

MapFace – An Editor for MetaMap Transfer (MMTx)

Katharina Kaiser, Theresia Gschwandtner, and Patrick Martini
Vienna University of Technology
Institute of Software Technology & Interactive Systems
Favoritenstrasse 9-11/188, 1040 Vienna, Austria
{kaiser, gschwandtner, martini}@ifs.tuwien.ac.at

Abstract

Mapping concepts from medical terminologies, such as the UMLS, to medical documents is a prerequisite for many tasks of (automatically) processing documents. Due to the nature of the UMLS and the tools to accomplish the mapping, it is not always possible to achieve a correct and unambiguous mapping. This drawback led us to the development of an editor for correcting the obtained information.

Our editor, *MapFace*, visualizes information received by the *MetaMap Transfer (MMTx)* program and enables users to edit and correct this information on both a conceptual and a syntactical level. By means of this functionality we are able to provide further processing steps with correct and appropriate input. Furthermore, the visualization features enable users to validate or even generate hypotheses, as well as support the better understanding of medical text.

1. Introduction

Mapping concepts from medical terminologies, such as the UMLS, to medical documents is a prerequisite for many tasks of automatically processing documents. A lot of work exists annotating text with medical concepts, such as MicroMeSH [4], Metaphrase [7], MetaMap [1], PhraseX [6], or KnowledgeMap [3], to name a few.

Due to the nature of the UMLS Metathesaurus, sometimes more than one concept seems to be applicable to a particular text chunk. Furthermore, sometimes the adequate concept is not available as the concept cannot be identified in the Metathesaurus, for instance due to different wordings. Thus, we have developed an editor, called *MapFace*, that supports the processing and adaptation of output generated by the *MetaMap Transfer (MMTx)* [1] tool.

MMTx maps text to UMLS Metathesaurus concepts. As part of this mapping process, MMTx tokenizes text into sections, sentences, phrases, terms, and words. MMTx maps

the noun phrases of the text to the best matching UMLS concept or set of concepts that best cover each phrase. See Figure 1 for an example mapping of MMTx.

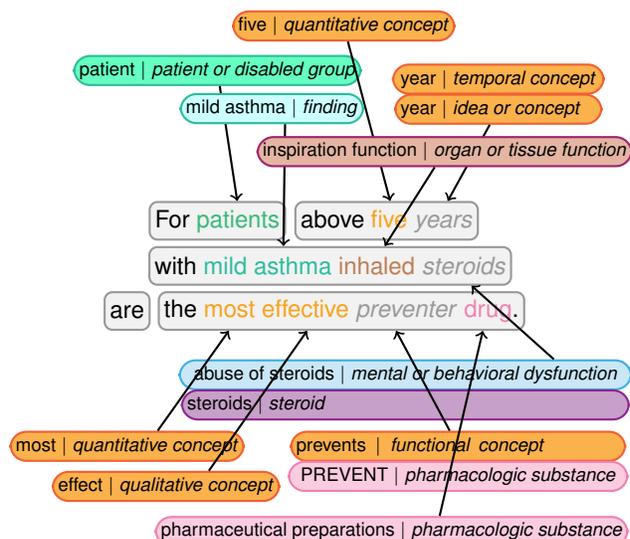


Figure 1. The MMTx program returns the input sentence tokenized into phrases and a set of best matching UMLS concepts for each medical concept identified within the sentence. See Appendix for details of color-coding, etc.

The editor supports the visualization of information obtained from MMTx as well as the adaptation of this information. The visualization of the information is accomplished by highlighting the particular text chunks and color-coding of their semantic information. Furthermore, additional information, such as concept information or relationships between concepts, is also displayed.

We will now describe the *MapFace* editor and its features in the following subsections.

2. MapFace

2.1. The GUI

We created the editor consisting of three sections: (1) editor pane, (2) annotation schema pane, and (3) candidates pane (see Fig. 2). With this framing we are able to display the additional information gathered by MMTx.

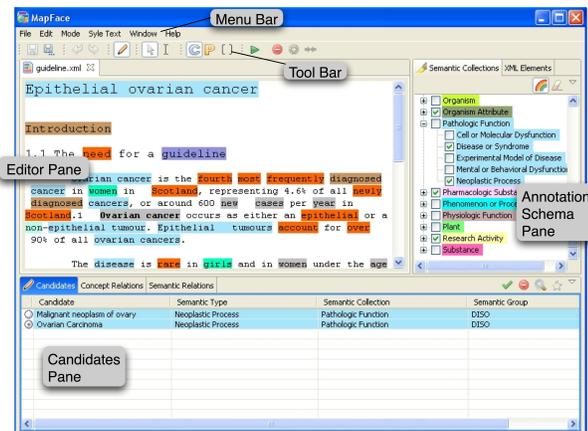


Figure 2. Screenshot of the MapFace editor.

The Editor Pane

The main window of the GUI is the *editor pane*. It displays the text of the clinical guideline. This is where the user can select text and process it by means of the MMTx program.

Once the text is processed by MMTx it is split into sections, sentences, phrases, and words. In case of a unique affiliation of a word or a phrase to a Metathesaurus concept, the text chunk is highlighted according to its semantic type affiliation. The color-codings of the semantic types are displayed in the *annotation schema pane*. If several concept candidates are considered for a word it is highlighted in grey to refer to manual selection of an appropriate concept. The user can select the concepts or phrases of processed text. The matching concept candidates suggested by MMTx are then listed in the *candidates pane*.

The Annotation Schema Pane

This pane displays the annotation schemas. There are two different schemas implemented at present: the *semantic collections view* and the *XML elements view*.

The former contains a list of all semantic types, grouped by *Semantic Collections* [2], each associated with a different color. By selecting a semantic type of interest, all concepts or phrases in the text corresponding to this type are

highlighted in the editor according to the semantic type's color-coding.

The *XML elements view* contains all tag names of XML elements occurring in the underlying XML document of the guideline.

The Candidates Pane

The *candidates pane* is at the bottom of the window. Currently, it contains three different views: (1) candidates view, (2) concept relations view, and (3) semantic relations view. In this pane additional information of the Metathesaurus is displayed.

The *candidates view* provides information about the concept candidates as well as possibilities to edit the list of candidates and select the appropriate one. It displays a list of best weighted concept candidates detected by MMTx for a selected concept or phrase in the editor. The concept candidates are listed together with their semantic types, semantic collections [2], and semantic groups [5].

The *concept relations view* and the *semantic relations view* provide information of conceptual relations and semantic relations, respectively, for a concept candidate selected in the *candidates view*. Each view displays a list of all relations between a selected concept candidate or its semantic type and other concepts or their semantic types occurring in the same section.

2.2. Editing Metathesaurus Information

Editing the Metathesaurus information is necessary, as on the one hand, MMTx cannot always determine an appropriate or distinct concept for a text chunk, and on the other hand, MMTx sometimes provides the wrong syntactical information, which causes errors in the concept assignment (see Fig. 1).

Among the mapping of concepts to text chunks it is often necessary to have information about phrases within the text. Thus, we propose two modes: the *concepts mode* and the *phrases mode*. These modes exclude each other, so only either one of them can be active at one time.

When the **concepts mode** is active, all information, editing and highlighting options refer to medical concepts assigned to text chunks by means of MMTx. The user can select a concept in the editor and read the assigned information or modify the assignment.

Phrases are bigger constructs, which contain a certain number of concepts (or none). When the **phrases mode** is active, all information, editing and highlighting options refer to phrases detected by the MMTx program.

In the *concepts mode* it is possible to create new concepts from selected text and to delete existing concepts. If several equally weighted concept candidates occur, the user

can choose the appropriate one among them or at least decrease the list by removing inappropriate candidates. If no appropriate candidate for a text chunk has been found in the UMLS, it is possible to search for another concept and assign it to the text.

In the *phrases mode* it is possible to create new phrases from selected text or to delete existing phrases. It is also possible to merge two adjacent phrases. For further processing it is sometimes necessary to assign a phrase one semantic type. If a phrase consists of several concepts with different semantic types it is possible to choose among all available semantic types in this phrase or at least to decrease the list of semantic type candidates to the most likely ones. The latter can also be done automatically by taking advantage of the information about relations between semantic types in the sentence.

With these features it is possible to change the MMTx output in order to receive a correct mapping (see Fig. 3).

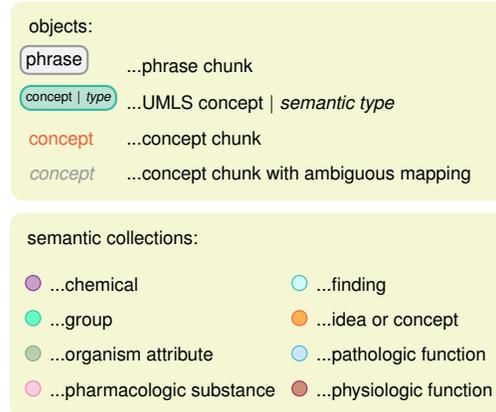


Figure 3. Corrected sentence. A phrase was split, concept chunks "five" and "years" where merged to a single concept chunk, and appropriate UMLS concepts have been assigned.

3. Conclusions

With the MapFace editor we provide an important and useful means to visualize and edit conceptual and syntactical information assigned by the MMTx program. This is important for tasks using information gained from MMTx as input. Using the MapFace editor we can correct and disambiguate the UMLS-specific information in order to improve the output of subsequent tasks. Furthermore, visualizing the semantic information of the UMLS can be useful for knowledge engineers in order to better understand the medical texts.

Appendix: Legend



Acknowledgements

This work is supported by "Fonds zur Förderung der wissenschaftlichen Forschung FWF" (Austrian Science Fund), grant L290-N04.

References

- [1] A. R. Aronson. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In *Proc. of the AMIA Symposium*, pages 17–21, 2001.
- [2] Z. Chen, Y. Perl, M. Halper, J. Geller, and H. Gu. Partitioning the UMLS Semantic Network. *IEEE Transactions on Information Technology in Biomedicine*, 6(2):102–108, June 2002.
- [3] J. C. Denny, J. D. Smithers, R. A. Miller, and A. Spickard. "Understanding" medical school curriculum content using KnowledgeMap. *J Am Med Inform Assoc*, 10(4):351–362, July/August 2003.
- [4] P. Elkin, J. J. Cimino, and H. J. Lowe. Mapping to MeSH: the art of trapping MeSH equivalence from within narrative text. In *Proc. of the 12th Annual Symp Comput Appl Med Care*, pages 185–190, 1988.
- [5] A. T. McCray, A. Burgun, and O. Bodenreider. Aggregating UMLS semantic types for reducing conceptual complexity. In V. L. Patel, R. Rogers, and R. Haux, editors, *Proceedings from the Medinfo 2001 World Congress on Medical Informatics*, volume 84 Studies in Health Technology and Informatics, pages 216–220. IMIA, IOS Press, 2001.
- [6] S. Srinivasan, T. C. Rindflesch, W. T. Hole, A. R. Aronson, and J. G. Mork. Finding UMLS metathesaurus concepts in MEDLINE. In *Proc. of the AMIA Annual Symposium*, pages 727–731. AMIA, 2002.
- [7] M. Tuttle, N. Olson, K. Keck, W. Cole, M. Erlbaum, D. Sherertz, C. Chute, P. L. Elkin, G. Atkin, B. Kaihoi, C. Safran, D. Rind, and V. Law. Metaphrase: an aid to the clinical conceptualization and formalization of patient problems in healthcare enterprises. *Methods of Information in Medicine*, 37(4-5):373–373, Nov 1998.