Reports and Communications

Unified Medical Language System: Progress Report[1]

1. Purpose

Staff of the National Library of Medicine (NLM) and contractors supporting the Unified Medical Language System (UMLS) met on Sept.16-17, 1987 to:

- define the probable features of the UMLS more clearly
- review work done over the last six months
- develop plans for the coming year's research efforts, and
- identify UMLS related products likely to be available by Sept.1988.

As described in previous articles of the "National Library of Medicine News" (April-May 1987, Sept.1986), the purpose of the UMLS is to overcome the lack of precise link among related biomedical information in different automated systems. The goal is not to impose a single medical vocabulary on all users and systems, but to make the myriad of classifications of medical knowledge invisible to the user while providing a single logical path to a broad range of biomedical information sources.

The figure below represents the principal elements of the UMLS as currently envisioned. There will be two categories of UMLS components: knowledge sources and functional features. Direct users of computer applications will be able to select individual components of combinations of components to interpret questions, identify appropriate information sources, translate queries into a variety of different target vocabularies, and retrieve, evaluate, and organize information relevant to a particular query.

2. UMLS Knowledge Sources

The UMLS will contain at least two new knowledge sources: a Metathesaurus and an Information Sources Map. Encompassing or complementing the current Medical Subject Headings or MeSH® file, the Metathesaurus will store medical concepts and terms in a canonical form to which multiple existing vocabularies and classification systems will be mapped. The Metathesaurus will include concept definitions and will represent a variety of relationships among terms, in addition to the hierarchical relationships in MeSH.

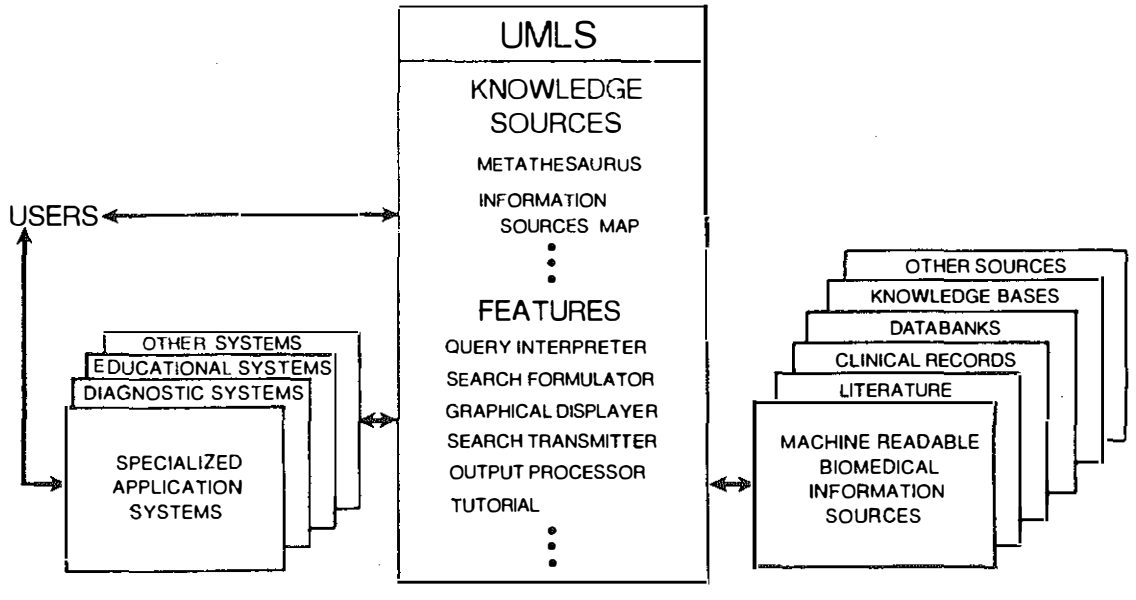
The Information Resources Map will include information about the scope, location, vocabulary, syntax rules, and access conditions of medically relevant databases of all kinds. The map will be used by other UMLS components or externally developed applications systems to determine the most appropriate information sources for a particular user query.

3. UMLS Features

A variety of UMLS features will interact both with the UMLS knowledge sources and with the broad array of machine-readable files of medically relevant bibliographic and factual information. The Query Interpreter will interact with the user to verify that the user's real intent has been understood. The Graphical Displayer may well be invoked in this verification step to allow the user to see and traverse the relationships among terms and concepts in the Metathesaurus. The display package will also be available for direct use by searchers who wish to explore concepts in the Metathesaurus.

An interactive search formulator, which may also make use of the Graphical Displayer will help users refine their queries, will consult the Information Sources Map to identify appropriate targets for the queries, and will use the Metathesaurus to translate the question into the vocabulary understood by appropriate information sources. A search transmitter will transmit the formatted

1 We gratefully acknowledge permission to reprint this report from National Library of Medicine News, Oct.1987, p.4-6.



UNIFIED MEDICAL LANGUAGE SYSTEM (UMLS)

search statements to relevant computer systems and receive the output from these systems. An output processor will merge, organize, evaluate, and rank the information retrieved according to its relevance to the user's query.

4. September Meeting: Contractor Reports

At the September meeting, contractor personnel and NLM staff reviewed progress on several projects which will either contribute to the design and development of these eventual UMLS components or increase our understanding of the medical information sources with which the UMLS must interact.

The University of California, San Francisco (UCSF) reported that structured machine readable descriptions for over 800 diseases would be available shortly for use by NLM and other research groups. UCSF has begun to analyze and classify the language used to describe various diseases to identify patterns of word and phrase usage.

Brigham and Women's Hospital described progress in the development of a semantic network structure for the UMLS and displayed preliminary tools for building and displaying this structure. Carnegie Mellon University and Massachusetts General Hospital have also contributed substantially to the semantic network effort over the past six months.

Yale University School of Medicine has moved from a study of domain independent semantic relationships, which may be useful in refining bibliographic retrieval, to focusing on domain specific semantic relationships. Yale is currently developing a test system which will permit users of a database of summaries of cases in a restricted domain to retrieve MEDLINE® citations related to clinical questions arising from those cases.

The University of Pittsburgh and the University of Utah have extended their work on a frame-based system for linking different clinical findings vocabularies by expanding the number of generic frames in the system and developing automated tools for mapping specific vocabulary terms to these generic concept representations.

Massachusetts General Hospital has expanded Micro-MeSH, its microcomputer-based tool for searching and displaying MsSH's hierarchical structure, to present information about allowable main heading/subheading combinations and to display information about other than hierarchical relationships among terms. MicroMeSH also contains features which assist users in formatting and running queries on MEDLARS and other database systems.

In the area of studies of the charateristics of medical information sources, UCSF reported that its comparative analysis of terms in standard medical education testing instruments with terms in MeSH and the medical literature had been extended to cover portions of the American Board of Family Practice examination.

Carnegie Mellon University discussed preliminary work on the analysis of characteristics of journal abstracts which can help or hinder bibliographic retrieval.

5. NLM Contributions

NLM itself has made progress toward the development of a test collection of user queries and corresponding citations with relevance judgments. This collection, as well as a recently established special one year subset of the MEDLINE file, will be used to evaluate potential system improvements suggested by the UMLS research effort.

The Library has also worked on a number of immediate improvements to current information services. The bulk of the mapping of the PDQ® (Physician's Data Query) Thesaurus to MeSH has been completed. NLM will shortly begin testing the use of the PDQ/MeSH map as a means of automatically retrieving MEDLINE citations relevant to the factual information a user has retrieved in PDQ. NLM has also mapped the most heavily used terminology in GenBank® , the National Institute of Health's gene sequence databank, to the MeSH vocabulary. The GenBank mapping has been incorporated in the 1988 version of MeSH. Gradual expansion and transformation of the current MeSH file will be one vehicle for moving toward the planned capabilities of the Metathesaurus. NLM has also created a special MeSHLINK file in which to store the mapping of external vocabulary terms which are equivalent to a combination of MsSH terms or a main heading/subheading combination. Such mappings cannot be accomodated in the MeSH file as presently constituted.

GRATEFULMED™, NLM's microcomputer-based search interface, already contains several of the desired UMLS features in a preliminary form. These will be enhanced in successive versions of the program. For example, as an aid to integrated access to different types of databases, the next version of GRATEFULMED™, currently undergoing testing at NLM, will include a "search engine" feature which will make it much easier for developers of special purpose medical information systems to embed access to the MEDLARS® databases in their application programs. The AI/RHEUM expert diagnostic system has already been modified to permit its users to employ GRATEFULMED to retrieve MEDLINE citations relevant to particular diagnoses.

Betsy L.Humphreys

Ms.B.L.Humphreys, Deputy Assoc.Dir., National Library of Medicine, Bethesda, MD 20894, USA

Systematics Association, 1987 Meeting

A symposium on "Prospects in Systematics" had been organized by the Systematics Association from July 1-3, 1987. Further information can be obtained from Prof.R.G.Davies, Dept.of Pure and Applied Biology, Imperial College of Science and Technology, Prince Consort Road, London SW7 2AZ, England.

CSNA-88 Annual Meeting

The Annual Meeting of the Classification Society of North America was to be held on June 16-18, 1988 at the Greenwich Village Campus of New York University in New York City. The plans foresaw that the keynote address was to be presented by Dr.Fionn Murtagh, a senior systems analyst in the European Space Agency at the European Southern Observatory, Munich, FRG on the topic: "Clustering of Astronomical Data: Some New Approaches and Recent Results". Two featured symposia

were on "Applications in Psychology and Sociology (organized by Jeff Tanaka) and "Tree Fitting and Clustering" (organized by Geert De Soete). James Ramsey and Suzanne Winsberg were to offer a workshop titled "How to Work with Splines". Sharon L.Weinberg was the meeting's host. For further information write to her at: Program of Educational Statistics, 933 Shimkin Hall, New York University, New York, NY 10003, USA. Recently Vol.16 (1987) of "Classification Literature Automated Search Service" (CLASS) (144p.) has been published by the Classification Society of North America. Members of the Society receive it as benefits together with the "Journal of Classification" in return for 1988 annual dues of USD 35.-
Applications for memberships should be addressed to the Business Manager, Prof.Glenn W.Milligan, Faculty of Management Sciences, The Ohio State University, Columbus, OH, 43210, USA.

2nd Conference of the International Federation of Classification Societies (IFCS), June 27-30, 1989, USA

This conference is to be held at the University of Virginia in Charlottesville, VA. It is devoted to the presentation of theoretical, methodological, and applied papers on classification, pattern recognition, and related methods of statistics and data analysis in the broad sense. It includes mathematical, statistical, and practical investigations in special fields of knowledge, and the interface between classification and the Information Sciences. Papers are invited for the meeting. Suitable topics include:

1. Classification, discrimination, aggregation and clustering methods.
2. Pattern recognition and image analysis methods.
3. Statistical and probabilistic methods for data analysis and classification.
4. Similarity and distance measures, data quality and reliability.
5. Multidimensional scaling and structure recovery methods.
6. Consensus methods and correpondence analysis methods.
7. Biological taxonomy, genome/molecular sequencing.
8. Analysis and comparison of tree and graph patterns.
9. Artificial intelligence and expert systems for classification.
10. Classification and clustering algorithms and algorithmic aspects.
11. Classification and clustering software for microcomputers and supercomputers.
12. Computer graphics for classificatory problems.
13. Practical applications in fields of biological sciences, information sciences, life sciences, mathematical sciences, medical science, and social sciences.

For further information write to: IFCS-89, Dept.of Mathematics, U.of Virginia, Charlottesville, VA 22903. Proposals for presentation of papers require that an English abstract of at most one page should be sent to: Robert F.Ling, Chairman, IFCS-89 Program Committee, Dept.of Math.Sciences, Clemson Univ., Clemson, SC 29634-1907. Submission of an abstract must be accompanied by a title, keywords, name(s) and institutional affiliation(s) of author(s), and the name of the contacting author for papers with multiple authors. Deadline for submitting of papers is January 15, 1989. The Federation had held its first International Conference at Aachen, June 29 -July 1, 1987. The Proceedings of this conference have been published recently by North-Holland (Editor: H.H.Bock) under the title "Classification and Related Methods of Data Analysis" (XIII,749p.). They are available for hfl 290.-

Call for Papers: 1989 European Meeting - Psychometric Society

The Meeting will be held at the University of Leuven, Belgium, July 17-19, 1989. Persons wishing to present a paper should send: Title; abstract of 150 words; name(s) and institutional affiliation(s) of the author(s); keywords; indication of how many minutes will be needed (20 min.is suggested, discussion included); electronic mail address, if available.
Submissions should be related to one of the following areas: Test theory, data analysis, multidimensional scaling, statistical methods, structural models, correspondence analysis, measurement theory, multivariate analysis, mathematical models, factor analysis, psychophysical scaling, classification. Deadline for receipt of paper and symposium proposals is March 1, 1989. Papers should be sent to the Chair of the Scientific Committee: Geert De Soete, Dept.of Psychology, University of Ghent, Henri Dunantlaan 2, B-9000 Ghent. A brochure with information on the conference can be obtained from the Chair of the Organizational Committee: Luc Delbeke, Dept.of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium.

British Classification Society

The Annual General Meeting of 1988 took place at Kew on Friday, June 10, 1988 at the Commonwealth Mycological Institute, starting 10:30 a.m. A tour of Kew Gardens with a visit to the Herbarium had been foreseen. Dr.S. BLINKHORN of Hatfield Polytechnic was to speak on "Classification in Psychology". Thereafter, the Annual General Meeting (AGM) was to follow.

In 1987 a successful meeting had taken place together with the Geological Information Group at the British Museum (Natural History) in London. The following nine papers were to be presented, rounded up by introductions to the meeting by W.T.C.SOWERBUTTS from the Geological Information Group and J.C.GOWER from the British Classification Society in the beginning and an open discussion on "Classification and geology" at the end. The papers: G.ROSS: What do we expect of classification schemes? - A.WOOLEY: Evolution of petrological classification schemes. - D.D.HAWKES: Expert systems for rock classification. -J.HOLLIS: Geological considerations in classifying soils. - W.B.HARLAND: Classification and geological time. -P.R.GRANT, R.PEARSE, K.BARTRAM: An expert system in Palaeopalynology. - G.M.VASEY: The classification of Carboniferous non-marine bivalves: systematic versus stratigraphy. - D.T.DONOVAN: Ammonite classification. - S.U.HULTBERG, M.D.SIMMONS: Classification of organisms: palaeontology versus biology/industry.

In March 1989 a Joint Meeting of the British Classification Society and the Chemometrics Society has been foreseen. For any further information please contact: Dr.J.C.Gower, Statistics Department, Rothamsted Experimental Station, Harpenden, Herts. AL5 2JQ, England.

German Society for Classification: Call for Papers of 13th Annual Conference

This Conference will take place at the University of Augsburg from April 10-12, 1989. On the many aspects of the topic "Contents-related and Numerical Analysis of Data" proposals for papers are invited, especially regarding the 27 thematical areas which have been grouped into the following 6 blocks: (1) Data, information and knowledge; Hierarchical and nonhierarchical data structures; Qualitative and quantitative data; Relational, longitudinal and spatial data; Formal and contents-related concept analysis; Incomplete and insecure information. - (2) Data analysis and scaling; Numerical classification, cluster analysis; Stochastic methods of data analysis; Evaluation and comparison of classifications; Pattern recognition. - (3) Software for data analysis and classification; Decision-supporting systems; Computerization of classification systems. - (4) Systematics and classification of knowledge; Knowledge representation and indexing languages; Library classification and information retrieval; Commodity and product classification and description; Classification in online retrieval; Analysis and structuring of concept-systems and classification systems. - (5) Data analysis in market research and psychology; Classification and data analysis in economics; Analysis of social science data. - (6) Biological taxonomy and sequential analysis; Classificatory problems in medicine; Applications in computer science and in engineering; Analysis of linguistic data. Whosoever wants to participate should write, possibly until Nov. 15, 1988 to Prof. Dr. Otto Opitz, Lehrstuhl für Mathematische Methoden der Wirtschaftswissenschaften, Universität Augsburg, Memminger Str. 14, D-8900 Augsburg, Tel.0821-598-385.

Proceedings of the German Society for Classification

The proceedings volume of the 11th Annual Conference in Aachen, June 29 - July 1, 1987 has recently been published (Editors: H.-J.Hermes, J.Hölzl) under the title: Wissensorganisation im Wandel. Dezimalklassifikation, Thesaurusfragen, Warenklassifikation. (Knowledge organisation in a changing world). It is available from Indeks Verlag, Frankfurt in the Series "Studien zur Klassifikation", Vol.18. The proceedings volume of the 12th Annual Conference, held March 17-19, 1988 in Darmstadt, entitled "Classification and Order" will be edited by R.Wille and is to be published by the end of 1988 as Vol.19 of Studien zur Klassifikation, Indeks Verlag.

Chemical Nomenclature and Terminological Science

On June 24, 1988 the terminology-bureau Topterm at Amsterdam organized a Symposium on "Chemical Nomenclature and Terminological Science: a useful blending of two disciplines". In the field of chemistry it is recognized that exactness in communication is essential to effective progress. A good system of nomenclature is a fundamental necessity for exact communication. Chemical nomenclature covers a much wider field than simply assigning names to chemical compounds; it embraces quantities and units, symbols, acronyms, abbreviations and even pronunciation. Even though chemists frequently communicate by means of structural formulae, the role of nomenclature and terminology is as important to chemistry as in any other discipline. In addition, the language of chemistry or chemical nomenclature is not only used by chemists. All kinds of scientists -biologists, physicists, environmentalists, engineers, medical personnel, journalists, lawyers, educators, publishers, even politicians - have a need for a chemical language. The Symposium aimed to bring together representatives from science, industry and government. The list of speakers included: Prof.Dr. F.ALDERWEIRELDT, University Centre, Antwerpen; Dr.K.L.LOENING, Chemical Abstracts Service, Columbus, OH; Dr.L.MAAT, Technical University Delft; J.R.MENGARDUQUE, Ministry of Foreign Affairs, The Hague; J.ROMPAY, Head of the Analytical Research Department of Janssen Research, Beerse, Belgium; Drs.W.B.SONNEVELD, Managing Director of Topterm; and Drs.A.J.VERVOORN, Technical University Eindhoven. For further information turn to Topterm, Postbus 7971, NL-1008 AD Amsterdam.

Concept Analysis and Artificial Intelligence

This is the topic of a Meeting of the Special Interest Group on Concept Analysis of the German Society for Classification to be held from Oct.6-8, 1988 at the Institute for Informatics, Technical University of Clausthal-Zellerfeld, FRG. For further information contact: Prof.Dr.W.Lex, Institut für Informatik, TU Clausthal, Erzstr. 1, D-3392 Clausthal-Zellerfeld.

E-mail Directory of Classifiers

William H.E. Day, (Computer Sci., Memorial Univ. of Newfoundland, Canada) is compiling a directory of network user addresses of researchers who are interested in classification, clustering, phylogenetic reconstruction, or related methods of data analysis. (Listings of graduate students are welcome!) On 12 July the directory contained 119 entries. Listing in this directory is NOT restricted to members of particular societies or to residents of particular countries. The directory usually doesn't include addresses for sending mail from one network domain to another since such addressing conventions usually depend on the originating node.

If you wish to be listed in the directory, please send to Dr. Day (at any of the e-mail addresses below): your full name as you wish it to appear in the directory,

e.g. Day, William H.E.

and the network user addresses by which you are known ALONG WITH identifications of the corresponding message domains,

e.g.	Bitnet/EARN/Netnorth:	whday@mun
	CDNnet:	whday@garfield.mun
	UUCP:	garfield!whday

If you agree to permit this information to appear in the directory, then Dr. Day will send you current versions of the directory. He sends the directory only to persons whose names appear in it.