MapFace - A Graphical Editor to Support the Semantic Annotation of Medical Text

Theresia Gschwandtner, Katharina Kaiser, and Silvia Miksch
Institute of Software Technology and Interactive Systems
Vienna University of Technology
Vienna, Austria
Email: {gschwandtner,kaiser,miksch}@ifs.tuwien.ac.at

Abstract — *The mapping of medical texts, i.e., clinical practice guidelines, to concepts of medical terminology systems is a prerequisite for many tasks of automatically processing these documents. The MetaMap Transfer (MMTx) program extracts medical concepts from plain text and maps them to corresponding concepts from the UMLS Metathesaurus. Due to the complex nature of this task, this mapping still contains erroneous bits of information.*
Our editor visualizes the annotation of the text and provides means to easily navigate and modify it. Thus we are able to shorten a cumbersome and time-consuming task and subsequently provide reliable and well-defined information for further processing steps. Even more, the visualization features support a better understanding of the medical texts.

I. Introduction

By providing the best available evidence at the point of care, computerized clinical practice guidelines (CPGs) have become an important tool to improve the quality of health care. A front end for developing such an application is the pre-processing of the CPG, in order to create a mapping of concepts from medical terminologies to medical concepts that exist in the text of the CPG. Thus, the plain text is enriched by useful metainformation which disambiguates the semantic meaning of the medical concepts included in the text. To this end, we use the MetaMap Transfer (MMTx) program [1] to create a rough mapping of concepts from the Unified Medical Language System (UMLS) Metathesaurus [2] to free medical text. Due to the complexity of unstructured text, is not always possible to automatically create a correct mapping. However, the complete reliability of information is crucial in medical care - an extremely sensitive discipline - which makes it necessary for medical experts to control and correct these results. This led us to meet the following challenges:

- Enabling medical experts without special programming skills to handle the MMTx program.
- Providing means to assure the correct affiliation of UMLS concepts to text chunks.
- Supporting the understanding of medical concepts in the text as well as relations between them.

II. The MapFace editor

A variety of systems exist annotating text with medical concepts. In contrast to most other systems, MMTx not only supports exact matches between a text token and a UMLS concept, but also considers term variants as well as partial matches. Additionally, it computes matchscores of found candidates by combining specific measured values. For a detailed description of annotation systems we refer to [3].

The MMTx program automatically tokenizes the text into phrase chunks and concept chunks; additionally, it maps UMLS concepts to the text (see Figure 1). Since manually correcting these results would be an extremely cumbersome and time-consuming task, and as there has been no satisfactory tool to support it, we have developed the MapFace editor (see Figure 2). By means of this editor we are able to

1. facilitate the handling of the MMTx program, by
   - providing a GUI, and
   - making the MMTx results easily navigable.
2. assure a correct mapping of the text to UMLS concepts, by providing means to
   - correct the tokenization of the text into chunks (see Figure 3), and to
   - edit the affiliation of UMLS concepts to these chunks (see Figure 3).
3. support the better understanding of the text, by
   - color-coding of semantic types [4],
   - highlighting concepts or phrases accordingly, and
   - listing and highlighting relations among them.

III. Evaluation

In order to ensure the usability of the MapFace editor we conducted an evaluation study. In doing so, we went for a heuristic approach according to Nielsen and Molich [5]. Four evaluators were asked to solve typical tasks with the help of the MapFace editor. They noted and rated each usability problem they encountered.

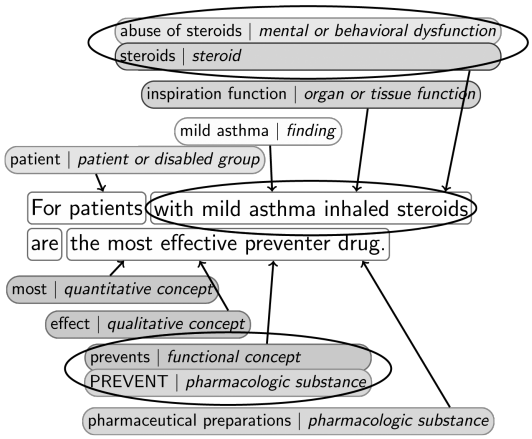

Figure 1: The MMTx program tokenizes the sentence into phrase chunks. A set of best matching UMLS concepts together with their associated semantic types (*in italics*) are returned for each concept identified within the text. The encircled objects are wrong or ambiguous results which have to be corrected by means of the MapFace editor.

The study discovered 32 usability problems, of which 78,12% could be fixed immediately. The other 21.88% either take more effort to fix or they require more general changes in the design. In any case, we will consider all problems carefully in order to improve the quality of the editor.

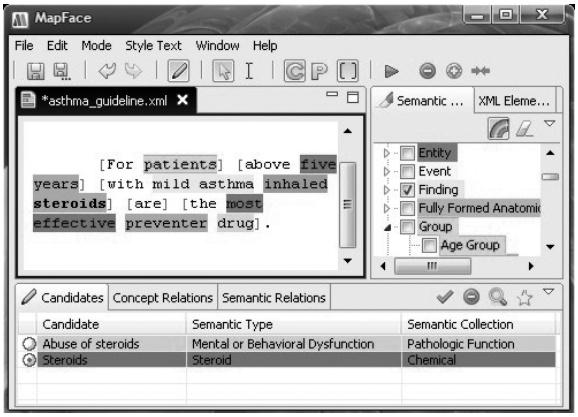

Figure 2: For a given concept chunk a list of best matching UMLS concepts is displayed in the bottom pane. In the right-hand pane the user can select color-coded semantic types to highlight associated text chunks.

IV. Conclusion

The MapFace editor provides important means to navigate, edit, and visualize the semantic annotation of medical documents automatically generated by the MMTx program. Enabling medical experts to easily control and correct the semantic annotation assures the quality of the outcome, which in turn improves the validity of any subsequent processing step. With respect to the outcome of the usability evaluation we will ensure that the MapFace editor is not only an important and time-saving means, but also a convenient tool to work with.

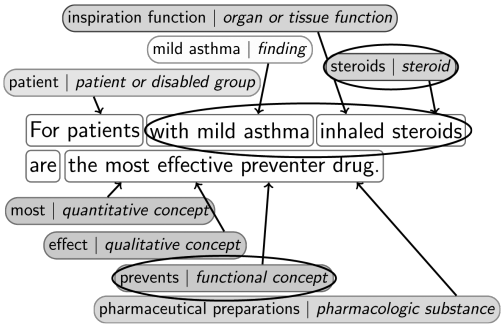

Figure 3: Corrections accomplished by means of the MapFace editor are encircled (compare Figure 1).

Acknowledgments

The research leading to these results has received funding from Fonds zur Förderung der wissenschaftlichen Forschung FWF (Austrian Science Fund), grant L290-N04 and from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 216134.

References

[1] A. R. Aronson. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In *Proceedings of the 25th Annual American Medical Informatics Association Symposium*, pages 17–21, Washington, D.C., November 2001.

[2] P. L. Schyler, W. T. Hole, M. S. Tuttle, and D. D. Sherertz. The UMLS Metathesaurus: representing different views of biomedical concepts. *Bulletin of the Medical Library Association*, 81(2):217–222, April 1993.

[3] L. H. Reeve. *Semantic Annotation and Summarization of Biomedical Text*. PhD Thesis. Drexel University, College of Information Science and Technology, Philadelphia, Pennsylvania, 2007.

[4] A. T. McCray and S. J. Nelson. The representation of meaning in the UMLS. *Methods of Information in Medicine*, 34(1-2):193–201, March 1995.

[5] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 249–256, Seattle, Washington, April 1990.