Roy Rada
National Library of Medicine, Bethesda, MD, USA

Connecting and Evaluating Thesauri: Issues and Cases

Rada, R.: **Connecting and evaluating thesauri. Issues and cases.** Int: Classif. 14 (1987) No. 2, p. 63–69, 36 refs.

Connecting and evaluating thesauri is an important task for the systematic development of better information retrieval systems. Connecting thesauri includes not only determining when terms in different thesauri are the same but also detennining what kinds of relationships can be transferred from one thesaurus to another. This paper first presents issues in connecting and evaluating thesauri. Various experiments in connecting a particular thesaurus, the Medical Subject Headings, with other medical thesauri are described. In these experiments, similar terms in two thesauri are recognized and then differences in two thesauri are exploited to create more powerful thesauri. Part of the evaluation requires the thesaurus to support automatic indexing and retrieving of documents.

(Author)

1. Introduction

Thesauri are fundamental components of many information retrieval systems. Synonyms to the term "thesaurus" are the terms "classification structure", "controlled vocabulary", and "ordering system" (1). Connecting of thesauri can help users get information from different information sources. Evaluating the success of such connections requires understanding the functions of the thesaurus within the information system.

A thesaurus is a set of concepts in which each concept is represented with at least synonymous terms, broader concepts, narrower concepts, and related concepts (2). A term is a word or sequence of words that refers to an atomic concept within a given domain of discourse. For instance, "shortness of breath" may be considered a term when discussing symptoms of patients. Each concept may also be associated with one term that serves as the name of that concept and which will here be called the *concept main term*.

Each concept in a thesaurus can be viewed as a frame in a frame-based system (3). Definitions of the slots in each frame of a thesaurus include (2):

1. *Broader.* This relation can mean:
a) Class inclusion, such as neoplasm is a disease.
b) Whole-part, such as hand is a part of arm.
c) Other connected concepts, such as "characteristic curve of an electron tube" is *broader-than* "electron tube".
2. *Narrower.* This relation includes the reverse relations of those listed for *broader.*
3. *Related.* Two concepts are considered to be *related* if they are related but neither is *broader than* the other. *Related* may be used to identify terms that are related to each other from a certain point of view, such as usage, action, or process.
4. *Synonymous.* Synonyms are terms that mean the same thing.

Other relations could be introduced (1), but for the sake of simplicity, the standards for thesauri restrict the relations to little more than the above.

Connecting two thesauri T_1 and T_2 allows a searcher to see directly or indirectly the contents of two thesauri or the documents to which the thesauri point without having to master the contents of each thesaurus. For this paper, "connecting" is considered broader-than "mapping" but synonymous with "merging". In mapping T_1 to T_2 a new slot has to be entered for each concept in T_1 which slot can be called the *nearest* slot. This *nearest* slot points to the concept(s) in T_2 which are semantically near the concept in T_1 (see Figure 1). In "merging" T_1 with T_2 the *nearest* slot is not needed but changes may be made to the values for the *synonym, broader, narrower*, and *related* slots for all concepts.

2. Mapping and Merging

2.1 Mapping

Concepts can be mapped between two thesauri with a variety of tools. First, direct lexical matching between concept main terms can be performed. Secondly, knowledge about the syntax or the morphosemantics of main terms can be employed. Finally, the knowledge in the relationships within the thesauri themselves can be the basis for sophisticated mapping of terms from one thesaurus to another.

Direct lexical matching of main terms can be done with common computer software tools. With readily available commands in computer editors or operating systems one can efficiently ask whether two strings or terms are identical or have substrings in common. The matching of substrings can accommodate certain simple syntactic and morphosemantic matchings.

2.1.1 With Knowledge Outside Thesaurus

Knowledge about syntax and morphology of terms is not explicit in typical thesauri. This type of knowledge is, however, available in the world and can be used to advantage in mapping thesauri. Furthermore, it should be possible to study a typical thesaurus and infer syntactic and morphosemantic information from the terms and relations in the thesaurus.

concept in T_1

main term
synonymous tenn(s)
broader concept(s)
narrower concept(s)
related concept(s)
nearest concept(s) in T_2

Fig. 1: For mapping T_1 to T_2 concepts in T_1 may be augmented with a *nearest* concept slot that points to the nearest or most similar concept(s) in T_2.

A variety of rearrangements of terms can be performed to determine whether a kind of syntactic match exists between two terms. Two examples illustrate the type of rewriting that can facilitate connecting two main terms:

- x of y ⟷ yx
  (as in shortness of breath ⟷ breath shortness)
- x, y ⟷ yx
  (as in cancer, lung ⟷ lung cancer).

Here "u ⟷ v" means that u is a *nearest* concept for v. There has been substantial work reported in the literature about this kind of syntactic connecting of thesauri under the title of switching vocabularies (4, 5) or phrase rewrite systems (6).

After decomposing the words within two terms into their components it is possible to do matching based on these components. Such matching may succeed where direct matching fails. For instance, start with the two terms "hypertension" and "high pressure". If the term "hypertension" has been decomposed into the components "high" and "pressure", then "hypertension" would map to "high pressure". Such connecting of terms has been widely employed in the processing of noun phrases in medical records and their translation into terms from medical classification schemes (7, 8).

2.1.2 Using Thesaurus Structure

A thesaurus is itself rich in information which can be used to help drive mapping. The synonymous relation can naturally assist in mapping. Assume that two thesauri, T_1 and T_2, are available. T_1 includes a concept main term called x which has a synonymous term called y. T_2 includes a concept main term called z which has a synonymous term called y. By going through the synonymous term y, an algorithm can connect x and z (9).

The *broader* and *narrower* relations in thesauri can also be used in mapping terms. Assume that T_1 has relationships for a term x that are the same as the relationships that T_2 has for a term y (see Figure 2). Based on the similarity in the neighborhood of x and y, it seems reasonable to conclude that x ⟷ y. This kind of determination can in theory be made arbitrarily complex to take into account all of the information in the thesaurus in the course of deciding the degree of similarity between any concept x in T_1 and y in T_2 (10).

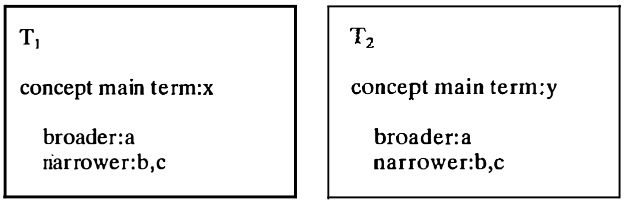

Fig. 2: Thesaurus T_1 has the same structure around concept x that thesaurus T_2 has around concept y. Accordingly, x and y may be suspected to be synonymous.

2.2 Merging

The determination of *nearest* concept in mapping two thesauri avoids certain decisions which could lead to more accurate and powerful thesaurus connections. Consider two thesauri such that in T_1 concept b is *narrower* than concept e and concept e is *narrower* than concept a, while in T_2 concept b is *narrower* than concept a (see Figure 3). How should T_1 be augmented to fill the *nearest to* T_2 concept slot for concept "e". Certain advantages in retrieval could accrue by showing in T_1 that "e" is narrower than "a" in T_2. Otherwise, one might have to say that "e" is nearest to "a" and "b" in T_2.

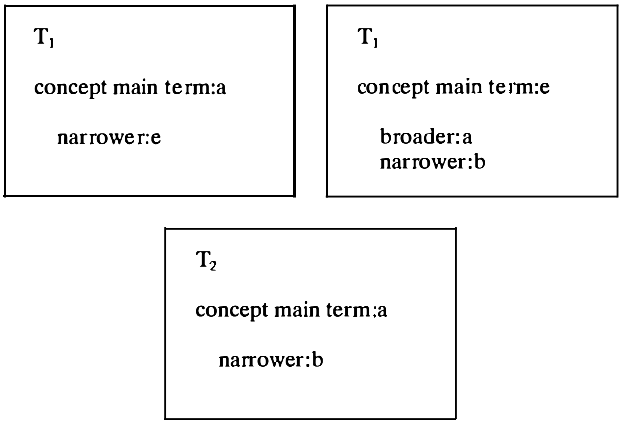

Fig. 3: Concept e in T_1 belongs between concept a and b in T_2.

The process of merging thesauri involves two crucial steps: finding similarities and exploiting differences. Finding similarities involves steps like those in mapping. In the first step, terms in T_1 are mapped to terms in T_2. Exploiting differences can involve a number of principles, but two which are most germane to thesauri could be called "learning by hierarchical transitivities" and "learning by analogy". Learning by hierarchical transitivities involves grafting a subtree from one thesaurus into another thesaurus (see Figure 4). Learning by analogy involves determining that similar terms in two thesauri are related in one but not the other thesaurus and then copying the relationship into the thesaurus that was lacking it. For instance, if in T_1 concept a has a causal relationship to concept b but the same concepts in T_2 have no direct connection, then it is reasonable to hypothesize that T_2 could be accurately augmented by copying the causal relation between concepts a and b from T_1 into T_2.

3. Evaluation

One hypothesis behind the effort to connect thesauri is that after T_1 is connected with T_2 that the resultant T_{1+2} will be better than T_1 alone. Demonstrating with a repeatable experiment that the connection of T_1 and T_2 has led to a better thesaurus is not necessarily straightforward but is facilitated by the fact that many computerized Information Retrieval Systems (IRS) are in routine use and depend heavily on thesauri. The value of a thesaurus is reflected in the value of the IRS of which the thesaurus is a part.

IRS performance characteristics should be related to the benefit to the user of the IRS. Connecting thesauri

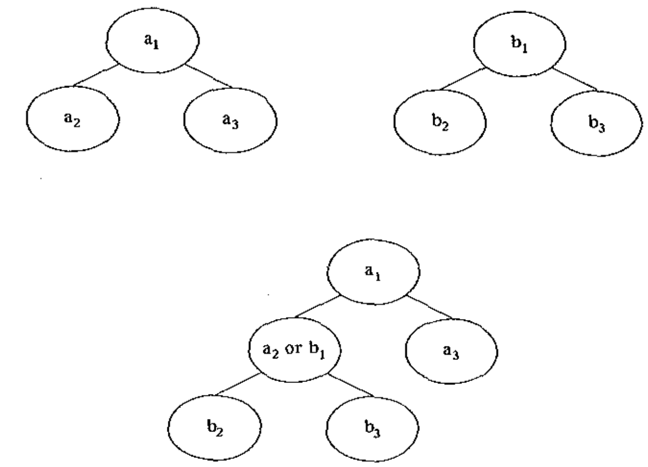

Fig. 4: Above – Thesauri T_1 (represented with terms a_i) and T_2 (represented with terms b_i). Below – Merged thesaurus formed by joining T_1 and T_2 at the common nodes a_2 and b_1.

could help all the performance characteristics of the IRS and thus improve user benefit. One way to evaluate the user benefit is to measure the performance characteristics of the IRS before and after connecting thesauri. Such experiments could survey actual users and ask them for feedback about completeness of retrieval, speed of retrieval, and such (see Figure 5) (9).

A thesaurus may be seen as serving two principal functions in an IRS (9):

- assisting indexers and searchers in the choice of appropriate terms and
- facilitating inclusive searches (if term x has y as a narrower term, then an inclusive search with x automatically includes y).

Evaluation of thesauri along these two dimensions can be done with or without computer assistance. For example, human indexers can be asked to find appropriate index terms from T_1 and T_{1+2} and to then say whether T_1 or T_{1+2} is better.

One way to avoid the subjectivity of asking people what they like about indexing or retrieval results is to axiomatize the desirable properties of a good thesaurus. These axioms can themselves be based on well-known psychological rules. A standard method of portraying the hierarchy in a thesaurus to a user is to list for any given *main concept* all the *narrower* concepts in a menu. As a novice searcher traverses a thesaurus in search of appropriate query terms, the searcher's short-term memory limitations make it desirable that each concept have a handful of narrower concepts so that the menu display is not too sparse or too cluttered (11, 12). Given that one can specify an ideal number of narrower terms for each concept in a thesaurus, a quantitative assessment of a thesaurus can be made of the extent to which the thesaurus meets that ideal. Thus T_{1+2} would be better than T_1, if T_{1+2}'s branching factor was closer to the ideal than T_1's branching factor was.

One can prove that document retrieval through two merged thesauri may lead to better retrieval than retrieval through one thesaurus (13). Assume that the *narrower* concepts of concept x in thesaurus T_1 are made

Information System Characteristic	Impact on Decision-Making	Processing Cost
completeness of information	+	
novelty of information to user	+	
appropriateness of information to user's background	+	+
concise (without redundancy) information		+
speed of retrieval		+

Fig. 5: Some standard factors to use in evaluating an information retrieval system.

narrower concepts of concept y in thesaurus T_2. A search that uses only concept y may retrieve fewer relevant documents than a search which uses y and the *narrower* concepts of x. To the extent that the concepts have been used in indexing documents and one only searches the indexed representation of documents, this argument for improved retrieval may fail. If the document space D_2 has been indexed with concepts from T_2 but not T_1, then to get more recall of documents from D_2 after merging T_1 with T_2, it might be necessary to search for strings of characters. For example, a search could be performed with the concept main terms of the *narrower* concepts of x that have been taken from T_1 and added to T_2. This search would treat a concept main term from T_1 as a string of characters and find occurrences of this string in the text of documents from D_2.

4. Work with a Particular Thesaurus: MeSH

4.1 The System

The National Library of Medicine (NLM) has long been concerned with the development, maintenance and improvement of document retrieval systems (14). NLM is responsible for MEDLINE, a computerized, bibliographic listing of a large segment of the documents in the biomedical, periodical literature (15). Each bibliographic reference to a document is associated with a set of indexing terms from a thesaurus called the Medical Subject Headings (MeSH) (16). A trained indexer scans a document and assigns indexing terms from MeSH based on a set of rules.

MeSH brings the vocabulary of the indexer and searcher into coincidence. A flowchart of the MEDLINE system emphasizes that documents and queries are encoded into MeSH terms (see Figure 6). A search on MEDLINE can be performed for documents represented by indexing terms satisfying any Boolean combination of terms in a query. Millions of *documents* hand-encoded into MeSH are stored on the computer. Thousands of *queries* hand-encoded into MeSH arrive each day from users around the world.

MeSH consists of a set of heading arranged in a 9-level hierarchy. Near the top of the hierarchy those terms include "anatomy", "disease", and "chemicals". Terms narrower-than "disease" include "neoplasm", "immune disease", and "infection". There are about 15,000 main headings in the primary structure of MeSH. Included in the set of headings which represent documents and queries are another 50,000 headings within a special thesaurus of chemicals (17).

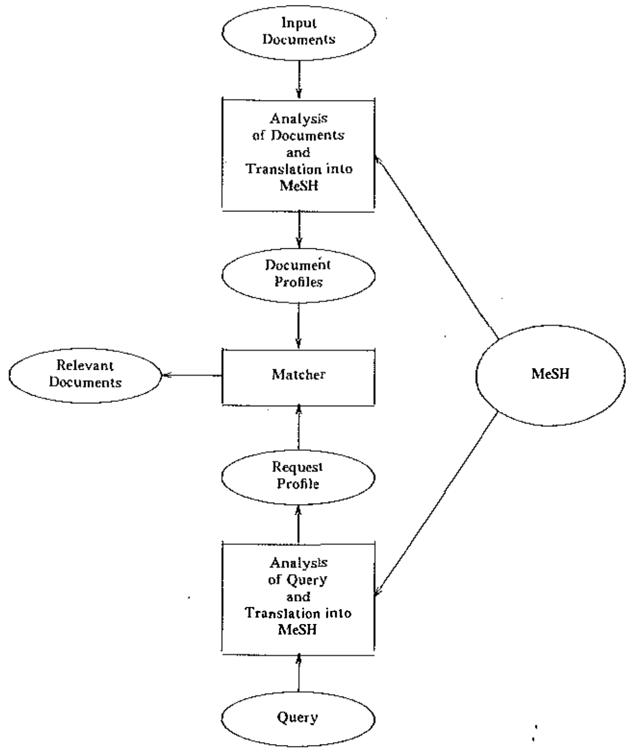

Fig. 6: Role of MeSH in storage and retrieval of documents from MEDLINE.

4.2 Office Use

NLM is trying to extend the use of MeSH by connecting it to other thesauri and information systems. The Library of Congress maintains a massive thesaurus, called the Library of Congress Subject Headings, and a mapping has been made between that thesaurus and MeSH. This mapping is to allow people searching for documents through either the computerized Library of Congress system or MEDLINE to be able to access material simultaneously from the other system (18). NLM is also trying to connect the genetics database, called GENBANK, to MEDLINE. Part of this connection has involved a mapping between the keywords of GENBANK and MeSH. GENBANK is the world's best known computerized source of genetic sequence information (19).

At Harvard Medical School the students are being introduced to a new mode of education that focuses on tutoring, self-pacing, and computers. Each student is provided with a personal workstation that facilitates communication with other students, with faculty, and with databases. The students are expected to take advantage of the computer to help themselves organize the vast amount of information that goes along with a medical education. MeSH is being explored as a tool to help this organization of information. Researchers at Harvard have developed a computer program that runs on a microcomputer and which provides an environment wherein the user may explore the MeSH vocabulary by browsing its hierarchical structure (20). The system is being augmented with terminology and techniques that make it increasingly useful to students who want to learn about medical care and to organize their own library. The medical school staff index all lectures, laboratory exercises, and patient cases with a controlled vocabulary based on MeSH.

NLM has sponsored Integrated Academic Information Management System projects at several institutions (21, 22). Baylor College of Medicine (BCM) is part of this project and is using MeSH in its office automation endeavors. BCM has about 800 microcomputers, 70 SUN workstations, and 5 mainframe computers on a network. Two Britton-Lee Intelligent Database Machines facilitate data access across the network. MeSH will be stored on the Database Machines and will be used to help classify messages (23).

MeSH has been placed on several computers at NLM for the exploration of graphic interfaces to MeSH. MeSH is displayed on a XEROX 1108 with the assistance of the programming package called GRAPHER. The user can enter a term and be graphically shown the hierarchy of terms around it or can choose to traverse the thesaurus from top to bottom by mouse-activating terms on the screen. MeSH has also been stored and graphically displayed on a MacIntosh and an IBM-PC AT.

4.3 Knowledge Building Experiments

The strategies for semi-automatic augmentation of MeSH have focused on finding the similarities between MeSH and other thesauri and then exploiting the differences (24). The kinds of changes to MeSH that have been considered include the adding of terms, the adding of relationships, and the readjusting of relationships. Extensive tests of the value of the augmentations have been performed by using MeSH as part of algorithms for

1) evaluating the similarity between documents and queries and
2) automatically indexing document titles into MeSH.

One large-scale experiment for the automatic addition of terms to MeSH referred to the Systematized Nomenclature of Medicine (SNOMED) (25). SNOMED is a 50,000 main term thesaurus that is used in the indexing of parts of the patient record (26). The strategy for augmentation was to find two main terms that were the same in both thesauri. Then the children of the SNOMED main term were added to MeSH as children of the MeSH main term (27). To evaluate this merge, automatic indexing of titles of medical journal articles was done. No improvement in indexing was observed, but this was shown to be true, in part, because the terms from SNOMED that were not already in MeSH tended not to occur in the titles of journal articles.

In experiments with the Computing Reviews Classification Structure (28) and MeSH the merge strategy allowed nodes to be inserted between other nodes as a function of their position in both thesauri. This merge played a key role in the subsequent construction of a thesaurus for the field of medical informatics (29).

Thesauri are rich in hierarchical relationships but poor in other kinds of relationships. In the medical domain there are several computerized knowledge bases which are rich in non-hierarchical relationships. One such knowledge base, called Current Medical Information and Terminology (CMIT), gives for each of about 4,000 diseases the etiology, signs, symptoms, laboratory findings, and more (30). The addition of such information to MeSH could be useful in many computer tasks, such as expectation-based parsing. If a parser first detects the etiology and symptoms of a disease in a paper, then the parser might expect that that disease will be discussed in the paper.

Experiments have been done to add CMIT relations to MeSH. The method found similarities between concepts in MeSH and CMIT and then added CMIT relationships to MeSH. For instance, since in CMIT "granuloma" is a *pathological finding* of "rheumatoid arthritis", the MeSH concepts "rheumatoid arthritis" and "granuloma" were connected by the relationship *pathological finding*. These experiments involved first manually translating CMIT phrases into MeSH concepts, then automatically connecting the appropriate pairs of MeSH concepts with relationships from CMIT (31, 32). To extent these efforts at adding the relationships of CMIT to MeSH, automatic translation of CMIT phrases into MeSH were attempted (33). Various word frequency and pattern recognition techniques were used to classify CMIT phrases as MeSH terms. Evaluations of the merged MeSH and CMIT revealed problems secondary to an inadequate parsing of the CMIT phrases into MeSH. CMIT is presented in a highly stylized type of medical sublanguage which makes parsing of it very difficult by other than medical professionals (34).

4.4 Consistency Challenge

Certain problems can arise in merging which we don't entirely know how to solve. If one thesaurus says that x is *broader-than* y but another thesaurus says that x and y have the same parent (i.e., are *siblings*), then an inconsistency has occurred. The PDQ thesaurus is used by the the PDQ cancer information retrieval system of the National Cancer Institute. The PDQ thesaurus provides access to current information on cancer treatment, research and prognosis (35). A merge of part of MeSH with part of the PDQ thesaurus was performed (with a method that required locating similar concepts in the two thesauri and then connecting subtrees of one thesaurus to the other thesaurus at the point of similarity). In 7 of 8 cases where similar concepts were identified, the merge occurred without conflict. But the 8th case demonstrated the problem of consistency in a way that had additional ramifications. In MeSH "anal neoplasms" are *narrower-than* "rectal neoplasms". In the PDQ thesaurus "anal neoplasms" and "rectal neoplasms" are both children of the same concept (see Figure 7). Since the PDQ thesaurus is, in general, more specific in the cancer terminology than is MeSH, one heuristic would say to follow the PDQ lead. More careful examination of MeSH reveals, however, another problem. In the anatomy section of MeSH the "anal canal" is listed as *narrower-than* "rectum". Thus to change the MeSH neoplasm terminology by making "anus neoplams" and "rectum neoplasms" *siblings*, without also changing the MeSH anatomy section, would be to introduce a kind of inconsistency between

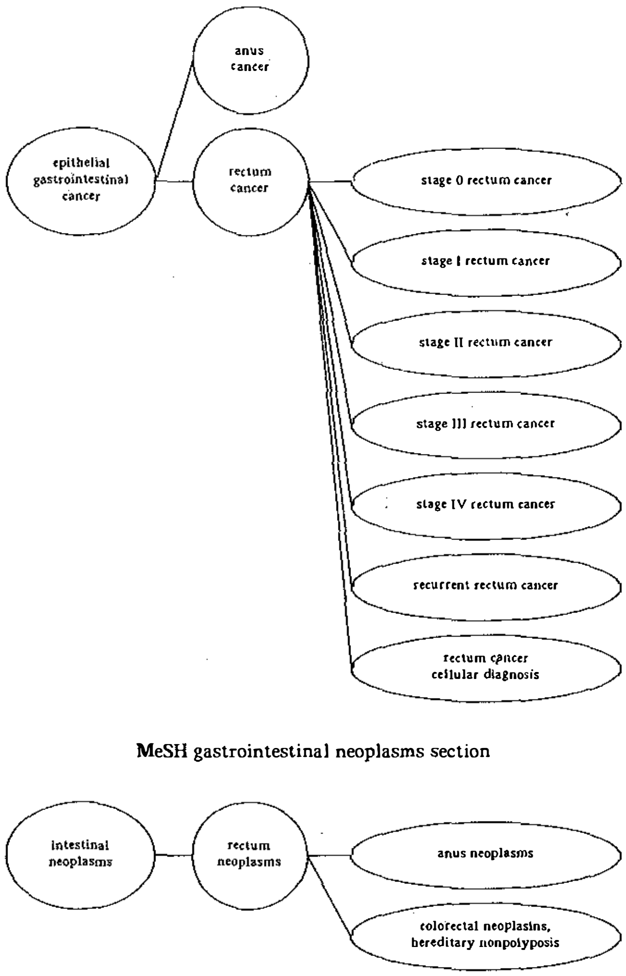

Fig. 7: Conflict between PDQ and MeSH thesauri.

the anatomy and disease sections of MeSH. Heuristics to properly handle such situations are not necessarily straightforward.

4.5 Computer Evaluation

Computer programs can be used to do indexing or searching. To the extent that the performance of the computer program depends on a thesaurus, the performance of the program on two different thesauri is a reflection of the value of the thesauri. Experiments have been used to test the choice of appropriate terms for indexing. Computer programs have been written which take titles from articles off MEDLINE and automatically index them into MeSH terms. These programs rely on MeSH or MeSH connected with another thesaurus (27). The performance of the computer indexer is assessed by comparing its output to the output of the human indexer as it is stored on MEDLINE for each article.

To test the role of the *broader* and *narrower* relationships in a thesaurus, a spreading activation model has been used to reason with a thesaurus in determining the relevance of documents to queries (32). The basic principle is that the number of *broader-than* and *narrower-than* relationships which separate two sets of concepts reflects on the conceptual distance between the sets. The simplest example follows:

- given a thesaurus where x is broader-than y and y is broader-than z,
- given a query represented by x,
- given a $document_y$ represented by y and a $document_z$ represented by z
- the query is conceptually closer to $document_y$ than it is to $document_z$ because x is one relationship from y but two relationships from z.

The computer algorithm ranks documents to a query based on conceptual distance. People are asked to perform the same ranking. The ranking of the computer with T_1 is compared to that of people. Then T_1 is merged with T_2, and the computer again does the ranking but now based on T_{1+2}. This second ranking is compared against the human ranking. If the computer does better with T_{1+2} than it does with T_1, then T_{1+2} is considered to be a better thesaurus.

5. Discussion

Making a thesaurus is typically a labor-intensive job. Furthermore, a new thesaurus often repeats substantial amounts of material already present in existing thesauri (36). Approximately 22 years ago. Clara E. Müller tried to establish a special concordance between the Universal Decimal Classification and some special classification systems without the assistance of a computer. She met with the same frustration as other researchers who embarked on that project. Too many parameters are involved which could hardly be made explicit without computer assistance (1).

As each new document space and thesaurus becomes available, the need for methods that allow users to search multiple document spaces without having to understand multiple thesauri increases. One approach to allowing users to take advantage of multiple thesauri at once is to map the terms of each thesaurus to the terms of the other (6). Furthermore, an information retrieval system with merged thesauri can lead to better information retrieval than an information retrieval system which only maps between terms of different thesauri (13).

While the merging of thesauri T_1 and T_2 would seem to necessarily produce a more powerful thesaurus than either T_1 or T_2 alone, demonstrating this for two, real-world thesauri with a repeatable experiment is not necessarily easy. First, the thesauri have to be stored on the computer in a form suitable for experimentation. Second, some method of using the thesauri has to be elaborated for which the value of that usage can be precisely assessed. Fortunately, information retrieval systems facilitate access to large numbers of queries, documents, and user interactions – all of which may be systematically related to the role of the thesaurus.

For researchers in information science and artificial intelligence thesauri are an attractive topic. Thesauri are more complex than databases but simpler than natural language. They represent a middle ground from which one can hope to gradually build towards an understanding of the knowledge that supports intelligent information retrieval.

Acknowledgements: A host of collaborators contributed to the work reported here. A partial listing in alphabetical order follows: Ellen Bicknell, John Eng, Bruce Greenwald, Steve Lester, Florence Lu, Joseph Maline, and Hafedh Mili. This work was partly supported by NSF Grant ECS-84-06683.

References

(1) Dahlberg, Ingetraut: Conceptual Compatibility of Ordering Systems. Int. Classif. 10, 1, pp. 5–8, 1983.
(2) National Library and Information Associations Council: Guidelines for Thesaurus Structure, Construction, and Use. New York: American National Standards Institute, 1980.
(3) Fikes, Richard, Kehler, Tom: The Role of Frame-Based Representation in Reasoning. Communic. ACM, 28, 9, pp. 904–920, Sept 1985.
(4) Niehoff, Robert: Development of an Integrated Energy Vocabulary and the Possibilities of Online Subject Switching. J. ASIS, p. 3, Jan.–Feb. 1976.
(5) Niehoff, Robert, Mack, Greg: The Vocabulary Switching System. Int. Classif. 12, 1, pp. 2–6, 1985.
(6) Klingbiel, P.H.: Phrase Structure Rewrite Systems in Information Retrieval. Inform. Process. & Management, 21, 2, pp. 113–126, 1985.
(7) Dunham, G., Pacak, M., Pratt, A.: Automatic Indexing of Pathology Data. J. ASIS, March 1978.
(8) Wingert, F.: An Indexing System for SNOMED. Meth. Inform. Med. 25, pp. 22–30, 1986.
(9) Soergel, Dagobert: Organizing Information: Principles of Data Base and Retrieval Systems. Orlando, Florida; Academic Press, 1985.
(10) Bonham, G.M., Nozicka, G.J., Stokman, F.N.: Cognitive Graphing and the Representation of Biomedical Knowledge. Proc. Expert Systems in Government Conf., pp. 397–403, IEEE Computer Society Press, 1985.
(11) Lee, Eric, MacGregor, James: Minimizing User Search Time in Menu Retrieval Systems. Human Factors, 27, 2, pp. 157–162, 1985.
(12) Miller, D.P.: The Depth/Breadth Tradeoff in Hierarchical Computer Menus. Proc. Human Factors Society – 25th Annual Meeting, 1981.
(13) Mazur, Zygmont: Organization of the Inverted Files in a Distributed Information Retrieval System Based on Thesauri. Inform. Process. & Management, 22, 2, pp. 243–250, 1986.
(14) MEDLARS Management Section: Online Services Reference Manual. Bethesda, MD: National Library of Medicine. July 1982.
(15) McCarn, D.B.: MEDLINE: an Introduction to On-Line Searching. J. ASIS, 31, 3, pp. 181–192, May 1980.
(16) National Library of Medicine, Medical Subject Headings Section: Medical Subject Headings, Annotated Alphabetical List, Springfield, Virgina: National Technical Information Service, 1986.
(17) National Library of Medicine, Medical Subject Headings Section: Medical Subject Headings – Supplementary Chemical Records. Springfield, Virgina: National Technical Information Service, 1986.
(18) National Library of Medicine: Unified Medical Language System. Nat. Libr. of Medicine News, 41, 11, pp. 1–4, November 1986.
(19) Burks, C., Fickett, J.W., Goad, W.B., Kanehisa, M., Lewitter, F.I., Rindone, W.P., Swindell, C.D., Tung, C., Bilofsky, H.S.: The GenBank Nucleic Acid Sequence Database. Cabios, 1, 4, pp. 225–233, 1985.
(20) Barnett, G.O.: An Interactive MeSH Evironment. NLM Contract Report PO#467-MZ-600562, August 1986.
(21) Matheson, N.W., Cooper, J.A.: Academic Information in the Academic Health Sciences Center J. Med. Educ., 57, 10 Part 2, pp. 1–93, October 1982.
(22) Warner, H.R., Detmer, D.E., Peay, W.J.: IAIMS Implementation and Administration at the University of Utah. Proc. Tenth Ann. Symp. on Computer Applications in Medical Care, pp. 111–113, IEEE Computer Society, 1986.
(23) Gorry, G.A.: IAIMS Report. NLM Grant Report LM#04478-02, June 1986.
(24) Forsyth, Richard, Rada, Roy: Machine Learning: ExpertSys-

tems and Information Retrieval. London: Ellis Horwood, 1986.
(25) Cote, Roger: Systematized Nomenclature of Medicine. Skokie, Illinois: College of American Pathologists, 1979.
(26) Cote, R.A.: Architecture of SNOMED: Its Contribution to Medical Language Processing. Proc. Tenth Ann. Symp. on Computer Applications in Medical Care, pp. 74–80. IEEE Computer Society, 1986.
(27) Rada, Roy, Darden, Lindley, Eng, John: Relating Two Knowledge Bases: The Role of Identity and Part-Whole. In: The Role of Language in Problem Solving, 2, ed. R. Jernigan, pp. 71–91, Amsterdam, Netherlands: Elsevier Science Publ., 1987.
(28) Denning, D., Minker, J., Parker, A., Ralston, A., Reilly, E., Rosenberg, A., Walston, C., Willoughby, T., Sammet, J., Blum, A.: The Proposed New Computing Reviews Classification Scheme. Communic. ACM, 24, 7, pp. 419–434, July 1981.
(29) Rada, Roy, Calhoun, Edith, Hafedh, Mili, Singer, Sarah, Blum, Bruce, Orthner, Helmuth: A Medical Informatics Thesaurus, Proc. MEDINFO '86, pp. 1164–1172, October 1986.
(30) Finkel, A., Gordon, B., Baker, M., Fanta, C.: Current Medical Information and Terminology. Chicago: American Medical Association, 1981.
(31) Rada, Roy, Humphrey, Susanne, Coccia, Craig.: A Knowledge-base for Retrieval Evaluation. Ann. Proc. ACM, pp. 360–367, Oct. 1985.
(32) Rada, Roy: Gradualness Facilitates Knowledge Refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence, 7, 5, pp. 523–530, September 1985.
(33) Lester, Steve; Rada, Roy: A Method of Medical Knowledge Base Augmentation. Meth. Inform. Med. 26 (1987) No. 1, p. 31–39.
(34) Dunham, George: The Role of Syntax in the Sublanguage of Medical Diagnostic Statements. In: Analyzing Language in Restricted Domains: Sublanguage Description and Processing, ed. Richard Kittredge, Hillsdale, New Jersey: Lawrence Erlbaum, 1986, pp. 175–194.
(35) Kreps, G.L., Maibach, E.W., Naughton, M.D., Day, S.H., Annett, D.: PDQ Usage: Implications for Evaluation. Proc. Amer. Assoc. Medical Systems and Informatics. 5th Annual Joint National Congress, pp. 71–75, 1986.
(36) Dextre, S.G., Clarke, T.M.: A System for Machine-Aided Thesaurus Construction. Aslib Proc, 33, 3, pp. 102–112, 1981.