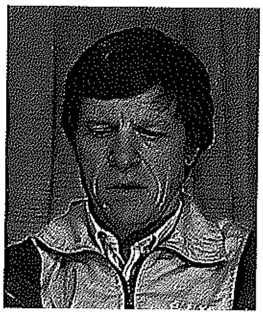
Robert Baud, Laurence Alphay, Anne-Marie Rassinoux, Judith Wagner
University State Hospital of Geneva, Switzerland



Natural Language Processing and GALEN

Baud, R., Alphay, L., Rassinoux, A.-M., Wagner, J.: **Natural language processing and GALEN.**
Int.Classif. 19(1992)No.4, p.193-194, 4 refs.
Within the next three years, analyzing, understanding, indexing, querying, and generating free texts in Natural Language for a couple of European languages will be available in a closed domain like amedical specialty. This means that physicians and nurses will be able to enter information for the patient database using ordinary sentences, more or less well-formed, including typing errors, bad syntax constructions and locally defined medical jargon. A new generation of human interfaces will be at hand, including voice recognition and speech generation, as complementary features of NLP applications. (Authors)

1. Introduction

The goal of our group at Hospital of Geneva is to build a Natural Language Processor within a medical software environment, and to provide NLP functionalities, embedded into this architecture. A Knowledge Representation of medical texts is built and stored within a database and may be queried when needed. This same knowledge representation is ready for use by inferential processes, such as a computer-aided diagnosis system or a critiquing system. Generation of natural language, starting from the Knowledge Representation, ends the loop of basic NLP functionalities, necessary for the system presented here to be complete. And when analysis and generation functions are available together, the subgoal of natural language translation is reachable, at least in a finite domain of knowledge like medicine.

2. History

The Geneva Hospital group has now a five years experience in Natural Language Processing, working with French and English texts. First a fruitful collaboration has been established with the Linguistic String Project of Naomi Sager at New York University. Later, promising results have also been obtained using non-traditional approaches, based on semantic information particular to the domain of knowledge, from which the free text is issued. Amongst those, a solution called "proximity processing" is relevant in medicine (1). The idea is to take advantage of the typical situation of a closed domain of knowledge, and to analyze sentences using semantic information, without implementing formal grammar and the subsequent parser. The proximity of words belonging to predefined classes of concepts drives the process; syntactical information is used, when available, to solve ambinguities as early as possible. However, the system is mainly able to analyze a text from semantic information (2, 3).

3. Plans, and relations to GALEN

Different companion objectives are defined as different states lead to the above goal. These objectives are all part of the NLP system, but, even though not all of them are developed within GALEN, all the resulting programs will be available within the consortium. They are the following:

1) *To analyze any kind of medical text*

This objective means that the analyzer should accept either well-formed sentences or medical "jargon". The NLP system should be powerful enough to enable the resolution of many common situations, including relative sentences, conjunctions, and the resolution of referents. In addition, ordinary human errors should not be an obstacle to achieve adequate processing.

2) *To obtain good response time*

It is necessary that the overall performance of this NLP system allows interactive processing of medical texts. This means that the response time should be of the order of magnitude of one second per line of text.

3) *To provide a sound Knowledge Representation of texts*

From the initial medical text, the NLP system builds a KR using the method of Conceptual Graphs (CG) based on First Order Logic (4). This sound KR is suitable for further processing, either querying a database of texts, or working with deductive or abductive inference methods.

4) *To store Conceptual Graphs into a database DB*

A suitable mechanism allows the storage and retrieval of CGs in a dababase.

5) *To query the DB of medical texts*

A pattern matching algorithm is provided to allow the search of a full DB according to conceptual graphs corresponding to users' questions. The set of matched

https://doi.org/10.5771/0943-7444-1992-4-193 https://www.inlibra.com/de/agb - Open Access -

texts is the answer to the query. Problems belonging to the field of common sense reasoning have to be dealt with at this stage.

6) *To develop dictionary maintenance tools*

Appropriate tools are necessary for the building and the maintenance of the dictionaries related to the chosen closed domain of knowledge. They have to be carefully designed, because a considerable bottleneck for NLP exists, when considering the construction of the basic semantic dictionary. The human interface should be excellent in order to minimize the manpower resources.

7) *To adapt the NLP system to other European languages*

The first version has been written for French, a second version will be developed for English, and a subsequent version is now being designed, for Italian or German.

8) *To generate free text from Conceptual Graphs*

The generation of natural language sentences from CGs is just the opposite to the analyzer. Because the KR is a well designed reference form of medical expressions, and valid as a unique recipient in a multilingual environment, it is necessary to be able to restore free text for a good user interface.

9) *To translate CG from one European language to another*

The translation of CG from one language to another should be possible for any European language. At this level only the instantiations of concepts need to be translated, and such a task is achieved without any natural ambiguity.

10) *To build KR of known nomenclatures in medicine*

Among the different nomenclatures in use, ICD-9 from WHO, must be represented using the CGs. When coupled with the above two objectives, it is possible to maintain a unique nomenclature, instead of one for each European language. An extension to other nomenclatures is also considered.

(11) *To index medical abstracts*

Automatic indexing of abstracts of medical publications is not a fully mastered task today. The KR obtained from NLP on such abstracts certainly contains enough information for indexing, and a dedicated program could perform such a task.

The development of an NLP system within GALEN will take advantage of the Structured Medical Knowledge Representation, which is not far from the conceptual graphs representation. The CORE model is though as a source of medical information when building dictionaries. The implementation of a GALEN prototype in a clinical setting will be realized at Hospital of Geneva.

4. Vision

What is our vision for the second half of the decade? New data bases will be developed, using medical texts as input. An adequate treatment of these texts will allow an intemal representation of them, from which any type of queries will be possible.

In a ten year vision, NLP will be of common use, as a mandatory interface to any software product. Programming languages will be reserved to specialists and natural language will become the most common way to handle computers. In the medical domain, the knowledge representation of texts will be a part of a General Knowledge Base used with Decision Support Systems and Expert Systems. The medical records are possibly fully computerized, including data, images and texts: those three facets are known as the comerstones of the complete mastering of the computerized medical records of the future.

5. Conclusion

The richness of patient to patient information on a text basis, when dealing with discharge letters and consultant reports, is unquestionably more promising than any other encoding methodology. To meet the stage of quasi scientific critique, without being concemed with detailed data, there is nothing really more relevant that can be done. The condition of entering into the detailed data gathering implies the representation of the contents of medical records as a final critical appraisal of the patient. It appears therefore that the NLP approach is not only feasible, but obviously profitable.

References

(1) Rassinoux, A.-M., Baud, R.H., Scherrer, J.R.: Proximity processing of medical texts. In: Proc. MIE. Glasgow, Scotland: Springer Verlag 1990.

(2) Baud, R.H., Rassinoux, A.-M., Scherrer, J.R.: Knowledge representation of discharge sununaries. In: Proc. AIME 1991. Maastricht, NL: Springer Verlag 1991.

(3) Baud, R.H., Rassinoux, A.-M., Scherrer, J.R.: Natural language processing and semantical representation of medical texts. Meth.Inform.Med. (1992)No.2

(4) Sowa, J.F.: Conceptual structures: Infonnation processing in mind and machine. Addison-Wesley Publ. 1984.

Address: Dr.Robert Baud, Hospital Cantonal Univ. de Genève, Centre d'Infonnatique Hospitalière 1211,CH-Geneve 4, Switzerland 9.