

The Tractable Representation of Utterance Meanings for Information Retrieval

Richard F. E. Sutcliffe*

Piek Vossen#

Peter Hellwig&

and

The SIFT Team*#&

Department of Computer Science*
and Information Systems
University of Limerick
Limerick, Ireland

sutcliff@ul.ie

+353 61 333644
+353 61 330876 (Fax)

Computer Centrum Letteren#
University of Amsterdam
Amsterdam, The Netherlands

University of Heidelberg,&
Karlstrasse 2,
D-6900 Heidelberg,
Germany

How can linguistic meanings can be represented in a form which is computationally tractable? We have developed a paradigm which allows both word and sentence meanings to be captured using distributed patterns, that is, vectors and matrices. These representations can be generated automatically. Word meanings are produced from machine readable dictionaries by a family of taxonomic traversal algorithms. Sentence meanings can then be constructed from parse trees by using the distributed lexical representations together with semantic case information.

In previous work we have reported on an initial algorithm for extracting meaning vectors from the Princeton WordNet (Sutcliffe, O'Sullivan and Meharg). The algorithm has also been applied to the Merriam Webster Compact Electronic (Sutcliffe, 1993) and to the Irish-Irish An Foclo/ir Beag (Sutcliffe, McElligott and O'Neill, 1993). The representations can be shown to be in accordance with human data (Sutcliffe, O'Sullivan, Sharkey, Vossen, Slator, McElligott and Bennis, 1994).

We have now built a full text information retrieval system called SIFT-1 which works with a technical computer manual. It uses the semantic lexicon generated using our MKII WordNet traversal in order to match the meaning of a query with those of utterances in the text. In operation, the user enters a query like "how do I install the wordprocessor". SIFT-1 responds by showing on the screen a portion of the manual which describes how to install the software.

We will describe SIFT-1, outline how the lexicon was generated and report on our findings so far.

References

Sutcliffe, R. F. E. (1993). Constructing Distributed Semantic Lexical Representations using a Machine Readable Dictionary. In K.T. Ryan and R. F. E. Sutcliffe (eds) Proceedings of AICS-91 - The Fifth Irish Conference on Artificial Intelligence and Cognitive Science, University of Limerick, 10-11 September 1992. London, UK, Berlin, FRG, Heidelberg, FRG, New York, NY: Springer-Verlag

Sutcliffe, R. F. E., McElligott, A., O'Neill, G. (1993). Using Distributed Patterns as Language Independent Lexical Representations. In Proceedings of the AAAI-93 Spring Symposium Series: Building Lexicons for Machine Translation, March 23--25, 1993, Stanford, University, CA.

Sutcliffe, R. F. E., O'Sullivan, D., Meharg, F. (1994). A Lexicon of Distributed Noun Representations Constructed by Taxonomic Traversal. Proceedings of the 15th International Conference on Computational Linguistics, (COLING'94), Kyoto, Japan.

Sutcliffe, R. F. E., O'Sullivan, D., Sharkey, N. E., Vossen, P., Slator, B. E. A., McElligott, A., & Bennis, L. (1994). Psychometric Performance Metrics for Semantic Lexicons. Proceedings of the International Workshop On Directions Of Lexical Research, 15-17th of August, 1994, Beijing, China.