Semantic Experts to Support Advanced Information Management Technologies

1. General methodology

Baud, RH; Lovis, C; Rassinoux, AM; Scherrer, JR. 1998. Alternative ways of knowledge collection, indexing and robust language retrieval. *Methods of Information in Medicine* 37(4-5):315-26.

Hahn, Udo. 1989. Making understanders out of parsers: Semantically driven parsing as a key concept for realistic text understanding applications. *International Journal of Intelligent Systems* 4(3):345-93.

Lytinen, Steven L. [1991.] Semantics-first natural language processing [get the complete reference: ??AAAI] [have]

Riesbeck, Christopher K. 1981. Perspectives on parsing issues. *Proceedings of the Nineteenth Annual Meeting of the Association for Computational Linguistics*, 105-6.

Schank, Roger C. 1975. *Conceptual information processing.* Amsterdam: North-Holland Publishing Co.

Schank, Roger C. 1982. *Dynamic memory.* Cambridge University Press.

Schank, Roger C., and Robert Abelson. 1977. *Scripts, goals, and understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Schank, Roger C.; Michael Lebowitz; and Lawrence Birnbaum. 1980. An integrated understander. *American Journal of Computational Linguistics* 6:13-30.

Wilks, Yorick A. 1972. *Grammar, meaning and the machine analysis of language.* London: Routledge & Kegan Paul Ltd.

Wilks, Yorick A. 1976. Parsing English II. In Eugene Charniak and Yorick Wilks (eds.) *Computational semantics: An introduction to artificial intelligence and natural language comprehension.* Amsterdam: North-Holland Publishing Co.

2. Semantic interpretation

2.1 References for semantic interpretation

Alshawi, Hiyan (ed.) 1992. *The core language engine.* Cambridge, MA: The MIT Press.

van Bakel, Bas, and Reiner T. Boon. 1998. Principle-based structured concept generation - A contribution to knowledge-based document indexing. In P. A. Coppen, et al. (eds.) *Computational linguistics in the Netherlands 1997*, 171-86. Amsterdam: Rudopi.

van Bakel, Bas. 1998. Condorcet Annual Report.

van Bakel, Bas. 2000. Condorcet Final Report

van Bakel, Jan. 1984. *Automatic semantic interpretation: A computer model of understanding natural language.* Dordrecht: Foris Publications.

Bonnema, Remko; Rens Bod; and Remko Scha. 1997. A DOP model for semantic interpretation. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 159-67.

Bos, Johan; Björn Gambäck; Christian Lieske; Yoshiki Mori; Manfred Pinkal; and Karsten Worm. 1996. Compositional semantics in VERB-MOBIL. *Proceedings of the 16th International Conference on Computational Linguistics (COLING),* 131-6.

Charniak, Eugene, and Robert Goldman. 1988. A logic for semantic interpretation. *Proceedings of the 26th Annual Meeting of Association for Computational Linguistics,* 87-94.

Charniak, Eugene, and Yorick Wilks. 1976. *Computational semantics.* Amsterdam: North-Holland Publishing Company.

Dörre, Jochen. 1997. Efficient construction of underspecified semantics under massive ambiguity. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics,* 386-93.

Gardent, Claire, and Michael Kohlhase. 1996. Higher-order coloured unification and natural language semantics. *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics,* 1-9.

Hahn, Udo. 1989. Making understanders out of parsers: semantically driven parsing as a key concept for realistic text understanding applications. *International Journal of Intelligent Systems,* 4(3):345-85.

Jackendoff, Ray. 1997. The architecture of the language faculty. Cambridge, MA: The MIT Press.

Jacobs, Paul S., and Lisa F. Rau. 1990. SCISOR: Extracting information from on-line news. *Communications of the ACM* 33(11):88-97.

Mauldin, Michael. 1991. *Conceptual information retrieval. A case study in adaptive partial parsing.* Boston: Kluwer Academic Publishers.

Moore, Robert C. 1989. Unification-based semantic interpretation. *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics,* 33-41.

Oltmans, Erik. 1999. *A knowledge-based approach to robust parsing*. Entschede, The Netherlands: Centre for Telematics and Information Technology.

Palmer, Martha. 1981. A case for rule-driven semantic processing. *Proceedings of the Nineteenth Annual Meeting of the Association for Computational Linguistics,* 125-31

Romacker, Martin, and Udo Hahn. 2000. An empirical assessment of semantic interpretation. *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational linguistics,* 327-34.

Romacker, Martin; Katja Markert; and Udo Hahn. 1999. Lean semantic interpretation. *Proceedings of the 16th International Joint Conference on Artificail Intelligence (IJCAI),* 868-75.

Rosé, Carolyn P. 2000. A framework for robust semantic interpretation. *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational linguistics,* 311-18. [also see works cited]

3. Aspects of semantic expertise

3.1 References for argument identification

Alshawi, Hiyan, and David Carter. 1992. Sortal restrictions. In H. Alshawi (ed.) *The core language engine,* Ch. 9, pp. 173-85.

Wechsler, Stephen. 1995. The semantic basis of argument structure. Stanford, CA: CSLI Publications.

Wilks, Yorick. 1975. An intelligent analyzer and understander of English. *Communications of the ACM* 18:264-74.

Wilks, Yorick. 1977. Good and bad arguments about semantic primitives. *Communication and Cognition* 10:181-221.

3.2 References for document structure

Baxendale, P. B. 1958. Man-made index for technical literature - an experiment. *IBM Journal of Research and Development* 2(4):354-61.

Braddock, Richard. 1974. The frequency and placement of topic sentences in expository prose. *Research in the Teaching of English,* vol. 8, 287-302.

Edmundson, H. P. 1969. New methods in automatic abstracting. *Journal of the ACM* 16(2):264-85.

Fukumoto, Fumiyo; Yshimo Suzuki; and Jun'ichi Fukumoto. 1997. An automatic extraction of key paragraphs based on context dependency. *Proceedings of the Fifth Conference on Applied Natural Language Processing,* 291-8. Association for Computational Linguistics.

Grosz, Barbara J., and Candace Sidner. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics* 12(3):175-204.

Hahn, Udo. 1990. Topic parsing: Accounting for text macro structures in full-text analysis. *Information Processing and Management* 26(1):135-70. [have]

Hahn, Udo, and Martin Romacker. 1997. Text structures in medical text processing: Empirical evidence and a text understanding prototype. *Proceedings of the AMIA Annual Fall Symposium,* 819-23. [have]

Hearst, Marti A. 1997. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics* 23(1):33-64.

Hearst, Marti A., and Christian Plaunt. 1993. Subtopic structuring for full-length document access. *Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* 59-68.

Kintsch, W., and T. A. van Dijk. 1978. Toward a model of text comprehension and production. *Psychological Review* 85:363-94.

Liddy, Elizabeth. 1991. The discouse level structure of empirical abstracts - an exploratory study. *Information Processing and Management* 27(1):55-81.

Lin, Chin-Yew. 1995. Knowledge-based automatic topic identification. *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics,* 308-10.

Lin, Chin-Yew, and Eduard Hovy. 1997. Identifying topics by position. *Proceedings of the Fifth Conference on Applied Natural Language Processing,* 283-90. Association for Computational Linguistics.

Luhn, H. P. 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development* 2(2), 159-65.

Marcu, Daniel. 1997. From discourse structures to text summaries. *Proceedings of ACL/EACL Workshop on Intelligent Scalable Text Summarization,* 82-8.

Moens, Marie-Francine, and Caroline Uyttendaele. 1997. Automatic text structuring and categorization as a first step in summarizing legal cases. *Information Processing & Management* 33(6):727-37.

Paice, Chris D. 1990. Constructing literature abstracts by computer: Techniques and prospects. *Information Processing and Management* 26(1):171-86.

Paice, Chris D. and Paul A. Jones. 1993. The identification of important concepts in highly structured technical papers. *Proceedings of SIGIR,* 69-78.

Saggion, Horacio, and Guy Lapalme. 2000. Concept identification and presentation in the context of technical text summarization. *Proceedings of the ANLP/NAACL Workshop on Automatic Summarization,* 1-10.

Salton, Gerard, and Chris Buckley. 1993. Approaches to passage retrieval in full text information systems. *Proceedings of SIGIR, 49-58.*

Sjostrom, Colleen L., and Victoria Chou Hare. 1984. Teaching high school students to identify main ideas in expository text. *Journal of Educational Research* 78(2):114-18.

Wilkinson, R. 1994. Effective retrieval of structured documents. *Proceedings of SIGIR,* 311-17.

Zadrozny, Wlodek, and Karen Jensen. 1991. Semantics of paragraphs. *Computational Linguistics* 17(2):171-209.

Zobel, J.; A. Moffat; R. Wilkinson; and R. Sacks-Davis. 1995. Efficient retrieval of partial documents. *Information Processing and Management* 31(3):361-77.

3.3 References for semantic domain experts

Dunham, George. 1986. The role of syntax in the sublanguage of medical diagnostic statements. In R. Grishman and R. Kittredge (eds.) *Analyzing language in restricted domains: Sublanguage description and processing,* 175-94.

Liddy, Elizabeth D.; Corinne L. Jorgensen; Ernest E. Sibert; and Edmund S. Yu. 1993. A sublanguage approach to natural language processing for an expert system. *Information Processing & Management* 29(5):633-645.

3.4 References for interaction of domain ontology and linguistics

Allen, James. 1993. Natural language, knowledge representation and logical form. In M. Bates and R. Weischedel (eds.) *Challenges in natural language processing.* Cambridge University Press.

Hirakawa, Hideki; Zhonghui Xu; and Kenneth Haase. 1996. Inherited feature-based similarity measure based on large semantic hierarchy and large text corpus. *Proceedings of the 16th International Conference on Computational Linguistics (COLING), 508-13.*

Iwanska, Lucja M., and Stuart C. Shapiro (eds.) 2000 Natural language processing and knowledge representation. Cambridge, MA: The MIT Press.

Markert, Katja, and Udo Hahn. 1997. On the interaction of metonymies and anaphora. *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI), 1010-15.*

Simmons, R. F. 1973. Semantic networks: Their computation and use for understanding English sentences. In Schank and Colby (eds.) *Computer models of thought and language.* San Francisco: W.H. Freeman and Co.

4. Aspects of domain expertise / domain knowledge

4.1 References for domain expertise, domain knowledge, and knowledge representation.

Campbell, Keith E.; Diane E. Oliver; Kent A. Spackman; and Eward H. Shortliffe. 1998. Representing thoughts, words, and things in the UMLS. *Journal of the American Medical Informatics Association* 5(5):421-31.

Schulz, Stefan, and Udo Hahn. 2000. Knowledge engineeering by large-scale knowledge reuse: Experience from the medical domain. *Proceedings of the 7th International Conference on Principles of Knowledge Representation and Reasoning.*

Sowa, John F. 2000. *Knowledge representation: Logical, philosophical, and computational foundations.* Pacific Grove, CA: Brooks/Cole.

5. Areas of overlap

6. General machinery and specific rules

6.1 Document structure

6.1.1 General machinery

6.1.2 Specific rules

6.2 Predicate processing; correspondence rules

6.3 Potential arguments

6.3.1 General rules

6.3.2 Specific rules

7. General linguistic procedures

7.1 Before potential argument identification

7.1.1 Primary Tokenization and sentence boundary identification

7.1.2 Secondary tokenization: expand local acronyms

7.1.3 Lexical look-up

7.1.4 Category label ambiguity resolution (stochastic tagger)

7.1.5 Underspecified syntactic parse

7.1.6 Map noun phrases to the domain model (MetaMap)

7.2 After potential argument identification

7.2.1 Coordination

7.2.2 Predicate identification

7.2.3 Argument identification

7.2.4 Relativization

7.2.5 Anaphora resolution