Semantic Approach to Text Understanding of Chief Complaints Data

Parsa Mirhaji, Sean Byrne, Narendra Kunapareddy, S. Ward Casscells

The University of Texas Health Science Center at Houston

OBJECTIVE

This paper proposes a semantic approach to processing free form text information such as chief complaints using formal knowledge representation and Description Logic reasoning. Our methods extract concepts and as much contextual information as is available in the text. Output consists of a computationally interpretable representation of this information using the Resource Definition Framework (RDF) and UMLS Metathesaurus [1].

BACKGROUND

Chief complaints are often represented textually and as a mixture of complex and context-dependant lexical symbols with little formal sentence structure [2]. Although human experts usually comprehend this information in its right context intuitively and effortlessly, use of chief complaint data by computers is a challenge.

Semantic approaches for text understanding are concerned with the meaning of terms and their relationships, driven from an explicit model rather than their syntactic forms. Explicit representation of domain concepts along with computer reasoning enables a knowledgeable computer agent to identify those concepts in a given text and pinpoint relevant relationships if they make sense according to an existing formal model available to the agent [3].

METHODS

Our methodology uses Resource Definition Framework (RDF) [4] and the Web Ontology Language (OWL) for knowledge representation [5]. Description Logic inferences are used for classification and case matching

Our methodology is implemented as follows: because there is no guarantee of having a formal sentence structure, the entire chief complaint is considered here as a single term. After a text preparation process that includes spell-checking and expanding known abbreviations and patterns, a syntactic term parser computes an index of all permutations of plausible subterms extractable from a given term based on word location, order, and word counts. From all plausible subterms, only those under five words long are processed further, assuming that the relevant context for a given concept might be found within 1-4 degrees of separation from the word(s) representing that concept. The MMTx linguistic analysis tool [6] from NLM is employed to map such eligible subterms to the UMLS Metathesaurus.

Outputs from MMTx include UMLS semantic types [1, 7] for each mapped concept and a mapping score. Only semantic types with a perfect mapping score of 1000 are processed further. An indexer then creates an RDF representation of the original term; its subterms are mapped to UMLS, their semantic types, their location in the term, and the order in which they appear. A subterm may have multiple UMLS maps and one UMLS map may occur more than once in the term or have more than one semantic type.

RESULTS

We have developed an OWL model that represents clinical evidence as a temporal event having spatial aspects, quantitative and qualitative modifiers, and contextual aspects such as age, presenter, causation, or negation. The model is an extension of the UMLS Semantic Net represented in OWL-DL. A computer agent uses this model, a set of rules, and DL reasoning to interpret the relationship between subterms and their semantic types according to the model.

CONCLUSIONS

Our method is able to extract important clinical observations in nearly all runs and the relevant contextual information in a majority of cases, if they exist. Failures are frequently related to semantically ambiguous or irregular iterations such as ‘referred by doc to check lab’ or ‘patient does not eat/drink/diarrhea’.

REFERENCES

[1] Lindberg D, Humphreys B, McCray A. The Unified Medical Language System. Methods Inf Med. 1993;32:281-91.

[2] Travers DA, Haas SW. Evaluation of Emergency Medical Text Processor, a System for Cleaning Chief Complaint Text Data. Acad Emerg Med 2004;11(11):1170-6.

[3] Allen J. Natural Language Understanding: Benjamin/Cummings; 1995.

[4] W3C. Resource Description Framework (RDF) Model and Syntax Specification. 1999.

[5] W3C. OWL Web Ontology Language, Semantics and Abstract Syntax. 2003.

[6] Aronson AR. Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program. Proc AMIA. 2001.

[7] Bodenreider O. Using UMLS semantics for classification purposes. Proc AMIA Symp. 2000:86-90.