In carrying out the project, the major problems were the acquisition of up to date UDC schedules and the lack of staff expertise in UDC.

Acquisition of Schedules

During the first phase of the research,difficulty was experienced in obtaining up to date schedules. This was largely overcome through the use of the magnetic tape containing the complete merged set of UDC schedules prepared by the American Institute of Physics Universal Decimal Classification Project, 1965 to 1967. Freeman (1968). For the present phase the full Special Schedules and the Abridged English Edition were used in printed form. However, there were many difficulties in acquiring all schedules* and in order to get on with the task schedules were borrowed, or photocopies obtained from the National Library, National Science Library (Building Research) and International Development Research Centre.

---

* In Canada all orders for English or French editions should be made through the Canadian Standards Association.

WATER THESAURUS/CONCORDANCE

Special editions were very helpful in this project; a list of those referred to were:

(a) Chiapetti, F. et al, 1971, An International Language for Information Retrieval: The UDC-Thesaurus in the Water Field C.N.R. Institute di Recerca Sulle Acque, Rapporto Interno 18, Roma.

(b) Deutsche Dokumentations-Zentrale Wasser (DZW), 1971, Ordungssystematik zur Dokumentation Wasser: Begriffsverzeichnis: Thesaurus, Erich Schmidt, Verlag, Berlin.

(c) Fédération Internationale de Documentation, 1955, ABC Abridged Building Classification, 2nd ed., Beuweentrum, Rotterdam (FID 261)

(d) Fédération Internationale de Documentation, 1963, Universal Decimal Classification for use in Polar Libraries, 2nd ed., revised, S.P.R.I. Occasional Paper No. 2 (FID 348)

(e) Food and Agriculture Organization, 1950, Handbook for World Fisheries Abstracts, Washington, D.C.

(f) International Committee for Atmospheric Sciences, 1966, ICAS Vocabulary (Preliminary Edition), Washington, D.C.

(g) National Research Council of Canada, 1965, Index of Report Files in the Building Research Library, compiled by Mara E. Karnupe, Technical Paper No. 197, Division of Building Research, Ottawa.

(h) U.S. Department of Housing and Urban Development, 1971, Urban Vocabulary, Washington, D.C.

(i) New schedules for Classes 551.48/.49 (Surface and Groundwater Hydrology within the Geomorphology Class) and Class 556 (General Hydrology) which have been acquired through correspondence with researchers in the field.

Lack of Staff Expertise in UDC

The individuals who participated in this phase of the project were a librarian and a professional assistant. No major problems arose although they were not experienced in the use of UDC. Programmed texts by Wellisch (1970) and Perreault (1969) were used to acquire familiarity with the principles and techniques of UDC.

WATER THESAURUS/CONCORDANCE

DEVELOPMENT OF THE CONCORDANCE

Phase 1

Three copies of the U.S. Water Resources Thesaurus were used, two as working copies and one as the master. The first step was to transfer notations derived in the Edmonton studies, Mercier (1971), to the second edition of the U.S. Thesaurus. Following this two major tasks were necessary. These were the further verification of the machine concorded notations derived in the Edmonton study and the manual assignment of notations to those descriptors not previously concorded. The basic steps are depicted in Figure 1; they are described as follows:

Step 1: The Abridged edition was checked to verify existing descriptor notations and as an initial step in assigning new notations for those descriptors previously left blank.

Step 2: If Step 1 was unsuccessful, the other tools on hand were utilized.

Step 3: If a match could not be made, an attempt was made to assign a notation to the term by synthesizing two UDC classes, or adding auxiliaries, otherwise no number was assigned.

Step 4: All notations were neatly written into the Master Copy for encoding. When completed, the various sections of the Master Copy were passed on to the keypunchers for keying. In this manner all steps were completed in parallel.

Phase 2

All terms and existing notations were punched onto cards so that various lists could be compiled and printed. The basic steps were as follows:

Step 1: The terms and their notations were punched in fixed fields.

Col. 2-35	Descriptor
Col. 47-68	UDC notation. If no notation occurred the field was left blank.

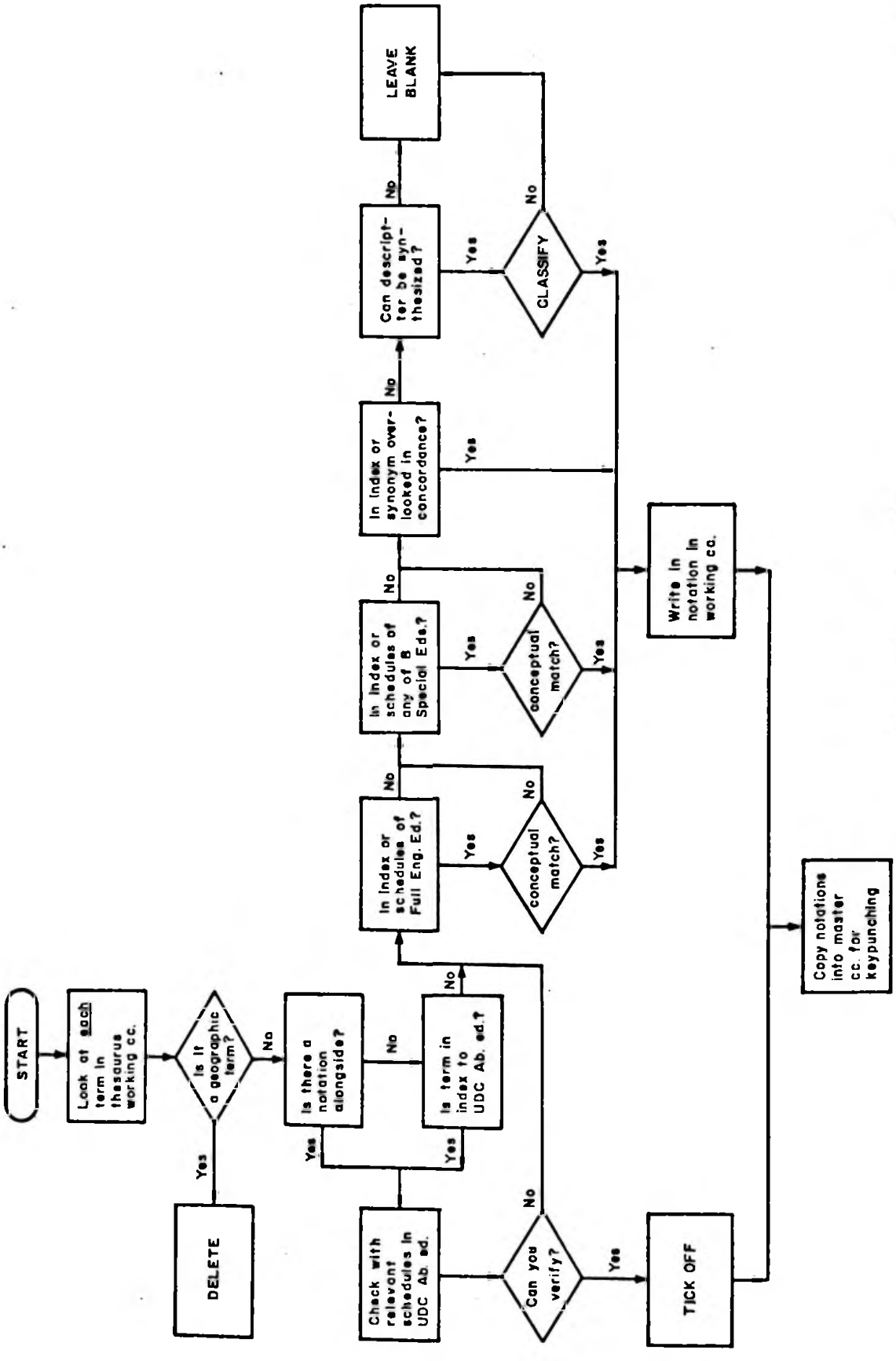

Figure I. Development of the Concordance - Phase I

Step 2: The cards were copied onto tape for ease of computer manipulation. Two listings were then produced:

1) An alphabetic listing
2) A numerical listing by UDC term

Both listings consisted of the descriptor term and the UDC notation. A sample of the alphabetic listing is given in Figure 2.

Step 3: For clarity, headings were added to the numerical listing. This was done by punching two cards for each heading.

Card 1,	Col.	16	*
	Col.	47-48	UDC notation
Card 2,	Col.	16-45	Heading
	Col.	47-48	UDC notation

These cards were added to the previous file. A sort was made and a numerical listing, containing headings, was printed. The headings were based on those occurring in the abridged and full editions of the UDC. They were selected on the basis of frequency of groups of notations occurring in the listing. A new heading was also inserted if a notation seemed remotely connected to the nearest heading. A sample of the resulting list is given in Figure 3.

Phase 3

To date, phase three of the project has not been completed. The object is to produce a printed subset of UDC schedules in machine readable form. Three basic steps have been defined, which are presently in parallel operation. These steps are as follows:

Step 1: Copies of the alphabetic and numerical listings and the UDC schedules have been distributed to professionals, within the Department, with special backgrounds such as biology, chemistry, economics, and engineering. They have been asked to examine the schedules of their specialty with particular reference to the two listings provided. Their task is to check the numerical listings with the special or abridged schedules and to select additional notations when expansion is considered necessary.

WATER THESAURUS/CONCORDANCE

BULKHEADS	623.92
BULLHEADS	
BULRUSHES	582.542.4
BUNDLED CONDUCTORS	
BUOYANCY	532.3
BUOYS	627.9
BURNING	614.84
BURNING	632.187
BURNING	631.6.068
BURROWS	591.52
BUS (ELECTRICAL)	
BUTTERFLY VALVES	621.646.25
BUTTRESS DAMS	627.828
BYPASSES	
BYPRODUCTS	.002.67
CACTI	582.85
CADDISFLIES	595.74
CADMIUM	546.48
CADMIUM RADIOISOTOPES	
CAISSONS	624.157
CALCAREOUS SOILS	631.411.2
CALCITE	549.742.111
CALCIUM	546.41
CALCIUM CARBONATE	661.652.5
CALCIUM CHLORIDE	661.842.321
CALCIUM COMPOUNDS	631.416.7
CALCIUM COMPOUNDS	661.842
CALCIUM HYDROXIDE	
CALCIUM SULFATE	
CALIBRATIONS	53.089
CALICHE	661.471.3
CAMERAS	771.3
CAMP SITES	796.54.006
CAMPING	796.54
CANADA GOOSE	598.413
CANAL CONSTRUCTION	626.13
CANAL DESIGN	626.12
CANAL EMBANKMENTS	626.17
CANAL LININGS	626.3
CANAL SEEPAGE	626.826
CANAL SEEPAGE	626.14
CANALS	556.53
CANALS	626.1
CANNERIES	639.2.068
CANOPY	551.584.2
CANTILEVERS	624.35

Figure 2 . Alphabetic List.

*	624.15
*FOUNDATIONS	624.15
FLEXIBLE FOUNDATIONS	624.15
FOUNDATIONS	624.15
DAM FOUNDATIONS	624.152.63
SHEET PILING	624.152.63
MOVABLE DAMS	624.152.632.1
PILES (FOUNDATIONS)	624.154
TIMBER PILES	624.154.2
CONCRETE PILES	624.154.3
STEEL PILES	624.154.7
FRICTION PILES	624.155
PIPE DRIVING	624.155
CAISSONS	624.157
EARTHQUAKE ENGINEERING	624.159.1
REINFORCEMENT	624.159.4
ABUTMENTS	624.16
PIERS	624.16
*	624.19
*TUNNELS & TUNNELLING	624.19
PRESSURE TUNNELS	624.19
TUNNEL CONSTRUCTION	624.19
TUNNELING	624.19
TUNNELS	624.19
TUNNEL FAILURE	624.19.004.64
TUNNEL LININGS	624.19.02
TUNNELING MACHINES	624.19.02
TUNNEL DESIGN	624.191.1
TUNNEL HYDRAULICS	624.196
BRIDGE CONSTRUCTION	624.21
BRIDGES	624.21
BRIDGE DESIGN	624.21:72.01
CANTILEVERS	624.35
STRUCTURAL DESIGN	624.4
STRUCTURAL MODELS	624.9.001.57
ROOFS	624.91
*	625
*RAILWAY . HIGHWAY	625
AGGREGATES	625.07
GRADING	625.071
RAILROADS	625.1/.6
ROAD CONSTRUCTION	625.7
ROADS	625.71
ROADBANKS	625.714
ROAD DESIGN	625.72
CULVERTS	625.745
GULLIES	625.745
PAVING	625.8
ASPHALTIC CONCRETE	625.855.3
SOIL AGGREGATES	625.855.5

Figure 3. Notational List

Step 2: Cards for the notations to be inserted are being punched as before but without descriptor. These will then be sorted with the previously punched cards.

Step 3: A print of all UDC notations and their English language equivalent will be made to produce a subset of UDC water schedules. This will be done by matching each notation with the AIP/UDC tape through a computer program. For those notations in which no match occurs, a default will be made to print what occurs on the card containing the notation. In this way a description will be printed in the subset of schedules for all new or derived notations.

OBSERVATIONS AND RESULTS TO DATE

The numerical printout resulting from phases 1 and 2 confirmed that the classes most frequently used for water resources information are Classes 5 and 6 (Science and Technology). The ubiquity of water is obvious in that notations from all classes, except Class 2 (Religion), appeared. Only small parts fell into the socio-economic areas.

Less that 20% of the terms were not given a notation. Ambiguous terms such as "comparative benefits" or "limiting factors" fell into this category. The classifier must ask: "Comparative benefits in what context?" or "Limiting factors in what area?". In spite of the knowledge that such terms are related to water and are defined through the synonyms and related terms listed below each descriptor in the thesaurus, the conceptual area or areas into which these descriptors fall are either too numerous or too ambiguous to be given a specific UDC notation. On the other hand, one of the most important features of UDC is that it is an aspect classification, showing subjects in different contexts. In some cases several notations can be listed, e.g.

		Aspect
CONDENSATION	536.423	Physics(Heat. Thermodynamics.)
CONDENSATION	551.574	Meteorology.
CONDENSATION	556.131.3	Hydrologic Cycle.

At this stage the work has resulted in the development of an effective water thesaurus/concordance. Because we have not yet had time to produce and publish a Canadian thesaurus/concordance, a copy of the U.S. Thesaurus with UDC notations typed alongside the descriptors is

currently being used. (Figure 4) Additional vocabulary not found in the U.S. Thesaurus has been added as an addendum.

The Thesaurus/Concordance and the Special Water Schedules, to be completed in phase 3, will be issued in two volumes. These will be available in draft form after a testing period of indexing and searching in the WATDOC System. It is planned to have students use these tools on a trial basis during the summer.

IMPLEMENTATION IN WATDOC SYSTEM

Indexing of documents will not be restricted by the concordance. Indexers will be instructed to follow it whenever possible and to use the concordance as a primary tool for the selection of terms and subject classes. In an automated system some degree of consistency is imperative for the efficient retrieval of documents. Further, the schedules will be kept up to date through computer programs being developed for the production of the special water schedules. If efficient programs can be developed for such updating they can contribute to the eventual development of program packages for the automated updating and printing of schedules. As suggested by Wellisch (1971), the weakness of UDC in updating would be largely eliminated if compact editions could be generated from a central magnetic tape store.

Aquatic Life
Fish
Wildlife
RT Anadromous Fish
Animal Groupings
Eels
Fish Migration
Fish Types
Freshwater Fish
Wildlife

Catalysts 54-4
RT Chemcontrol 577.1
Chemical Reactions
Chemicals 66.097
Enzymes
Inhibitors 678.044
Retardants

Catastomids
USE Suckers

Catch Crops
BT Crops
RT Disasters
Drought Tolerance
Droughts
Dry Farming
Farm Management
Freezing
Land Use
Storms
Water Injury
Winterkilling

Catchment Basins
USE Watersheds (Basins)

Catfishes 597.554.4
UF Blindcats
Blue Catfish
Headwater Catfish
Ictalurus Catus
Ictalurus Furcatus
Ictalurus Lupus
Noturus Species
White Catfish
BT Aquatic Animals
Aquatic Life
Fish
Freshwater Fish
Pan Fish
Wildlife
NT Bullheads
Channel Catfish
Madtoms
RT Rough Fish

Cathodes 621.3.035.22
BT Electrodes
Equipment
NT Cathodic Protection
RT Corrosion Control
Electrochemistry
Electrolysis

Cathodic Protection 620.197.5
BT Cathodes
Electrodes
Equipment
RT Coatings
Corrosion
Linings
Pitting (Corrosion)

Cation Adsorption
BT Adsorption
Chemical Properties
Ions
Properties
RT Anion Adsorption
Cations
Clay Minerals
Ions

Cation Exchange
BT Chemical Reactions
Ion Exchange
Ions
Separation Techniques
Treatment
Water Treatment
RT Anion Adsorption
Anion Exchange
Cations
Control
Demineralization
Permselective Membranes
Quality Control
Water Quality
Water Quality Control
Water Treatment

Cations
BT Ions
RT Cation Adsorption
Cation Exchange
Electrolytes

Cattails 582.522.1
BT Amphibious Plants
Aquatic Life
Aquatic Plants
Monocots
Rooted Aquatic Plants
RT Aquatic Weeds
Riparian Plants

Cattle 636.2
BT Domestic Animals
Livestock
Mammals
Ruminants
RT Feed Lots

Cavefishes 597.24
BT Aquatic Animals
Aquatic Life
Fish
Freshwater Fish
Wildlife
RT Caves

Caverns
USE Caves

Caves 551.35.054.12
UF Caverns 551.442
BT Geology
Geomorphology
RT Cavefishes
Karst
Karst Hydrology
Limestones
Sinks
Travertine

Cavitation 532.528
BT Fluid Mechanics
Hydrology
RT Aeration
Bubbles
Corrosion
Erosion
Flow
Flow Separation
Flumes
Hydraulic Structures
Hydrofoils
Hydrologic Aspects
Impellers
Irrigation Water
Negative Pressure
Pressure
Reaeration
Scour
Stream Erosion
Vortices

Celerity
RT Rates
Standing Waves
Waves (Water)

Cells (Biological)
USE Cytological Studies

Cellulose 547.458.8
BT Carbohydrates 661.728
Organic Compounds
RT Fibers (Plant) 676.16
Lignins
Lumber
Pulp Wastes
Resins
Vascular Tissues
Wood Wastes

Cement Grouting 666.97.033.14
BT Grouting
RT Cements
Mortar

Cements 666.94
BT Construction Materials
NT Asbestos Cement 691.54
Portland Cements
RT Adhesives
Aggregates
Alkali-Aggregate Reactions
Asphalt
Cement Grouting
Clays
Concrete Additives
Concrete Mixes
Concrete Technology
Concretes
Diatomaceous Earth
Grouting
Linings
Masonry
Mortar
Paints
Pozzolans

Figure 4 . Example of Concordance

REFERENCES

AITCHESON, Jean, 1970, "The Thesaurofacet: A Multipurpose Retrieval Language Tool", Journal of Documentation, v. 26, No. 3, p.187-203.

BRITISH STANDARDS INSTITUTION, 1963, Guide to the Universal Decimal Classification (UDC), British Standards House, London.

FOSKETT, A.C., 1969, The Subject Approach to Information, London, Bingley.

FREEMAN, R.R., ATHERTON, P., 1968, Final Report of the Research Project for the Evaluation of the UDC on the Indexing Language for a Mechanized Reference Retrieval System, Report AIP/UDC-9.

LLOYD, G.A., 1972, "The Universal Decimal Classification as a Switching Language", Subject Retrieval in the Seventies, School of Library and Information Services, University of Maryland, Greenwood, p. 116-125.

----------, 1972, Head, Classification Dept., Fédération Internationale de Documentation, The Hague, Personal Communication, 26 May, 1972.

MERCIER, Marcel A., 1971, "Study of UDC and other Indexing Languages through Computer Manipulation of Machine Readable Data Bases", University of Alberta, Edmonton, Master's Thesis.

MILLS, Jack, 1964, The Universal Decimal Classification, Rutgers Series on Systems for the Intellectual Organization of Information, v. 1, Graduate School of Library Service, Rutgers,New Brunswick, New Jersey.

OHMAN, E., OLIVECRONA, C., 1971, "Some Notational, Hierarchic and Syntactic Problems in connection with Concordance between UDC and Thesauri", Symposium on UDC in relation to other Indexing Languages, Herceg Novi, Yugoslavia.

PERREAULT, Jean M., 1969, An Introduction to UDC: A Programmed Text, London, Bingley.

RIGBY, M., 1970, "Computers and the UDC: A Decade of Progress", prepared for FID/CCC-M and 35th FID Congress, Buenos Aires.

SAVELJEVA, A.N., 1971, "On Revising the UDC Schedules, Scientific and Technical Information", Nauch-Tech. Inf. 2, U.S.S.R., Ser. 1, No. 8

SCRIBOR, E., 1971, "UDC as a Classification Connecting Various Information Languages", Current Problems in Information and Documentation, Central Institute of Scientific, Technical and Economic Information, May-June, Poland.

STUEART, R.D., 1971, "An Analysis of the UDC as a Term System for Nuclear Science and Technology", Library Resources and Technical Services, v. 15, no. 3, p. 399-411.

WELLISCH, H., 1970, The Universal Decimal Classification: A Programmed Instruction Course, School of Library and Information Services, University of Maryland.

----------, 1971, "Organisatorische Neurodunung des DK-Systems" (Organizational Renewal of the UDC System) Nachrichten für Dokumentation 22 nr. 2, p. 55-63.

----------, 1972, "A Flow Chart for Indexing with a Thesaurus", Journal of American Society for Information Science, v. 23, no. 3, p. 185-194.

WILSON, T.D., 1972, "The Work of the British Classification Research Group" Subject Retrieval in the Seventies, School of Library and Information Services, University of Maryland, Greenwood, p. 62-71.

WRIGHT, J.E., 1969, "Steps in Development of BS 1000", ASLIB Proceedings, v. 21, no. 5.