

MapFace – An Aid for Medical Experts to Easily Annotate Documents with MetaMap Transfer

Theresia
Gschwandtner

Katharina
Kaiser

Patrick
Martini

Silvia
Miksch

Vienna University of Technology,
Institute of Software Technology & Interactive Systems
Favoritenstrasse 9-11/188, 1040 Vienna, Austria

{lastname}@ifs.tuwien.ac.at

ABSTRACT

Mapping medical concepts from a terminology system, to the concepts in the narrative text of a medical document is necessary to provide semantically accurate information for further processing steps. The MetaMap Transfer (MMTx) program is a semantic annotation system that generates a rough mapping of UMLS concepts to free medical text, but this mapping still contains erroneous and ambiguous bits of information. Since manually correcting the mapping and annotating the text chunks of a document accordingly would be an extremely cumbersome and time-consuming task, and since there has been no satisfactory tool to support it, we have developed the MapFace editor. The editor not only greatly facilitates the handling of the MMTx program, but it also provides a convenient way of navigating the annotated information gained from the MMTx output, and enables users to correct this information on both a conceptual and a syntactical level. Thus, we are able to provide reliable and well-defined information for further processing steps. Finally, a usability study was conducted to derive an informative basis for improvement.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – *Graphical user interfaces (GUI), Prototyping, Screen design (e.g., text, graphics, color)*

I.7.2 [Document and Text Processing]: Document Preparation – *Markup languages*

General Terms

Design

Keywords

Semantic Annotation, Annotation of Medical Documents, Graphical Editor, MetaMap Transfer (MMTx), UMLS, Concept Mapping.

1. INTRODUCTION

By annotating a free text document, unstructured text is enriched by meta-information. The identification of medical concepts in the narrative text of a medical document and the annotation of appropriate concept information, as provided by the UMLS Metathesaurus, is necessary to disambiguate the semantic meaning of the medical concepts included in the text and hence, it is a prerequisite for correctly interpreting the guideline while translating it into a computer-executable model. To enrich medical documents like this, it is necessary

- (a) to have access to a thesaurus of medical concepts (providing meta-information), and
- (b) to map these thesaurus concepts to corresponding text chunks of the document in order to annotate these text chunks with appropriate meta-information.

To this end, we use medical concepts provided by the Unified Medical Language System (UMLS) Metathesaurus [11] – the largest thesaurus in the biomedical domain (see (a)).

Subsequently, a mapping of the narrative text of a document to the corresponding concepts in the UMLS Metathesaurus has to be created. A variety of systems exist annotating text with medical concepts, such as MicroMeSH [4], Metaphrase [13], MetaMap [1], PhraseX [12], or KnowledgeMap [3], just to name a few.

We use the MetaMap Transfer (MMTx) program which is considered a state-of-the-art system for semantic annotation [3]. It assigns concepts from the UMLS Metathesaurus to corresponding text chunks of the document (see (b)). 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 scores found candidates by combining different measures. For a detailed description of annotation systems we refer to [10].

Due to the ambiguity of free text, the correct and automatic creation of an unambiguous mapping of UMLS concepts to medical concepts in the text of a document will probably never be accomplished totally correctly by means of a semantic annotation system. Hence, the reliability of the MMTx results is not granted, 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 Figure 1).

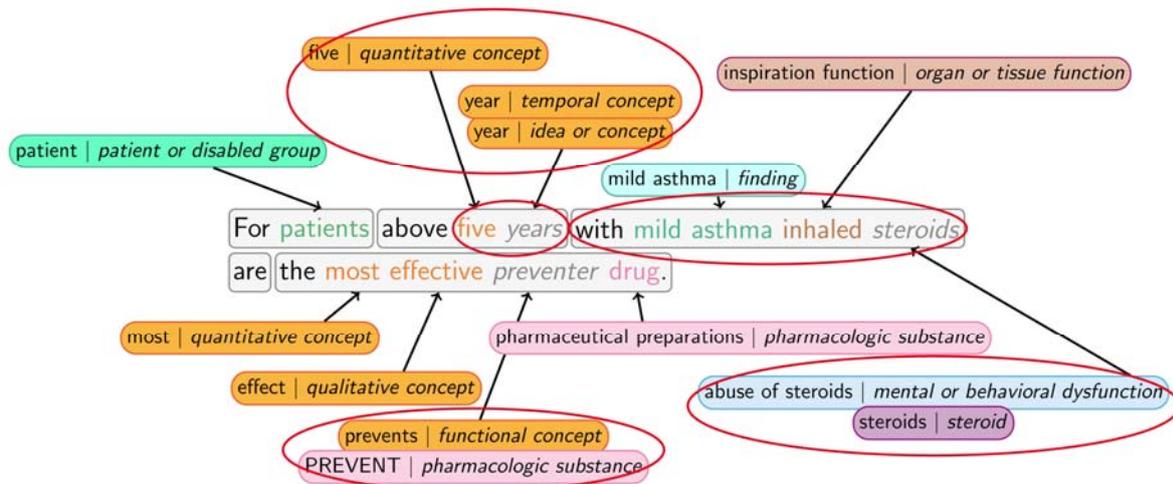


Figure 1. MMTx tokenizes a sentence into phrase chunks and maps medical concepts available in the UMLS Metathesaurus to text chunks. The encircled objects are wrong or ambiguous results of the MMTx which have to be corrected by means of the MapFace editor.

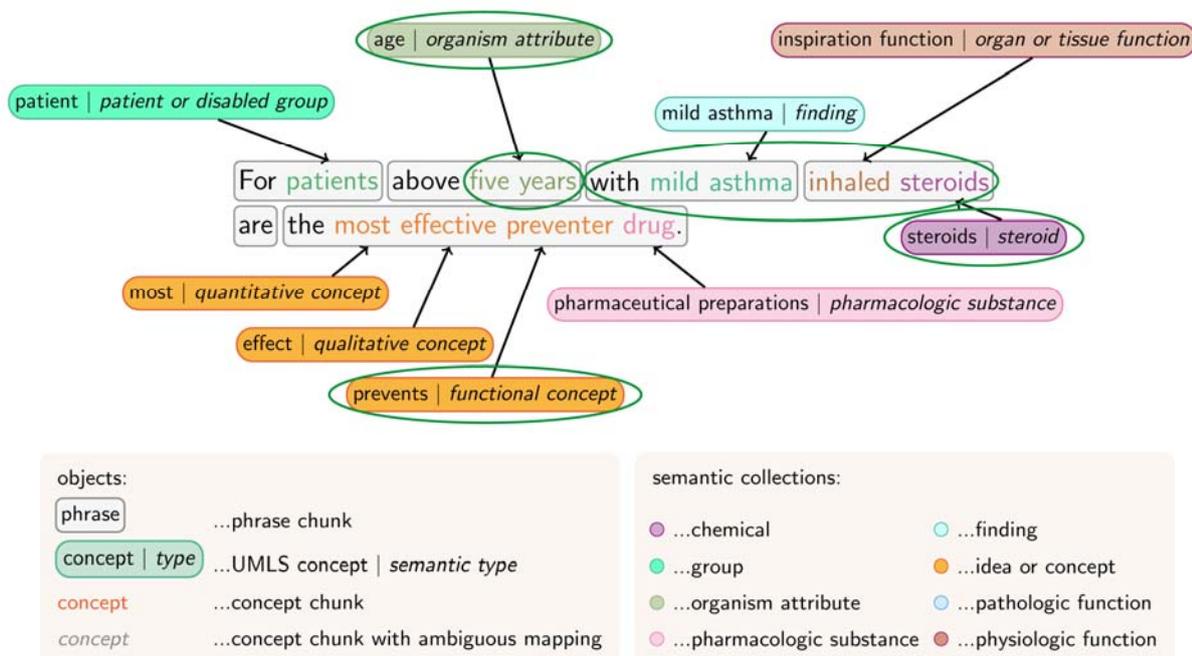


Figure 2. Corrections accomplished using MapFace are encircled (comp. Figure 1): disambiguation of the mapping of UMLS concepts to the text chunks years, steroids, and preventer. In addition, the tokenization of the phrase chunk with mild asthma inhaled steroids has been corrected and instead of the two concept chunks five and years we prefer a single concept chunk with the semantic meaning of "age".

Since medical care is an extremely sensitive discipline, the complete reliability of the annotated information, i.e., UMLS concepts correctly fitting the medical concepts in the text, is crucial for its usability for further processing. Thus, it is absolutely necessary for medical experts to control these results and to modify them, if necessary (see Figure 2). However, controlling and modifying the output of the MMTx program is hard to handle for physicians, since the use of MMTx is command line based and requires a moderate amount of programming knowledge. An example of the original output of MMTx for (only) the phrase "the most effective preventer drug" is given in the following.

Phrase: "the most effective preventer drug."

Meta Candidates (13)

812 Drug (Pharmaceutical Preparations) [Pharmacologic Substance]

756 Pharmaceutical [Intellectual Product]

719 Pharmaceutics (Pharmacy) [Biomedical Occupation or Discipline]

719 Medicament [Pharmacologic Substance]

645 Most [Quantitative Concept]

645 Effective (Effect) [Qualitative Concept]

574 PREVENT [Pharmacologic Substance]

574 Effectiveness [Qualitative Concept]
 574 Prevents [Functional Concept]
 545 Prevention (Prophylactic treatment) [Therapeutic or Preventive Procedure]
 545 Prevention [Therapeutic or Preventive Procedure]
 545 Preventable [Intellectual Product]
 530 Preventive (Preventive intent) [Qualitative Concept]

Meta Mapping (815)
 645 Most [Quantitative Concept]
 645 Effective (Effect) [Qualitative Concept]
 574 Prevents [Functional Concept]
 812 Drug (Pharmaceutical Preparations) [Pharmacologic Substance]

Meta Mapping (815)
 645 Most [Quantitative Concept]
 645 Effective (Effect) [Qualitative Concept]
 574 PREVENT [Pharmacologic Substance]
 812 Drug (Pharmaceutical Preparations) [Pharmacologic Substance]

It is evident that it is not easy to assess and occasionally correct the information contained in this text list. This drawback led us to develop an editor with a graphical user interface, which enables physicians to solve this task without requiring special skills in information processing. With the MapFace editor we provide the following features:

- **Providing a GUI for MMTx:** We facilitate the creation of an automatic mapping of UMLS concepts to text chunks of a document with the help of the MMTx program by providing a graphical environment enabling the user to execute the MMTx program by clicking a button instead of having to cope with the command line based execution.
- **Automatic annotation of the document text:** The text chunks of the document are automatically annotated with information, based on the output of the MMTx program.
- **Navigation and visualization of the MMTx output:** We provide a convenient way to navigate and visualize the results of the MMTx program.
- **Modification on both syntactic and semantic level:** The editor provides means to easily modify the annotated information at a syntactic as well as at a semantic level.
- **Save and reload of annotated documents:** The annotated information – generated by MMTx and modified by the user – can be saved and reloaded which is a very useful means in respect that generating this information once again with MMTx takes a long time.

Finally, we evaluated the usability of the editor (see Section 3), to identify possible shortcomings of the design. The results of the evaluation provide an informative basis for improvement. Considering these outcomes we will ensure an intuitive and convenient way of working with the editor.

2. THE MAPFACE EDITOR

To begin with, here is a rough outline of how to use the MapFace editor:

1. The user loads a medical document (XML document) into the MapFace editor and clicks a button to process its text by means of the MMTx program. Whereupon, MMTx tokenizes the text into sections, sentences, and phrases, and additionally, assigns the concepts from the UMLS Metathesaurus to corresponding text chunks in the document. Thus MMTx provides not only conceptual/semantic information but also syntactic information.
2. Annotated text chunks – concept chunks and phrase chunks (see Section 2.1) – are now identifiable in the editor; on selection, the assigned meta-information (e.g., fitting UMLS concepts) is displayed.
3. Text chunks for which no distinct UMLS concept could be determined are marked by gray background to indicate the user to manually assign a UMLS concept to this chunk.
4. The user navigates the results to control the automatically generated annotation and – if necessary – modifies them by means of the MapFace editor (see Section 2.2).

The following subsections overview which kind of information is annotated (see Section 2.1), how the user modifies the annotation (see Section 2.2), what features are provided to support the correct affiliation of equally matching meta-information to text chunks (see Section 2.3), why there are different annotation schemes (see Section 2.4), and how the semantic information is visualized (see Section 2.5).

2.1 Annotated Information

The annotation of the text is accomplished by inserting into the document XML tags. It concerns the tokenization of the text into sections, sentences, phrases, and concepts as well as the semantic information about assigned UMLS concepts.

A **concept chunk** is annotated with the best fitting UMLS concept together with its *semantic type* [7], *semantic collection* [2], and *semantic group* [6]. For instance, the text chunk patients is annotated with the UMLS concept Patient, with the *semantic type Patient or Disabled Group* (see Figure 1), which belongs to the *semantic group Living Beings*, as well as to the *semantic collection Group*.

Phrases are groups of words that function as a single unit, which can contain several concepts. To annotate a **phrase chunk**, the user manually chooses one of the *semantic types* assigned to the concept chunks within this phrase. For instance, the phrase the most effective preventer drug contains the concept chunks (with associated *semantic types*): most (*Quantitative Concept*), effective (*Qualitative Concept*), preventer (*Functional Concept*), and drug (*Pharmacologic Substance*) (see Figure 2); the appropriate *semantic type* for this phrase chunk is *Pharmacologic Substance*.

2.2 Editing the Annotation

If the user selects a text chunk in the editor, the annotated information is displayed in the *candidates pane* at the bottom of the window (see Figure 3). This pane displays information about the assigned UMLS concept or – in case of ambiguity –

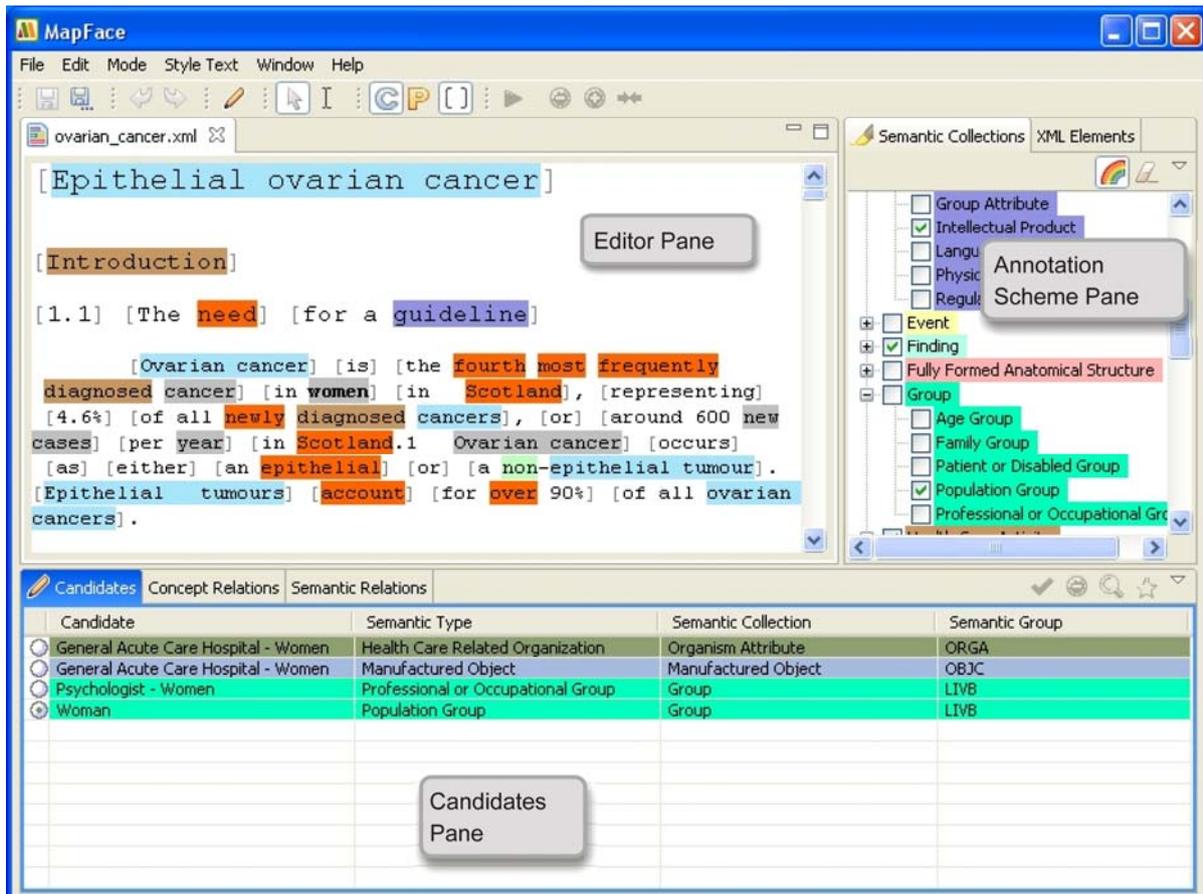


Figure 3. A screenshot of the MapFace editor: concept chunks are highlighted according to the color-code of the semantic type associated with the annotated UMLS concept; phrase chunks are delimited by brackets. The Candidates pane displays a list of best matching UMLS concepts for the text chunk women, which is selected in the editor.

a list of possible matching UMLS concept candidates. Furthermore, possibilities to modify the annotation at a semantic level are provided:

- **Choose a distinct UMLS concept candidate:** In case of an ambiguous mapping to more than one UMLS concept, a list of matching UMLS concept candidates is displayed, from which the user can choose the appropriate concept.
- **Look for additional UMLS concept candidates:** If the appropriate candidate does not appear in the candidates list, the user can search the Metathesaurus for additional UMLS concepts by entering an alternative expression for the concept text.
- **Delete the annotation:** The annotated information can be deleted if it is not appropriate, i.e., on the one hand, selected concept and phrase tags, and on the other hand, the UMLS concept information assigned to a concept chunk or – in case of a phrase chunk – the assigned semantic type.

Furthermore, the editor provides features to easily modify the tokenization of the text into phrase chunks and concept chunks at a syntactic level:

- **Modify the bounds of annotated text chunks:** Wrongly tokenized phrase chunks and concept chunks can be deleted and in turn, new phrase and

concept chunks can be created. This enables the user to modify the delimitations of these text chunks.

- **Merge two phrase chunks:** In addition, two adjacent phrase chunks can be merged in one step (to shorten this frequently used action).

2.3 Supporting the Correct Assignment

To facilitate the selection of the correct UMLS concept for a given concept chunk (in case of ambiguity), or the selection of the appropriate semantic type for a phrase chunk, the editor provides some special features.

- **Relations between UMLS concepts:** A list of relations between each possible UMLS concept candidate for a given concept chunk and the UMLS concepts assigned to other concept chunks in the same section of the text can be displayed. On selection of a specific relation, the corresponding concept chunks are highlighted in the editor.
- **Relations between semantic types:** Additional decision support can be derived from exploiting the relations between semantic types of UMLS concepts.
- **Automatic decrease of semantic types list:** The list of semantic type candidates for a phrase chunk can be automatically decreased to the most likely ones by taking advantage of the information about relations of

each candidate to the *semantic types* assigned to text chunks in the same sentence.

2.4 Annotation Schemes

The basic idea was to create an editor supporting different kinds of annotated information relevant for the processing of medical documents. This is why the *annotation scheme pane* at the right of the user interface (see Figure 3) contains diverse views for different annotation schemes.

Each view displays a list of subjects relevant for a specific annotation context. There are two annotation schemes implemented at present: the *semantic types annotation scheme* and the *XML elements annotation scheme*. The former is concerned with the annotation of UMLS concepts and their *semantic types* as described in this paper and the latter simply deals with the visualization of all XML elements occurring in the underlying XML document.

In order to be easily extendable, the editor has a plugin based architecture which allows the application of additional annotation schemes, e.g., annotation schemes for co-reference or negation detection [5].

2.5 Visualizing the Annotation

The view of the *semantic types annotation scheme* contains a comprehensive list of *semantic types* of UMLS concepts, grouped by *semantic collections* [2]. Since there are 135 different *semantic types*, the grouping into 28 *semantic collections* facilitates the inspection of the *semantic types* as well as enables the association of each collection with a unique color. This color-code is used to highlight the annotated text chunks accordingly.

In case of an ambiguous mapping to more than one UMLS concept, the text chunk in the editor is marked by a gray background (in contrast to the highlighting of text chunks which are already associated with *semantic types*, it cannot be turned off), which serves as reminder that the user should manually assign a UMLS concept to this chunk.

3. USABILITY EVALUATION

In order to assure the usability of the MapFace editor, a usability evaluation was conducted to gather feedback about its design. We went for a heuristic evaluation approach according to Nielsen [8][9], because it is focused on usability aspects like user control and freedom, aesthetic design, information structuring, consistency of the terminology and of the interaction mechanisms, etc.

A heuristic usability evaluation study was conducted with four evaluators, who have considerable knowledge about usability principles. The evaluators were asked to solve typical tasks by means of the MapFace editor. They noted every usability problem they encountered during the session and rated the severity of the problem on a scale from 1 to 5 whereas 5 represents the highest severity.

The total number of usability problems found is 32; their distribution according to their average severity – in case more than one evaluator encountered and rated one problem – is given in Table 1.

Table 1. The Distribution of problems according to their average severity (5 represents the highest severity).

Avg. Severity	1	2	3	4	5
Problems	2	6	17	4	3

- (1) **43,75 %** of all found problems take only **low effort** to fix them, e.g.,
 - Some information text should be changed.
 - Some features should be active in the first place.
- (2) **34,37 %** of all found problems take **moderate effort** to fix them, e.g.,
 - Shortcuts should be provided for frequently used actions.
 - Some features could be combined to a single button.
- (3) **21,88 %** of all found problems take **more effort to fix them**, e.g.,

- **The necessity to frequently change selection tools:** So far, the MapFace editor provides two selection tools for different purposes. The first tool is an “arrow cursor” to select already annotated text chunks in the editor by a double-click in order to read or modify the annotated information. The second selection tool is an “I-beam” or “text cursor” to select any text of the document, e.g., in order to process it by means of MMTx, to create a new concept chunk from the selected text, or to select two adjacent phrase chunks for merging. Switching the selection tools is accomplished by buttons in the tool bar.

However, while correcting the output of the MMTx program the user has to switch the active selection tool frequently which is cumbersome and unfavorable to the workflow. Thus, we intend to realize both means – selecting annotated text chunks by a double-click as well as selecting any text of the document by dragging the cursor – within one tool.

- **Switching between concepts mode and phrases mode:** The annotation of text chunks is carried out at two levels, i.e., at a concept level concept chunks in the text are annotated with fitting UMLS concepts and at a phrase level phrase chunks are annotated with appropriate semantic types. Hence, the editor provides two modes, i.e., a concepts mode and a phrases mode, which can be switched in the tool bar too. When the concepts mode is active only concept chunks can be selected or modified. In turn the user has to switch to the phrases mode to be able to deal with phrase chunks.

Solving the task of correcting the output of MMTx requires a basic idea about what presents a concept and what presents a phrase, and additionally, it demands for switching between these two modes. Two of four usability evaluators found that irritating.

However, the advantages and shortcomings of merging these two modes into one mode, which would allow the user to deal with both concept chunks and phrase chunks, have to be carefully examined. Since the text of these two different kinds

of chunks overlaps, it is supposable that visualizing both kinds of chunks at the same time could lead to a very crowded appearance. In addition, allowing the user to select either a phrase chunk or one of the concept chunks within the phrase cannot be accomplished by a double-click on the corresponding text anymore (due to the overlapping text). More sophisticated ways of selection would have to be realized, which in turn could affect the usability of the editor adversely. Still we intend to deal with this problem intensively in order to find a way to ensure a convenient workflow.

Problems of the categories (1) and (2) can be fixed immediately. Problems of the category (3) take more effort to fix them or they require more general changes in the design. However, all found usability problems will be considered carefully in order to improve the editor.

4. CONCLUSION

The MapFace editor is an important and useful means to create, visualize, and edit the semantic annotation of a medical document. Using the MapFace editor enables medical experts to automatically annotate the document by means of the MMTx program, to easily navigate through the annotated information, and to modify the annotation both at a semantic as well as at a syntactic level. To do so without the MapFace editor would be a very cumbersome task requiring not only moderate skills in programming, but also an enormous amount of working hours. Controlling and correcting the annotation by means of the MapFace editor ensures the quality of the outcome, which in turn is imperative for the validity of the outcome of any subsequent processing step. With respect to the results of the usability evaluation, we will ensure that the MapFace editor is not only a useful and time-saving means, but also an intuitive and convenient tool to work with.

5. ACKNOWLEDGMENTS

This work is supported by "Fonds zur Förderung der wissenschaftlichen Forschung FWF" (Austrian Science Fund), grant L290-N04 and by the European Commission's IST program, under contract FP7-ICT-216134 ReMINE.

6. REFERENCES

- [1] Aronson, A. R. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In *Proceedings of the 25th Annual American Medical Informatics Association Symposium (AMIA '01)* (Washington, D.C., Nov. 3-7, 2001). AMIA, Washington, D. C., 2001, 17-21.
- [2] Chen, Z., Perl, Y., Halper, M., Geller, J., and Gu, H. Partitioning the UMLS Semantic Network. *IEEE Transactions on Information Technology in Biomedicine (IEEE TITB)*, 6, 2 (June 2002), 102–108.
- [3] Denny, J. D., Smithers, J. D., Miller, R. A., and Spickard, A. Understanding medical school curriculum content using KnowledgeMap. *Journal of the American Medical Informatics Association (JAMIA)*, 10, 4 (July/Aug. 2003), 351–362.
- [4] Elkin, P., Cimino, J. J., and Lowe, H. J. Mapping to MeSH: the art of trapping MeSH equivalence from within narrative text. In *Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care (SCAMC '88)* (Washington, D.C., Nov. 6-9, 1988). IEEE Computer Society Press, 1988, 185-190.
- [5] Gindl, S., Kaiser, K., and Miksch, S. Syntactical Negation Detection in Clinical Practice Guidelines. In *Proceedings of the 21st International Congress of the European Federation for Medical Informatics (MIE 2008)* (Göteborg, Sweden, May 25-28, 2008). IOS Press, Amsterdam, The Netherlands, 2008, 136, 187–192.
- [6] McCray, A. T., Burgun, A., and Bodenreider, O. Aggregating UMLS Semantic Types for Reducing Conceptual Complexity. In *Proceedings of the 10th World Congress on Medical Informatics, 84 Studies in Health Technology and Informatics (MEDINFO '01)* (London, UK, Sept. 2 - 5, 2001). IMIA, IOS Press, Amsterdam, The Netherlands, 2001, 84, 216–220.
- [7] McCray, A. T. and Nelson, S. J. The representation of meaning in the UMLS. *Methods of Information in Medicine*, 34, 1-2 (Mar. 1995), 193–201.
- [8] Nielsen, J. Heuristic evaluation. *Usability Inspection Methods*. John Wiley & Sons, New York, NY, 1994.
- [9] Nielsen, J. and Molich, R. Heuristic evaluation of user interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 1990)* (Seattle, Washington, April 1-5, 1990). ACM Press, New York, NY, 1990, 249-256.
- [10] Reeve, L. H. *Semantic Annotation and Summarization of Biomedical Text*. PhD Thesis, Drexel University, College of Information Science and Technology, Philadelphia, PA, 2007.
- [11] Schuyler, P. L., Hole, W. T., Tuttle, M. S., and Sherertz, D. D. The UMLS Metathesaurus: representing different views of biomedical concepts. *Bulletin of the Medical Library Association*, 81, 2 (April 1993), 217–222.
- [12] Srinivasan, S., Rindflesch, T. C., Hole, W. T., Aronson, A. R., and Mork, J. G. Finding UMLS metathesaurus concepts in MEDLINE. In *Proceedings of the 26th Annual American Medical Informatics Association Symposium (AMIA '02)* (San Antonio, Texas, Nov. 9-13, 2002). AMIA, Washington, D. C., 2002, 727–731.
- [13] Tuttle, M., Olson, N., Keck, K., Cole, W., Erlbaum, M., Sherertz, D., Chute, C., Elkin, P. L., Atkin, G., Kaihoi, B., Safran, C., Rind, D., and Law, V. Metaphrase: an aid to the clinical conceptualization and formalization of patient problems in healthcare enterprises. *Methods of Information in Medicine*, 37, 4-5 (Nov. 1998), 373–373.