Semantic Knowledge Representation Project

A Report to the Board of Scientific Counselors

May 14, 1998

Investigators:
Alan R. Aronson, PhD
Thomas C. Rindflesch, PhD
Cognitive Science Branch
Lister Hill National Center for Biomedical Communications

1. Background

Since an overwhelming amount of human knowledge is encoded in natural language texts (as opposed to databases), the grand challenge in information technology is reliable and effective management of this knowledge, and this notion is reflected in a Report of the Board of Regents of the National Library of Medicine, Long Range Plan (1987):

> Processing information faster or more efficiently--which today's technology can easily accomplish--is not sufficient. More intelligent processing, logical aggregation of information, synthesis and analysis, and the development of knowledge systems that serve purposeful ends are needed. That is the fundamental task of medical informatics. (p. 31)

Lindberg (1996) supplements the Board's position by noting the importance of the representation of knowledge in text for biomedical information management.

The practical application of automatic techniques to the task of recovering meaning from text is rooted in the philosophy of language, formal linguistics, and natural language processing as a specialty in computer science. Since antiquity the philosophy of language has concerned itself with the nature of meaning (Waismann 1971). Plato and St. Augustine had essentially the same concept of meaning and language, which can be summarized as: (a) the words of language are names for entities; and (b) the essence of language consists of combining these names. More recently, these notions have been formalized in semantic theory by Montague (1974) based heavily on the work of Frege (see Geach and Black 1960). (Also see, for example, Chierchia and McConnell-Ginet 1990 for an overview of formal semantics.)

Maida and Shapiro (1982) provide one viewpoint which sets the theoretical scene for implementing the above notions. In particular they provide a formal mechanism for representing the relationship between the assertions made in text and the interaction of entities expressed in a domain model of some possible world:

> ...a semantic network models the belief structure of a thinking, reasoning, language using being (e.g., a human). In this case, nodes represent the concepts and beliefs such a being would have. (p. 172)

Three Unified Medical Language System (UMLS) knowledge sources, the SPECIALIST Lexicon, the Metathesaurus, and the Semantic Network (Lindberg et al. 1993, Humphreys et al. 1998) embody the semantic network referred to by Maida and Shapiro and represent concepts and relationships in the biomedical domain. Looked at from another point of view, these UMLS knowledge sources instantiate the Platonic model of meaning in that the Lexicon and the Metathesaurus taken together represent names for things, while the Metathesaurus and the Semantic Network together represent relationships. Natural language processing techniques are then called upon to provide a link between this domain model and text.

Natural language processing is a cross-disciplinary endeavor bringing together linguistics, computer science, and cognitive psychology; it has as its goal automatically interpreting natural language text and thereby recovering the information encoded therein (see Rindflesch 1996 for an overview of current research). Bates and Weischedel (1993) emphasize the importance of domain knowledge as a basis for significant progress in natural language processing effectiveness, while Sowa (1991, for example) discusses a number of important aspects of the interaction of domain knowledge and linguistic analysis. Rosner and Johnson (1992) provide an overview of current research aimed at implementing notions of formal semantics in natural language processing systems, while Saint-Dizier and Viegas (1995) concentrate on lexical semantics in this regard. Zernik (1992) discusses a practical semantic analysis in a particular subject area.

There is currently a considerable amount of interest in natural language processing of biomedical text. Several approaches are being explored to provide reliable automatic analyses which can support practical applications (See, for example, Haug et al. 1995, Hripcsak et al. 1995, Friedman et al. 1995, Rassinoux et al. 1995, Zweigenbaum et al. 1995.).

2. Project Objectives

The Semantic Knowledge Representation Project was recently initiated at NLM in order to develop programs to provide usable semantic representation of biomedical free text by building on resources currently available at the Library.

Access to biomedical information depends on reliable representation of the knowledge contained in text. For significant advances to be achieved a richer representation will be required than is currently available. As an example of the type of enhanced representation we are proposing, (2) contains the semantic propositions which represent some of the information contained in the text in (1).

(1) We used hemofiltration to treat a patient with digoxin overdose complicated by refractory hyperkalemia.

(2) Hemofiltration-TREATS-Overdose
Digoxin-CAUSES-Overdose
Overdose-OCCURS_IN-Patients
Hyperkalemia-COMPLICATES-Overdose

Each of the propositions in (2) is a predication whose predicate (in upper case) is a relation from the UMLS Semantic Network. Each of the arguments is a concept from the UMLS Metathesaurus. The set of propositions in (2) considered as the semantic representation of (1) is not complete; however, it represents the major relationships and concepts contained in the text.

3. Project Significance

We can consider some specific examples of applications that become possible on the basis of enhanced semantic representation of text. A representation of semantic knowledge, such as that given in (2) above, supports automatic indexing for enhanced information retrieval (IR) effectiveness and document profiling. If such representation could be reliably and accurately imposed on free text, improvement in other applications would also be possible: such as question answering (Chakravarthy and Haase 1995), retrieval of images (Chakravarthy 1994), structured browsing and navigation facilities (Wiesener, Kowarschick, and Bayer 1996; Schatz 1995), and knowledge mining, for example, gathering information from the research literature and computerized patient records. In addition, semantic knowledge representation applied to complex Metathesaurus concepts can play a role in the continued development of the UMLS.

4. Methods and Procedures

4.1 Introduction

This project fits within the research context of the Natural Language Systems Program and draws on the resources being developed in the SPECIALIST system (McCray et al. 1993), which provides a framework for exploiting the resources of the UMLS in processing biomedical text. In addition to the Metathesaurus and Semantic Network, the SPECIALIST Lexicon and associated lexical variant programs (McCray, Srinivasan, and Browne 1994) as well as the Knowledge Source Server (McCray et al. 1996) support syntactic analysis and semantic interpretation of free text in the biomedical domain.

Preliminary to the current project, we have developed two programs, MetaMap (Aronson, Rindflesch, and Browne 1994, Rindflesch and Aronson 1994, Aronson 1996) and SemRep (Rindflesch and Aronson 1993, Rindflesch 1995), which work in concert to provide the semantic representation given in (2) above. These programs work well enough to provide proof of concept but require improvement for reliability in the general case.

4.2 Underlying technology

The SPECIALIST system begins analysis of biomedical text by consulting the Lexicon to determine syntactic information for each lexical entry in the input. A stochastic tagger (Cutting et al. 1992) is called to resolve part-of-speech ambiguities, and an underspecified syntactic analysis is produced as the basis for further processing. For example, input text *ablation of pituitary gland* is given the following analysis:

(3) [[head(ablation)] [prep(of), head(pituitary gland)]]

Although noun phrases are correctly identified, this analysis is underspecified in the sense that overall structure is not provided. That is, no commitment has been made to the exact relationship between the two constituent phrases produced. A further example of the characteristics of this type of analysis is given in (4), which is the underspecified analysis of the input text *pancreatic secretory trypsin inhibitor.*

(4) [[mod(pancreatic), mod(secretory), mod(trypsin), head(inhibitor)]]

In particular, note that, although the head of the noun phrase and its modifiers have been identified, no indication is given of the internal syntactic structure of such phrases. It is our hypothesis that this attenuated analysis is sufficient to serve as the basis for usable semantic interpretation.

The next step in processing calls MetaMap to get to concepts in the Metathesaurus. This program takes advantage of syntactic analysis and considers each noun phrase individually as it proceeds. For example, it takes as input the underspecified syntactic analysis of *ablation of pituitary gland* and finds the following Metathesaurus concepts:

(5) Excision, NOS (‘Therapeutic or Preventive Procedure’, ‘Research Activity’)
Pituitary Gland (‘Body Part, Organ, or Organ Component’)

MetaMap accomplishes its task in four steps:

- Variant generation: Each input word (or multi-word item, such as *wood alcohol*) generates a list of morphological variants, synonyms, and (optionally) acronyms/abbreviations plus meaningful combinations of these variants. For example, *aortic* and *arteria aorta* are variants of *aorta*;
- Candidate retrieval: Metathesaurus strings containing one or more of the input words are retrieved as candidates for a mapping. Some candidates for *aorta* (in browse mode) are "Aorta operations," "Aorta, abdominal," and "Aneurysm, aorta, congenital";
- Candidate evaluation: Each candidate is evaluated for how closely it matches the input text according to a function with four components, centrality, variation, coverage and cohesiveness; and
- Mapping formation: Finally, candidates matching different parts of the input text are combined into a single mapping and re-evaluated to compute a total mapping score.

SemRep is called next and depends on both syntactic analysis and the Metathesaurus concepts provided by MetaMap. In addition, it consults the Semantic Network as part of the process of producing a final semantic interpretation. For example, in assigning an interpretation to *ablation of pituitary gland,* SemRep notes the syntactic analysis given for this input and then consults a rule which states that the preposition *of* corresponds to the Semantic Network relation LOCATION_OF, and further notes that one of the relationships in the Semantic Network with this predicate is

(6)	Semantic Type 1:	'Body Part, Organ, or Organ Component'
	Relation:	LOCATION_OF
	Semantic Type 2:	'Therapeutic or Preventive Procedure'

The MetaMap output for this input is then consulted, and it is noted that the Metathesaurus concept for the text phrase *ablation* is "Excision, NOS." The semantic type for this concept is 'Therapeutic or Preventive Procedure,' while the type for "Pituitary Gland" is 'Body Part, Organ, or Organ Component.' Since these semantic types match those found in the relationship indicated by the preposition *of* (LOCATION_OF) and since the relevant noun phrases are allowable arguments of the preposition *of*, (7) is produced as the semantic interpretation for this phrase, where the corresponding Metathesaurus concepts are substituted for the semantic types in the Semantic Network relationship.

(7) Pituitary Gland-LOCATION_OF-Excision, NOS

The preceding methodology serves as the basis for a number of research projects which investigate the use of semantic knowledge representation of text for enhanced management of biomedical information. These projects have often been conducted in collaboration with investigators both in the Library and at other institutions.

4.3 Completed work

MetaMap was initially developed for improved retrieval of MEDLINE citations. The methodology was tested by applying this program to the queries and citations of the NLM Test Collection, replacing text with the Metathesaurus concepts discovered by MetaMap. Retrieval experiments using SMART were performed both on the unmodified test collection and on the MetaMapped version of the collection. The result was a 4% increase in average precision (Aronson, Rindflesch and Browne 1994).

Other information retrieval projects involving the application of MetaMap include its use as a prototype interface to the Information Sources Map (with Lawrence W. Wright), where users' free text queries are expanded with Metathesaurus concepts. Focused use of MetaMap along with a number of rules describing findings in medical text formed the core of the FINDX program, which was developed to identify findings in patient records (Sneiderman, Rindflesch, and Aronson 1995) and MEDLINE abstracts (Sneiderman, Rindflesch, and Aronson 1996). MetaMap also served as the basis for research exploring full-text retrieval in a recent Library associate project. Combined with techniques for hierarchical indexing, MetaMap was applied to a subset of NLM's Health Services/Technology Assessment Text (HSTAT) database. The results are reported in Wright, Grossetta Nardini, Aronson, and Rindflesch, submitted 1998.

In UMLS development, MetaMap has been used (with William T. Hole) for finding unmarked synonymy in the Metathesaurus. This method was applied to four categories of 1995 Metathesaurus strings: disorders, procedures, functions, and findings and aided manual review during preparation of the 1996 edition of the Metathesaurus. Finally, the algorithm underlying MetaMap served as the basis for the approximate matching methodology (Guy Divita) employed in the Large Scale Vocabulary Test (McCray et al. 1997).

4.4 Ongoing projects

The MetaMap Indexing (MMI) Project is part of the Indexing Initiative project. In this project, MetaMap is applied to the task of automatically indexing the biomedical literature, especially MEDLINE citations. Concepts discovered by MetaMap for a given citation are proposed as indexing terms on the basis of a ranking function which emphasizes presence in the title and specificity of the concept. Extensive testing of this methodology has been conducted. As an alternative to this completely automatic approach, semiautomatic indexing experiments are currently being conducted (with James Marcetich and Toby Port) in which NLM indexers are shown MMI-ranked concepts as suggested indexing terms during the process of MEDLINE indexing. An indexing prototype is currently being developed to incorporate the results of several of the constituent projects of the Indexing Initiative, and MMI is one of the projects which will contribute suggested indexing terms for the prototype.

Recent work (including that of Srinivasan 1996a - 1996d) has demonstrated the importance of query expansion based on retrieval feedback for improving retrieval effectiveness when applying statistically-based systems to MEDLINE citations. As an alternative method of query expansion, we have used MetaMap for associating Metathesaurus concepts with the original query. Our experiments show that this methodology compares favorably with retrieval feedback (Aronson and Rindflesch 1997). We are currently pursuing research with Srinivasan which combines the two query expansion techniques.

MetaMap and SemRep together serve as the basis for several ongoing projects which explore the application of focused semantic interpretation to problems in biomedical information management. Sneiderman, Rindflesch, and Bean (submitted 1998) report on the formative evaluation of a program for accurately identifying terminology associated with the coronary arteries as expressed in coronary catheterization reports, while Bean, Rindflesch, and Sneiderman (submitted 1998) discuss two experiments which assess the ability of an enhanced version of SemRep to identify and characterize physico-spatial semantic relationships in these same reports. A separate application (with Larry Hunter) addresses the ability of MetaMap and a special version of SemRep to discover molecular binding relations in biomedical text. Finally, Rindflesch, Aronson, and Hole (submitted 1998) report on the use of MetaMap and SemRep together to process syntactically complex Metathesaurus strings, where the goal is to determine the relationships which obtain among the components of such strings.

5. Evaluation

Evaluation is a central concern in this project and is applied separately to MetaMap and SemRep. Because MetaMap development has been driven by IR considerations, formal evaluation procedures have primarily consisted of standard measurements of its effect on retrieval performance. On the basis of retrieval experiments, a recall-precision curve is computed, and a summary measure (usually average precision) is used to characterize the experimental result with a single number. Comparison with comparable baseline measures shows the degree of improvement due to the experiment. In addition to formal evaluation measures, informal evaluation consisting of a questionnaire given to participating indexers was used in the MEDLINE indexing experiments conducted as part of the MMI project. Finally, the evaluation procedures being developed for the indexing prototype of the Indexing Initiative will serve as an indirect evaluation of the MetaMap indexing methodology.

Evaluation of SemRep is based on the number of semantic propositions retrieved when processing text. Increasing precision by eliminating incorrect interpretations is currently being addressed. In a recent test, SemRep was run on 970 sentences, which were drawn from 70 MEDLINE abstracts (500 sentences) and extracts from four medical textbooks (470 sentences). For these sentences, SemRep identified 843 semantic relationships. Each interpretation was checked by hand to determine whether it accurately reflected the content of the text. The overall precision score was 76% (81% for MEDLINE only, 75% for the textbook extracts). An initial analysis of the etiology of the errors noted that 50% were due to word-sense ambiguity, 20% were due to the rules which map from syntactic phenomena to Semantic Network relations, and 30% were due to incorrect identification of syntactic arguments. Addressing these error types drives current development of SemRep.

6. Project Plan

In addition to ongoing evaluation efforts, short-term project objectives are to extend the capabilities of the two programs, MetaMap and SemRep, which together produce semantic representation of biomedical text; to continue development of the MetaMap Indexing methodology; and to apply the programs and methods being developed.

The current tokenization regime used by MetaMap and other NLS programs is being enhanced based on an analysis of a variety of token sequences occurring in four collections of MEDLINE citations. The ultimate goal is the specification and implementation of a hierarchical tokenization scheme which provides multiple, coherent tokenizations of text for related but different purposes. In order to increase the efficiency of MetaMap, Metathesaurus strings inappropriate as a basis for semantic interpretation are being considered for possible elimination from the underlying indexes which serve MetaMap. This task is aided by the overt marking in the Metathesaurus of some strings as *suppressible synonyms*.

The ranking function used by MetaMap Indexing is being improved in various ways. First, by normalizing frequency counts based on global counts, in a way analogous to inverse document frequency used in statistically-based retrieval systems. Second, the computation of specificity of a concept is being extended to additional UMLS vocabularies possessing hierarchies, not just MeSH as is currently done. Finally, the ranking function can adjust scores of concepts based on the scores of semantically similar concepts by taking advantage of the knowledge gained in the Semantic Locality project (which is part of the Indexing Initiative project).

Development in SemRep concentrates on the three areas of deficiency noted during the evaluation process. These include addressing word-sense ambiguity, expanding the rules which map from syntactic phenomena to Semantic Network relations, and improving syntactic argument identification. We are currently exploring machine-learning and other statistically-based techniques for resolving word-sense ambiguity. We have developed a semi-automated methodology for identifying semantic rules in a training corpus (the NLM Test Collection) and are using this to expand the set of rules which map from syntactic phenomena to Semantic Network relations. More aggressive use of the SPECIALIST Lexicon serves as the basis for improved identification of syntactic arguments. There is a wealth of syntactic information regarding argument identification in the Lexicon which is relevant to the problems being addressed by SemRep and methods are currently being developed for directly exploiting this information.

7. Summary

The Semantic Knowledge Representation project seeks to provide usable semantic representation of biomedical text by building on resources currently available at the Library, especially the UMLS knowledge sources and the natural language processing tools provided by the SPECIALIST system. Two existing programs, MetaMap and SemRep, are being evaluated, enhanced, and applied to a variety of problems in the management of biomedical information. These include automatic indexing of MEDLINE citations, concept-based query expansion, analysis of complex Metathesaurus strings, accurate identification of anatomical terminology and relationships in clinical records, and the mining of biomedical text for chemical binding relations.

8. References

Aronson, Alan R. 1996. The effect of textual variation on concept-based information retrieval. In James J. Cimino (ed.) *Proceedings of the 1996 AMIA Annual Fall Symposium*, 373-377.

Aronson, Alan R.; Thomas C. Rindflesch; and Allen C. Browne. 1994. Exploiting a large thesaurus for information retrieval. In *Proceedings of RIAO*, 197-216.

Aronson, Alan R., and Thomas C. Rindflesch. 1997. Query expansion using the UMLS Metathesaurus. In Daniel R. Masys (ed.) *Proceedings of the 1997 AMIA Annual Fall Symposium*, 485-489.

Bates, M., and Weischedel R. M. 1993. *Challenges in natural language processing.* Cambridge: Cambridge University Press.

Bean, Carol A.; Thomas C. Rindflesch; and Charles A. Sneiderman. 1998. Automatic semantic interpretation of anatomic spatial relationships in clinical text. Submitted to AMIA 98.

Brachman, Ronald J., and Hector J. Levesque (eds.) 1985. Readings in knowledge representation. Los Altos, CA: Morgan Kaufman Publishers, Inc.

Chakravarthy, Anil S., and Kenneth B. Haase. 1995. NetSerf: Using semantic knowledge to find Internet information archives. In Edward A. Fox, Peter Ingwersen, and Raya Fidel (eds.) *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 4-11.

Chakravarthy, Anil S. 1994. Toward semantic retrieval of pictures and video. In *Proceedings of RIAO 94,* 676-686.

Chierchia, Gennaro, and Sally McConnell-Ginet. 1990. *Meaning and grammar: An introduction to semantics*. Cambridge, MA: The MIT Press.

Cutting D.; J. Kupiec; J. Pedersen; and P. Sibun. 1992. A practical part-of-speech tagger. In *Proceedings of the Third Conference on Applied Natural Language Processing*.

Friedman, Carol; Stephen B. Johnson; Bruce Forman; and Justin Starren. 1995. Architectural requirements for a multipurpose natural language processor in the clinical environment. In Reed M. Gardner (ed.) *Proceedings of the 19th Annual SCAMC*, 347-351

Geach, P., and M. Black (eds.) 1960. *Translations from the philosophical writings of Gottlob Frege*. Oxford: Basil Blackwell.

Haug, Peter J.; Spence Koehler; Lee Min Lau; Ping Wang; Roberto Rocha; and Stanley M. Huff. 1995. Experience with a mixed semantic/syntactic parser. In Reed M. Gardner (ed.) *Proceedings of the 19th Annual Symposium on Computer Applications in Medical Care*, 284-288.

Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. 1995. Unlocking clinical data from narrative reports: a study of natural language processing. *Annals of Internal Medicine* 122(9):681-8.

McCray, Alexa T.; May L. Cheh; Anantha K. Bangalore; K. Rafei; Amir M. Razi; Guy Divita; and P. Zoe Stavri. 1997. Conducting the NLM/AHCPR Large Scale vocabulary Test: A distributed Internet-based experiment. In Daniel R. Masys (ed.) *Proceedings of the 1997 AMIA Annual Fall Symposium,* 560-64.

Humphreys, Betsy L.: Donald A. B. Lindberg; Harold M. Schoolman; and G. Octo Barnett. 1998. The Unified Medical language System: An informatics research collaboration. *Journal of American Medical Informatics Association* 5(1):1-13.

Lindberg, Donald A.B. 1996. Applying knowledge in the digital landscape. Keynote Address, Third Forum on Research and Technology Advances in Digital Libraries, May 13 -15, Washington, D. C.

Lindberg, Donald A. B.; Betsy L. Humphreys; and Alexa T. McCray. 1993. The Unified Medical Language System. *Methods of Information in Medicine* 32:281-291.

Maida, Anthony S., and Stuart C. Shapiro. 1982. Intensional concepts in propositional semantic networks. *Cognitive Science* 6:291-330. [reprinted in Brachman and Levesque 1985]

McCray, Alexa T.; Alan R. Aronson; Allen C. Browne; Thomas C. Rindflesch; Amir Razi; and Suresh Srinivasan. 1993. UMLS knowledge for biomedical language processing. *Bulletin of the Medical Library Association* 1:184-194.

McCray Alexa T.; Suresh Srinivasan; and Allen C. Browne. 1994 Lexical methods for managing variation in biomedical terminologies. In Ozbolt JG (ed.) *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care*, 235-239.

McCray Alexa T.; Amir M. Razi, Anantha K. Bangalore; Allen C. Browne; and P. Zoe Stavri. 1996. The UMLS Knowledge Source Server: A Versatile Internet-Based Research Tool. In Cimino JJ (ed.) *Proceedings of the 1996 AMIA Annual Fall Symposium*, 164-168.

Montague, R. (1974). *Formal philosophy: Selected papers of Richard Montague.* Ed. by R. H. Thomason. New Haven: Yale University Press.

Rassinoux, Anne-Marie; Judith C. Wagner; Christian Lovis; Robert H. Baud; Alan Rector; and Jean-Raoul Scherrer. 1995. Analysis of medical texts based on a sound medical model. In Reed M. Gardner (ed.) *Proceedings of the 19th Annual Symposium on Computer Applications in Medical Care*, 27-31.

Report of the Board of Regents, National Library of Medicine, Long Range Plan. 1987. Office of Inquires and Publications Management, National Library of Medicine, Bethesda, MD.

Rindflesch, Thomas C. 1995. Integrating natural language processing and biomedical domain knowledge for increased information retrieval effectiveness. *Proceedings of the 5th Annual Dual-use Technologies and Applications Conference,* 260-265.

Rindflesch, Thomas C. 1996. Natural language processing. *Annual Review of Applied Linguistics* 16, 71-85.

Rindflesch, Thomas C., and Alan R. Aronson. 1993. Semantic processing in information retrieval. Charles Safran (ed.) *Proceedings of the 17th Annual Symposium on Computer Applications in Medical Care*, 611-615.

Rindflesch, Thomas C., and Alan R. Aronson. 1994. Ambiguity resolution while mapping free text to the UMLS Metathesaurus. Judy G. Ozbolt (ed.) *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care,* 240-244.

Rindflesch, Thomas C.; Alan R. Aronson; and William T. Hole. 1998. The UMLS as a domain model of medicine. Submitted to AMIA 98.

Rosner, Michael and Roderick Johnson (eds.) 1992. *Computational linguistics and formal semantics*. Cambridge: Cambridge University Press.

Saint-Dizier, Patrick, and Evelyne Viegas (eds). 1995. *Computational lexical semantics*. Cambridge: Cambridge University Press.

Schatz, Bruce. 1995. Information analysis in the net: The Interspace of the twenty-first century. White Paper for America in the Age of Information: A Forum on Federal Information and Communications R&D Committee on Information and Communications (CIC). National Science and Technology Council July, National Library of Medicine. [http://csl.ncsa.uiuc.edu/ISWWW/america21.html]

Sneiderman Charles A.; Thomas C. Rindflesch; and Alan R. Aronson. 1995. Extracting physical findings from free-text patient records. 1995 AMIA Spring Congress, Cambridge, Massachusetts, June 25-28.

Sneiderman Charles A.; Thomas C. Rindflesch; and Alan R. Aronson. 1996. Finding the findings: Identification of findings in medical literature using restricted natural language processing. In Cimino JJ (ed.) *Proceedings of the 1996 AMIA Annual Fall Symposium*, 239-243.

Sneiderman, Charles A.; Thomas C. Rindflesch; and Carol A. Bean. 1998 Identification of anatomical terminology in medical text. Submitted to AMIA 98.

Sowa, John F. 1991. Logical structures in the lexicon. In Pustejovsky, James and Sabine Bergler (eds.) *Lexical Semantics and Knowledge Representation*. (Proceedings of a Workshop sponsored by the Special Interest Group on the Lexicon of the Association for Computational Linguistics.)

Srinivasan, Padmini. 1996a. Query expansion and MEDLINE. *Information Processing and Management* 32(4):431-443.

Srinivasan, Padmini. 1996b. Optimal document-indexing vocabulary for MEDLINE. *Information Processing and Management* 32(5):503-514.

Srinivasan Padmini. 1996c. Retrieval feedback for query design in MEDLINE. A Comparison with Expert Network and LLSF Approaches. In James J. Cimino (ed.) *Proceedings of the 1996 AMIA Annual Fall Symposium,* 353-356.

Srinivasan Padmini. Retrieval feedback in MEDLINE. 1996d. *Journal of the American Medical Informatics Association* 3(2):157-167.

Waismann, F. 1971. Meaning. In Jay F. Rosenberg and Charles Travis (eds.) *Readings in the philosophy of language,* 395-402. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Wiesener, S.; W. Kowarschick; and R. Bayer. 1996. SemaLink: An approach for semantic browsing through large distributed document spaces. *Proceedings of the Third Forum on Research and Technology Advances in Digital Libraries*. Los Alamitos, CA:IEEE Computer Society Press.

Wright, Lawrence W.; Holly K. Grossetta Nardini; Alan R. Aronson; and Thomas C. Rindflesch. 1998. Hierarchical concept indexing of full-text documents in the UMLS Information Sources Map. Submitted to JASIS.

Zernik, Uri. 1992. Corpus-based thematic analysis. In Paul S. Jacobs (ed.) *Text-based intelligent systems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Zweigenbaum, P.; B. Bachimont; J. Bouaud; J. Charlet; and J. F. Boisvieuz. 1995. A multi-lingual architecture for building a normalized conceptual representation from medical language. In Reed M. Gardner (ed.) *Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care,* 357-361.