MapFace v1.0

User Manual

Theresia Gschwandtner

February 2008

Contents

1 Introduction

MapFace is a Graphical User Interface, developed to provide a simple possiblity to process a textual clinical guideline by means of the MetaMap Transfer program (in the following also referred to as the MMTx program)[1, 2], an application which automatically maps free text of a guideline to corresponding concepts in the UMLS®Metathesaurus®[7], the largest thesaurus in the biomedical domain. MapFace was developed to provide an intuitive access to these UMLS concepts together with all the conceptual information, and to provide a tool to modify them.

The aim of MapFace is to serve the needs of two different user groups: physicians and knowledge engineers. On the one hand it should provide a comfortable way for physicians (who in general are not familiar with computer languages or programming) to process the text of a clinical guideline by means of the MMTx program and to edit the results. Thus, MapFace provides possiblities to correct, delete, or confirm the UMLS concepts assigned to text chunks by the MMTx program and to pick the right concept from two or more evenly matched UMLS concepts.

On the other hand, MapFace serves the needs of knowledge engineers who need to correctly interpret the medical text for further processing of the guideline to a computer interpretable model. To achive this, MapFace displays the processed guideline together with all the assigned information, and provides easy access to the semantic meanings of the different constructs of the guideline. Thus, MapFace facilitates the correct interpretation of the text, which is imperative for further processing.

1.1 Procedure

The main object of MapFace is to create a correct mapping of text of a clinical guideline to medical concepts using the MMTx program. An arbitrary sentence of a guideline could be:
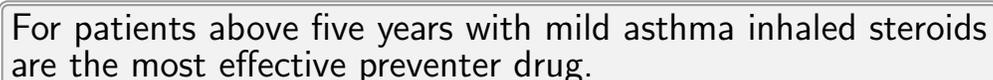
For patients above five years with mild asthma inhaled steroids are the most effective preventer drug.

Figure 1: A possible sentence of a guideline.

The MMTx program tokenizes the sentence into phrase chunks and maps the text to medical concepts available in the UMLS Metathesaurus:

Figure 2: 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.

Corrigenda:
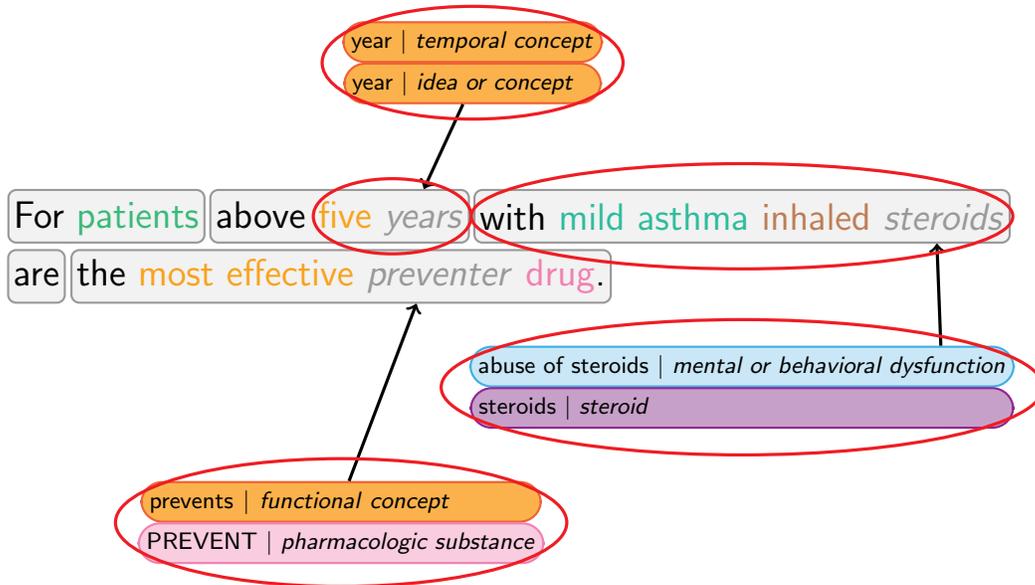
Figure 3: Wrong or ambiguous MMTx results are to be corrected by means of the MapFace editor.

As we can see above the MMTx program is not able to create an unambiguous mapping for the text chunks "years", "steroids", and "preventer". In addition, the tokenization of the phrase chunk "with mild asthma inhaled steroids" is not correct and instead of the two concept chunks "five" and "years" we would prefer a single concept chunk with the semantic meaning "age".

We can correct the mapping and tokenization with the help of MapFace. To do so, we delete the wrong phrase chunk and create two new chunks from selected text.

This is how the tokenization looks afterwards:
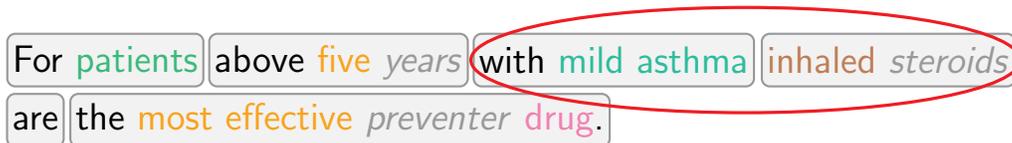
Figure 4: The phrase chunk "with mild asthma inhaled steroids" has been split.

Next we need to choose the correct UMLS concept in case of ambiguity and affiliate it to the corresponding concept chunk:
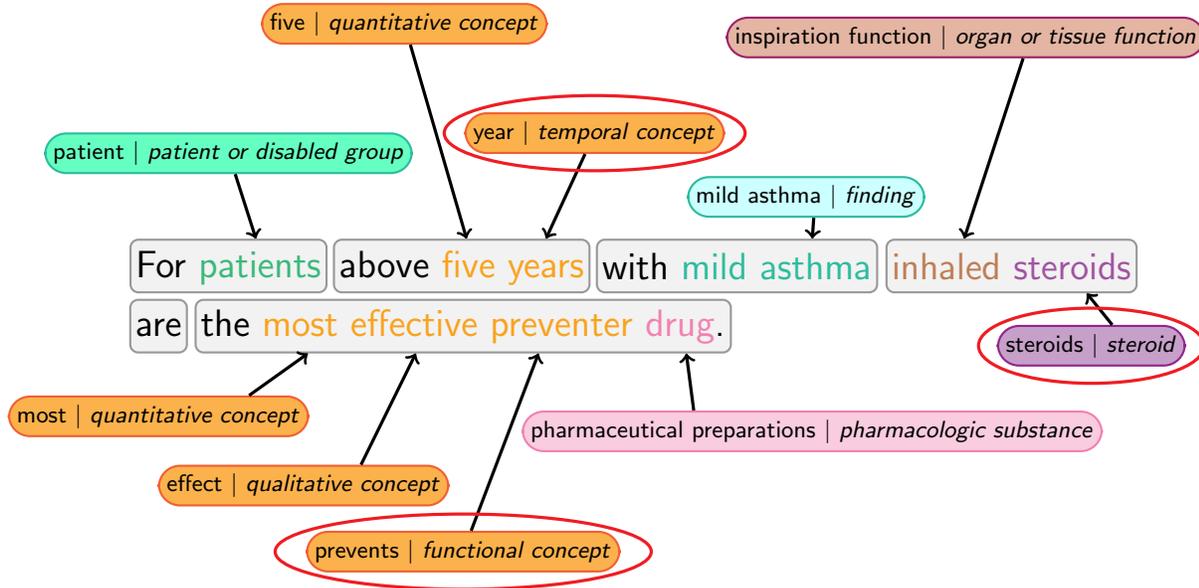
Figure 5: Appropriate UMLS concepts have been selected in case of ambiguity.

Still we don't agree with the mapping of the text chunks `five` and `years`. To create a single concept chunk with the semantic meaning "age", we delete the two chunks and create a new one from the text `five years`, whereupon we search for the UMLS concepts matching the text "age" and affiliate the correct match to the created concept chunk:
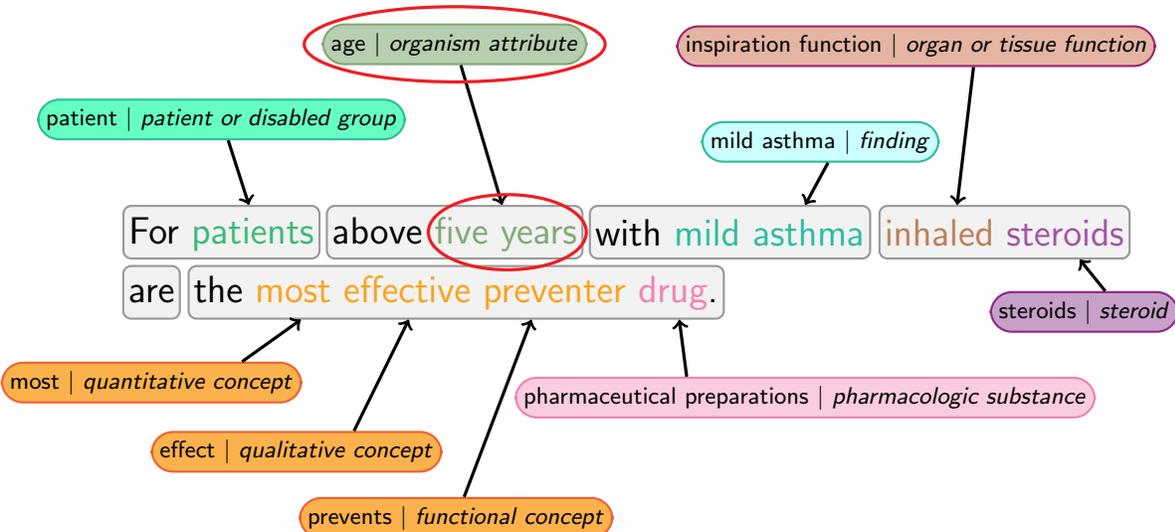
Figure 6: The concept chunks "five" and "years" where merged to a single concept chunk and an appropriate UMLS concept has been assigned.

Now, for each concept chunk in the sentence the correct UMLS concept has been found. Last but not least we assign a semantic type to each phrase containing at least one medical concept:
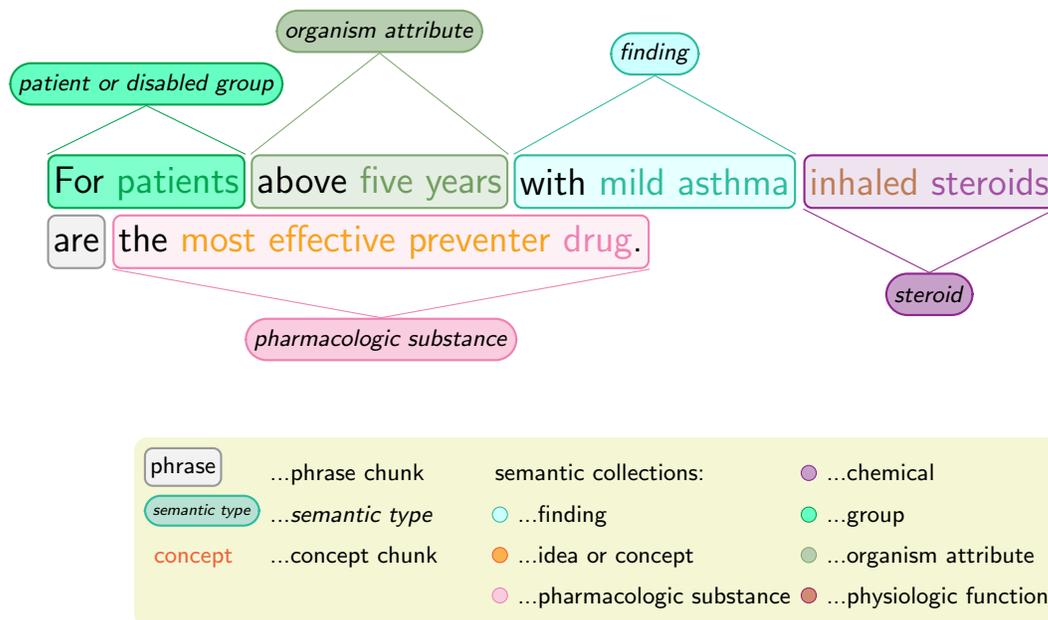
Figure 7: A semantic type has been assigned to each phrase chunk containing a medical concept.

2 Preliminaries

2.1 Clinical Practice Guidelines (CPGs)

> "Clinical practice guidelines are systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances." [4]

CPGs are instructions and recommendations on appropriate and effective treatment of patients with specific diseases, with the intention to improve the quality of health care.

2.1.1 Computerized Guidelines

The benefits of using CPGs are numerous and widely recognized, but the availability of patient-specific advice at the point-of-care turned out to be crucial for effectively using guidelines in clinical care [3]. Therefore, several groups and organizations are working at the development of computerized guidelines and decision support systems that incorporate these guidelines. Guidelines in a computer interpretable form offer great benefits, like

- providing recommendations for individual patients,

- not interrupting clinical workflow,

- improving the clarity of the text,

- helping to reveal inconsistencies and errors in guidelines,

- and providing automatical reminders for treatment or tests.

2.2 UMLS®- Unified Medical Language System®

The Unified Medical Language System is developed by the *National Library of Medicine*, USA, within the UMLS R&D project, initiated in 1986. It is a controlled compendium of many vocabularies and classifications of the biomedical domain and also provides a mapping structure between them. The UMLS was created to facilitate the developement of computer systems that process biomedical text by offering access to this knowledge.

2.2.1 The Metathesaurus®

The UMLS Metathesaurus is the largest thesaurus in the biomedical domain containing medical concepts from more than 100 vocabularies. It is built from numerous thesauri (e.g., MeSH, CRISP NCI), classifications (e.g., ICD-9-CM), clinical coding systems (e.g., SNOMED CT), and lists of controlled terms used in various biomedical documents. Consequently it is not built to be a single standard vocabulary, but enables exchange of information between different clinical databases and systems, in accordance with contextual and inter-contextual relationships between these diverse coding systems and vocabularies.

2.3 The MetaMap Program

Based on the SPECIALIST NLP tools NLM has developed the MetaMap programm (also referred to as the MMTx program, which stands for "Meta Map Transfer") [1, 2]. As a first step towards implementing a computer based application of a guideline, MetaMap detects noun phrases in biomedical text and assigns them to corresponding concepts in the UMLS Metathesaurus [7]. In doing so it goes through five steps.

1. Tokenization of arbitrary text into sections, sentences, and noun phrases.

2. Generation of variants for each phrase, i.e. all its spelling variants, acronyms, abbreviations, synonyms, inflectional and derivational variants and meaningful combinations of these.

3. Retrieving a candidate set of all Metathesaurus concepts containing at least one of the variants.

4. Evaluation of each candidate against the input text by computing the mapping strength of the candidate using a linguistically principled evaluation function and then arraying the candidates according to their mapping strength.

5. Combining candidates from disjoint parts of the noun phrase, recomputing the mapping strength for the combined candidates and forming a set of best Metathesaurus mappings for the original phrase.

   The output of MetaMap provides important information for knowledge engineers to better understand the medical text and its underlying concepts, which is a prerequisite for the further steps of transforming the text into a computer interpretable format.

3   Additional Tools

Besides the MMTx program and UMLS Metathesaurus, MapFace profits by the iUMLS program, developed by Katharina Kaiser, at the *Institute of Software Technology and Interactive Systems, Vienna University of Technology*. The iUMLS is an interface which provides a clearly arranged access to the MMTx program and its results as well as to information obtained from the UMLS database, such as relations between concepts and between semantic types.

In addition, MapFace takes advantage of the IR (Information Retrieval) program, developed by Karl-Michael Edlinger, Alex Hörmandinger, Matteo Savio, and Johannes Strodl, at the *Vienna University of Technology*. IR evaluates the most likely semantic types for a phrase from a given list of possible types, by testing each semantic type for relations to semantic types assigned to other phrases in the same sentence.

4   Paper Specific Expressions

In this user manual I will use the following expressions:

UMLS concept: A *UMLS concept* is a medical concept which is part of the UMLS Metathesaurus (see section 2.2.1). It is associated with one or more *semantic types*.

UMLS concept-candidate: The UMLS concept-candidates MapFace deals with are the best matching *UMLS concepts* for a given *concept chunk* or *phrase chunk* (see section 2.3).

Concept chunk: A *concept chunk* is a text chunk of a guideline matching one or more *UMLS concepts*.

Phrase chunk: A *phrase chunk* is a text chunk of a guideline detected to be a phrase by the MMTx program (see section 2.3). It may contain one *concept chunk*, many *concept chunks*, or no *concept chunk* at all.

Semantic type: A *semantic type* describes a medical concept, e.g. the *semantic type* for the medical concept "adrenaline" is "hormone".

Semantic collection: A *semantic collection* is a group of semantic types, e.g. the *semantic collection* "finding" contains the *semantic types* "finding", "laboratory or test result", and "sign or symptom". MapFace associates each *semantic collection* with a different color.

Semantic group: *Semantic groups* arrange *semantic types* in a way different from *semantic collections*.

Concept relation: A *concept relation* is a relation between two *UMLS concepts* obtained from the UMLS database.

Semantic relation: A *semantic relation* is a relation between two *semantic types* obtained from the UMLS database.

5 Installation Instructions

Installation instructions for Windows:

1. Install the MMTx program on your system. In the following its path will be refered to as
   "%MMTx_Path%".

2. Unzip the mapface.zip file. In the following its path will be refered to as "%MapFace_Path%".

3. Create a file "config.jar" from the contents of the directory
   %MMTx_Path%/nls/mmtx/config/.

4. Replace the "config.jar" file in the directory
   %MapFace_Path%/MapFace/plugins/at.ac.tuwien.ifs.ieg.mapface_1.0.0/
   with the previously created "config.jar" file.

6 Getting Started

After MapFace has been successfully installed, you open the editor by simply double-clicking
the mapface.exe application.

6.1 Tell MapFace about Your Document-Structure

In order to ensure clearness and readability, MapFace displays the text of the guideline (an
xml-document) in a structured way. Since the name of the structuring elements of an xml-file,
such as headlines and list-elements, can differ from one file to another, you can tell MapFace
the names of the xml-tags identifying headlines, list-elements, or simple text in the xml-file you
want to process.

   To do this you need to edit the "document_structure.txt" file in the directory
%MapFace_Path%/MapFace/plugins/at.ac.tuwien.ifs.ieg.mapface_1.0.0/,
where you can define the name of the xml-tag representing the first headline (h1), the second
headline (h2), and the third headline (h3). The text of these elements will have a bigger font-size
than the rest of the text. In addition it is possible to define the names of xml-tags representing
list-elements (iteration) and normal text (text).

   This is how the "document_structure.txt" file initially looks like:

```
h1 = guideline_title

h2 = null

h3 = section_title

text = ElementText

iteration = ElementList
```

   If you don't want to define a certain element, please set it to null, f. e. "h2 = null".
Furthermore, you should take care that there is no new line after the last element. Bear in mind
that MapFace only displays text between xml-tags (no tag-names or -attributes) to keep the
text tidy.

6.2 Edit the MapFace-Colors

MapFace reads the colors to use for highlighting from the "colors.txt" file at the directory `%MapFace_Path%/MapFace/plugins/at.ac.tuwien.ifs.ieg.mapface_1.0.0/`.
This file contains the RGB-values together with a color-name for each color. If you want to delete, add or change colors, you need to edit this file accordingly. Be sure to have the separation symbol "|" at the beginning and end of each line, as well as after the color-name and between the RGB-values.

For example, a line of the "colors.txt" file can look like this:

`|blue_velvet|214|181|255|`

For reasons of readability, don't use very dark colors, since the font-color is black. And remember: no new line after the last element!

7 The GUI

If you open MapFace, a window with a **menubar**, a **toolbar**, and three different panes appears: the **editor pane**, the **candidates pane**, and the *annotation scheme pane*.
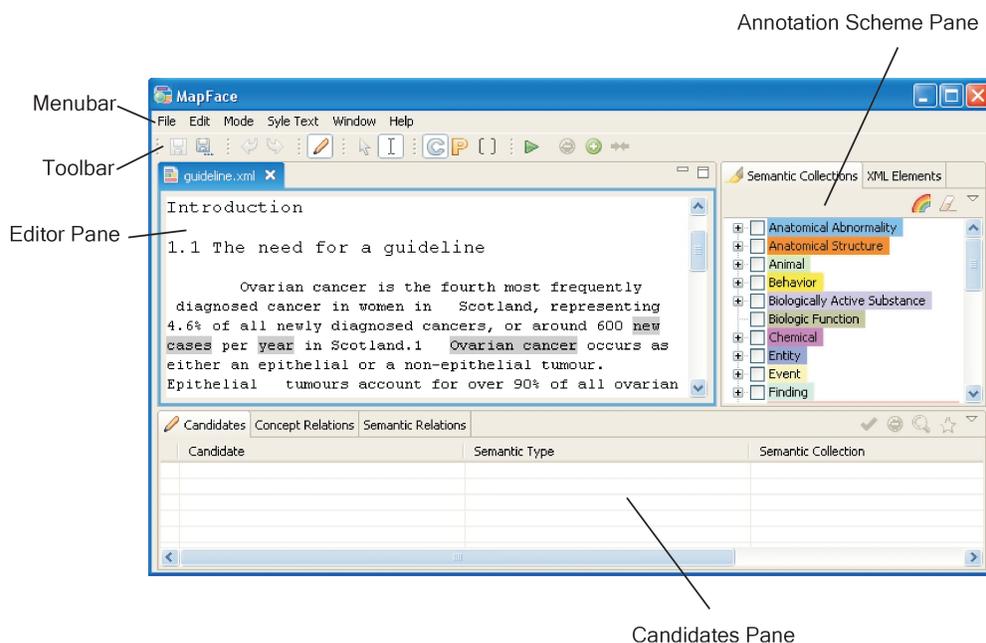
Figure 8: Components of the GUI.

7.1 The Menu Bar

All actions provided by the main toolbar of the window and in the local toolbars of the *annotation scheme pane* can be accessed through the menubar as well.

The menubar includes the following menus:

1. "File":

   - "Save" / "Save As.." (see section 9.4.5),
   - "Exit", and
   - "Open File..".

2. "Edit":

   - "Undo"/ "Redo" (see section 9.4.4),
   - "Run MetaMap" (see section 9.1.1),
   - "Delete Concept / Phrase" (see section 9.1.2),
   - "Create Concept / Phrase" (see section 9.1.3),
   - "Merge Phrases" (see section 9.1.4),
   - "Arrow Cursor" (see section 9.4.2), and
   - "Text Selection" (see section 9.4.3).

3. "Mode":

   - "Editing Mode" (see section 8.1.2),
   - "Concepts Mode" (see section 8.3.1), and
   - "Phrases Mode" (see section 9.1.2).

4. "Style Text":

   - "Highlight Text" (submenu, containing "Highlight All" and "No Highlighting" for semantic types and xml-elements (see section 9.3)), and
   - "Mark Phrases" (see section 9.4.1).

5. "Window":

   - "Reset Perspective" (see section 9.4.6).

6. "Help":

   - "FAQ" (see section 9.4.6), and
   - "User Manual".

7.2   The Tool Bar
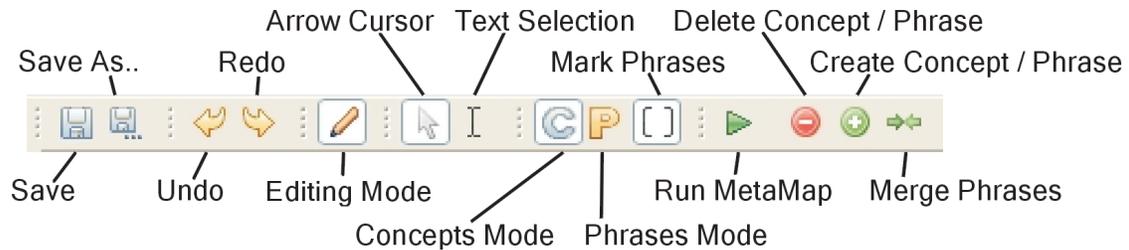
Figure 9: The toolbar.

7.3   The Editor Pane

The main window of the Graphical User Interface of MapFace is the **editor pane**. It displays the **text of the clinical guideline**. This is where you can select the text you want to process by means of the MMTx program.

In addition, you can select the concepts or phrases of processed text by double-clicking the text chunk of the concept or phrase in the editor, whereupon a list of best matching UMLS concept-candidates will be displayed in the *candidates view* (see section 7.4.1).

7.4   The Candidates Pane

The **candidates pane** is at the bottom of the user interface. It contains three different views, the *candidates view*, the *concept relations view*, and the *semantic relations view*.

7.4.1   The Candidates View

The main view of the *candidates pane* is the **candidates view**. It provides information about the best matching UMLS concept-candidates for a given text chunk and possibilities for editing. Therefore the *candidates view* displays a list of candidates detected by the MMTx program for a selected concept chunk or phrase chunk in the editor, together with their semantic types, semantic collections, and semantic groups. The lines of the table are highlighted according to the color coding of the corresponding semantic collections(see section 7.5.1).
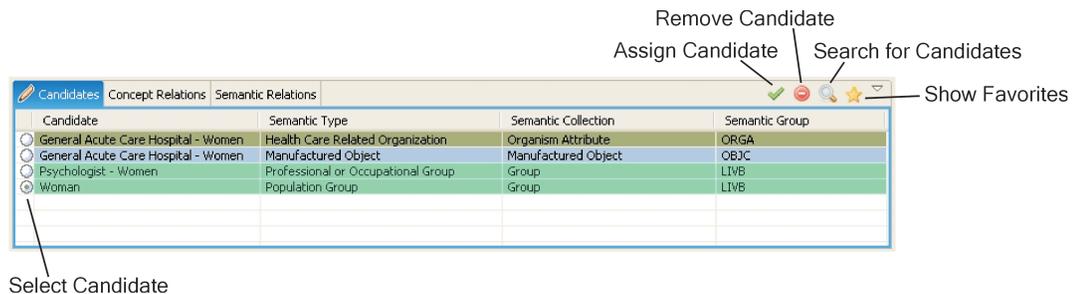
Figure 10: Candidates view.

14

Now you can select the right candidate from the list and assign it to the concept chunk or phrase chunk by clicking the "Assign Candidate" button (see section 9.2.1). If the appropriate candidate does not appear in the list you can look for additional candidates by clicking the "Search for Candidates" button (see section 9.2.3) and then enter an alternative expression for the text of the concept chunk. Furthermore you can remove candidates from the list by clicking the "Remove Candidate" button (see section 9.2.2). If the *phrases mode* (see section 8.3.2) is active, a decreased list containing the most likely candidates for the selected phrase chunk is displayed when clicking the "Show Favorites" button (see section 9.2.4).

7.4.2   The Concept Relations View

The **concept relations view** provides additional information for a UMLS concept-candidate selected in the *candidates view*. It displays a list of all relations between the concept-candidate selected in the *candidates view* and the UMLS concepts affiliated to concept chunks in the same section of the text. By selecting a relation from the list the two concerned concept chunks are highlighted in the editor. Relations between concepts are only available when the *concepts mode* (see section 8.3.1) is active.

7.4.3   The Semantic Relations View

The **semantic relations view** provides a list of all relations between the semantic type of the UMLS concept-candidate selected in the *candidates view* and the semantic types occuring in the same section of the text. The two concerned concept chunks or phrase chunks are highlighted in the editor by selecting a relation from the list.

7.5   The Annotation Scheme Pane

The **annotation scheme pane** is at the right of the user interface. The *annotation scheme pane* of the basic MapFace program contains two different views, the *semantic collections view*, and the *xml-elements view*. In both views you can select all elements in the view and accordingly highlight all associated elements in the editor by clicking the "Highlight All" button (see section 9.3.1), and also deselect all elements by clicking the "No Highlighting" button (see section 9.3.2).

7.5.1   The Semantic Collections View

The *semantic collections view* contains a **list of all semantic types**, grouped by semantic collections, each associated with a different color. Here you can select semantic types of interest, which will **highlight** every text chunk in the editor assotiated with one of these semantic types.
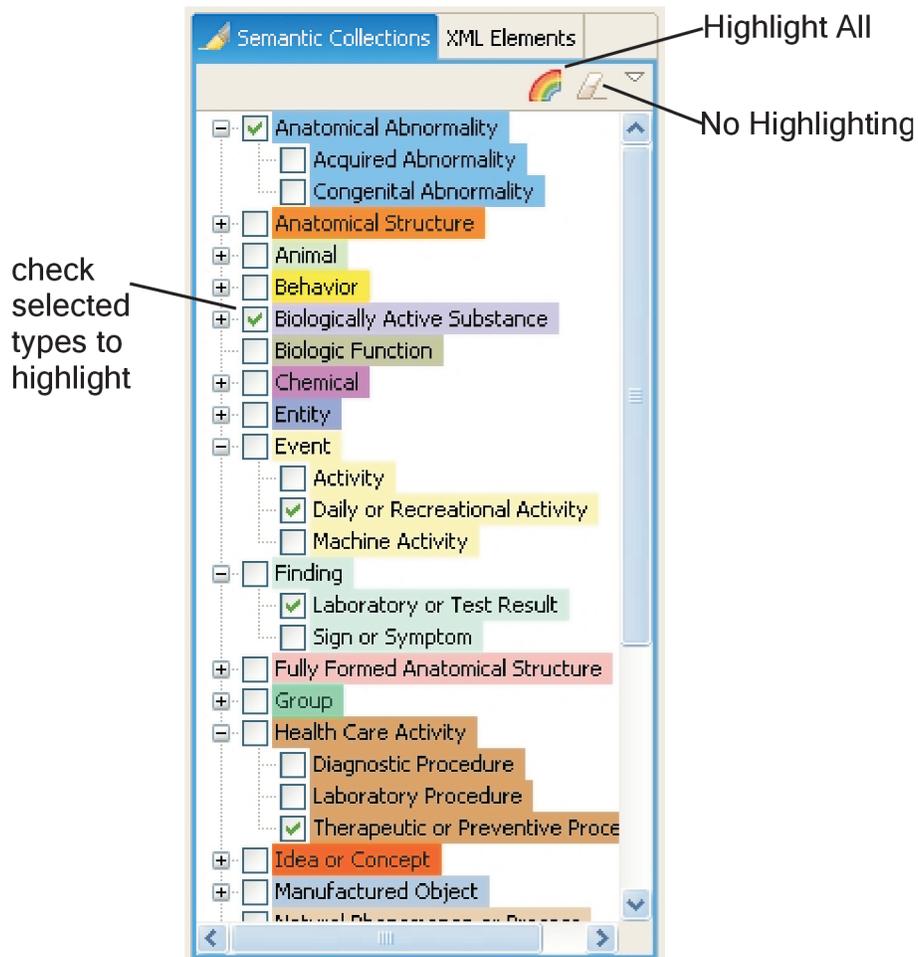
Figure 11: The semantic collections view.

7.5.2 The XML-Elements View

The *xml-elements view* contains all names of **xml-tags** occuring in the xml-document you are processing. By selecting an xml-element in the view you **highlight** the corresponding text chunk of the guideline in the editor.

8 Modes

8.1 Information Mode vs. Editing Mode

The **information mode** and the **editing mode** exclude each other, so either the *editing mode* or the *information mode* can be active at one time. Whenever the "Editing Mode" button in the toolbar, or the "Editing Mode" subitem in the *Mode* menu, is checked, the *editing mode* is active; otherwise the *information mode* is active.
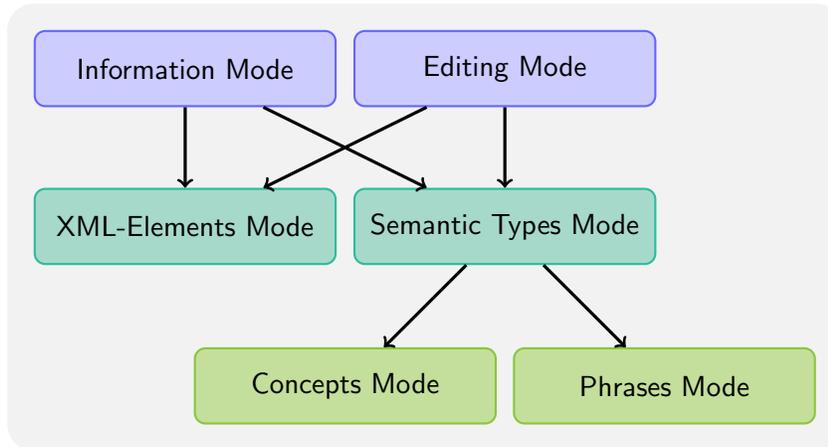
Figure 12: Modes

8.1.1 The Information Mode

The **information mode** is to serve the needs of knowledge engineers, when working with MapFace. Therefore, a straighforward representation of the information is provided by the following features:

- the **clinical guideline text** is displayed in the editor,

- the **assigned information** for a selected concept or phrase in the text is displayed in the *candidates view* (see section 7.4.1),

- concepts and phrases can be **highlighted** according to their **semantic types** (see section 9.3),

- all phrases existing in the document can be delimited by brackets (see section 9.4.1),

- **xml-elements** defined by **xml-tags** in the current xml-file can be **hightlighted** (see section 9.3), and

- **relations** between concepts and between semantic types can be **displayed** in the *candidates pane*, in combination with **highlighting** them in the text.

8.1.2 The Editing Mode

The **editing mode** is to serve the needs of physicians when working with MapFace. In addition to the features of the *information mode*, it offers possibilities to edit the generated phrases and concepts.

1. The following actions are available when working in the **concepts mode** (see section 8.3.1):

    - **running** the MetaMap program (see section 9.1.1),

17

- **deleting** an existing concept (see section 9.1.2),

- **creating** a new concept from selected text (see section 9.1.3),

- **choosing** the right concept from a list of equally matched metathesaurus concepts,

- **searching** for other metathesaurus concepts (see section 9.2.3),

- **assigning** candidates to concepts (see section 9.2.1), and

- **removing** concept-candidates from the list (see section 9.2.2).

2. The following actions are available when working in the **phrases mode** (see section 8.3.2):

- **running** the MetaMap program (see section 9.1.1),

- **deleting** an existing phrase (see section 9.1.2),

- **creating** a new phrase from selected text (see section 9.1.3),

- **merging** two adjacent phrases (see section 9.1.4),

- **choosing** the correct match from a list of semantic types for the phrase,

- **assigning** semantic types to the phrases of the guideline (see section 9.2.1), and

- **reducing** the list of candidates by using the "Show Favorites" action (see section 9.2.4).

8.2   Semantic Types Mode vs. XML-Elements Mode

8.2.1   The Semantic Types Mode

The *semantic types mode* is active every time the *semantic collections view* is visible in the *annotation scheme pane*. In this mode, you can further choose between the *concepts mode* and the *phrases mode* (see section 8.3).

8.2.2   The XML-Elements Mode

The **xml-elements mode** is active whenever the *xml-elements view* (see section 7.5.2) is visible in the *annotation scheme pane*. The *xml-elements view* displays a treestructure, representing all xml-elements occurring in the open xml-document (except the elements inserted by MapFace). In this mode you can select different xml-elements and accordingly highlight the text in the editor.

8.3   Concepts Mode vs. Phrases Mode

The **concepts mode** and **phrases mode** are only available if the *semantic types mode* (see section 8.2.1)is active. The two modes exclude each other, so either the *concepts mode* or the *phrases mode* can be active at one time.

These modes are accessible in the toolbar of the MapFace window and as subitems in the *Mode* menu.

8.3.1 Concepts Mode

When the **concepts mode** is active, all information, editing and highlighting actions refer to medical concepts assigned to the text by MetaMap. You can select a concept by double-clicking it in the editor. Then you may read the assigned information (see section 8.1.1), or process it (see section 8.1.2).

8.3.2 Phrases Mode

A phrase is a meaningfully arranged combination of words within a sentence, which may contain medical concepts. When the *phrases mode* is active, all information, editing and highlighting actions refer to phrase chunks detected in the text by MetaMap.

9 Features

9.1 Editor Actions

You can find the editor actions in the main toolbar of the MapFace window.

9.1.1 "Run MetaMap"

"Run MetaMap" is **the first step** to process a new xml-file. If you process text of your document with MetaMap, the text will be tokenized into sections, sentences, phrases, and concepts, and an amount of **syntactic and semantic information** will be computed for these elements. MetaMap primarily deals with the semantic information, the detected UMLS-Metathesaurus concepts. However, all the information will be saved in the ".idoc" file for later processing.

You can either **process the whole document at once**, by simply clicking the "Run MetaMap" button, or you can select text in the editor first (with the text selection cursor; see section 9.4.3), and then make MetaMap **process only the selected text**. This action will take some time, depending on the amount of text to process. If you don't want to wait so long, select smaller parts of the text.

You can undo and redo this action by using "Undo" / "Redo" (see section 9.4.4) in the toolbar of the MapFace window. After running MetaMap, you can edit the detected concepts or phrases by selecting the corresponding text chunks in the editor (using the *arrow cursor*; see section 9.4.2).

9.1.2 "Delete Concept / Phrase"

This action is thought to correct the tokenizarion of concept chunks and phrase chunks detected by MetaMap. To **delete an existing concept chunk or phrase chunk**, you select it in the editor and simply click the "Delete Concept / Phrase" button. If you delete a phrase this way, all included concepts will be deleted as well.

You can undo and redo this action by using "Undo" / "Redo" (see section 9.4.4) in the toolbar of the MapFace window.

9.1.3 "Create Concept / Phrase"

This action is thought to correct the tokenization too.

To **create a new concept chunk from a selected text**, the selected text must not contain any concept chunks. If this is the case, delete the concept chunks first (see section 9.1.2). Then select the appropriate text chunk with the *text selection cursor* (see section 9.4.3) and click the "Create Concept / Phrase" button.

To **create a new phrase chunk**, proceed the same way in the *phrases mode*. If you create a new phrase this way, included concept chunks will be created automatically.

You can undo and redo this action by using "Undo" / "Redo" (see section 9.4.4) in the toolbar of the MapFace window.

9.1.4 "Merge Phrases"

To **merge two adjacent phrase chunks into one phrase chunk**, you select the two phrases with the *text selection cursor* (see section 9.4.3) and then click the "Merge Phrases" button in the toolbar. You can undo and redo this action by using "Undo" / "Redo" (see section 9.4.4) in the toolbar of the MapFace window.

9.2 Candidates View Actions

9.2.1 "Assign Candidate"

To **assign a UMLS concept to a concept chunk** or a **semantic type to a phrase chunk**, you need to select the text chunk in the editor. A list of best matching candidates will appear in the *candidates view*, where you can choose the correct candidate by selecting the radio button to its left, and then assign it to the concept chunk or phrase chunk by clicking the "Assign Candidate" button.

You can undo and redo this action by using "Undo" / "Redo" (see section 9.4.4) in the toolbar of the MapFace window.

9.2.2 "Remove Candidate"

If you are sure that a UMLS concept-candidate does not match the selected concept chunk, and you want to keep the candidates list tidy, you can **remove the candidate from the list** by selecting the radio button to its left and clicking the "Remove Candidate" button.

This action is only available in the *concepts mode*. The list of UMLS concept candidates for a phrase consists of the candidates for the concept chunks included in the phrase.

You can undo and redo this action by using "Undo" / "Redo" (see section 9.4.4) in the toolbar of the MapFace window.

9.2.3 "Search for Candidates"

If the appropriate candidate does not appear in the candidates list, you can **search for additional UMLS concepts** for the selected concept chunk by clicking the "Search for Candidates" button and then entering an alternative expression for the concept text. This action is only available in the *concepts mode*.

You can undo and redo this action by using "Undo" / "Redo" (see section 9.4.4) in the toolbar of the MapFace window.

9.2.4 "Show Favorites"

"Show Favorites" is available in the toolbar of the *candidates view*, when working in the *editing mode* combined with the *phrases mode*. It is thought to facilitate choosing the correct semantic type for a phrase, by decreasing the number of semantic types in the list automatically, which is accomplished by taking advantage of the information about the semantic relations of each type. **Semantic types with no relation to other semantic types occurring in the same sentence will be removed from the list**. This may take a few minutes, since the favorite semantic types will be computed for all phrases in the same sentence at once. This action won't modify the original semantic type lists of the phrases, so if you select the phrase again, the original list will be displayed again.

9.3 Highlighting

For both views of the *annotation scheme pane* - the *semantic collections view* and the *xml-elements view* - there are options for selecting or deselecting all elements in the view by clicking a button, which effects the **highlighting of the corresponding text chunks in the editor**.

9.3.1 "Highlight All"

If you click the "Highlight All" button in the *semantic collections view*, or the *xml-elements view*, **all elements** in the view will be selected and the text chunks in the editor will be **hightlighted** accordingly.

In the *semantic collections view*, the "Highlight All" button stays checked, if you click it once, and will be unchecked, if you click it again. A checked "Highlight All" button has the effect that all text chunks in the editor with affiliated semantic type will be highlighted constantly. However, clicking the "Highlight All" button twice will once highlight all the text chunks in the editor assiciated with a semantic type, but assigning a semantic type to a text chunk afterwards will not affect the highlighting of this chunk.

9.3.2 "No Highlighting"

If you click the "No Highlighting" button in the *semantic collections view*, or the *xml-elements view*, all elements in the view will be deselected and the **highlighting** in the editor will be **cleared**.

In the *semantic types mode* (the *semantic collections view* is visible), concept chunks or phrase chunks with no assigned UMLS concept stay highlighted by means of a gray background, even if you clear the highlighting by clicking the button. The gray background indicates you to manually assign an UMLS concept to this chunk.

9.4 Additional Features

9.4.1 "Mark Phrases"

This action is available in the main toolbar of the MapFace window and in the *Style Text* menu. It deliminates all phrase chunks in the processed text by surrounding them with brackets, which adds quite a bit of clearity as to which concept chunks belong to which phrase chunk.

9.4.2 Arrow Cursor

Choosing the **arrow cursor**, you can select concepts or phrases in the editor by double-clicking the corresponding text chunk of the guideline. Depending on the mode (*concepts mode* or *phrases mode*), a list of UMLS candidates for a selected concept chunk, or all UMLS candidates of concepts included in a selected phrase chunk, will be listed in the *candidates view*.

9.4.3 Text Selection Cursor

Choosing the **text selection cursor** in the toolbar of the MapFace window or in the *Edit* menu of the menubar, you can select text in the editor for various purposes, e.g. processing it by the means of the MMTx program, creating a new concept chunk or prase chunk from selected text, or selecting two adjacent phrase chunks to merge them.

9.4.4 "Undo" and "Redo"

You can **undo and redo** each executed action by clicking the "Undo" or the "Redo" button in the toolbar or in the *Edit* menu of the menubar.

9.4.5 "Save" and "Save As.."

Saving the processed xml-document inserts xml-tags for sections ("mf_section"), sentences ("mf_sentence"), phrases ("mf_phrases") and concepts ("mf_concepts"), as well as for the text between these constructs ("mf_text"). For concept chunks and phrase chunks with affiliated UMLS concepts, these xml-tags contain the attributes "cui" (concept unique identifier) and "sem_type" (semantic type). Thus, the arrangement of these constructs in the documen-text is saved, together with some information about assigned UMLS concepts and semantic types. Since the output of the MMTx program provides much more syntactic and semantic information for processed text, a second file containing all computed information about these constructs is created. This file is saved to the same directory and has the same name as the xml-document with the extension "idoc". When you open the xml-document with MapFace again, the information of both files will be merged.

9.4.6 "Reset Perspective"

This action resets the MapFace GUI to the initial perspective, i.e. all initially visible views and their arrangement.

10   Detailed Examples

10.1   An Example for Physicians

This is an example of how a physician can use MapFace to process a clinical guideline.

1. Open an xml-document: To **open a file**, select the *File* menu in the menubar and choose
   "Open File..". Navigate through your disk system and choose an xml-document of a
   clinical guideline to open.

2. Get an impression of the **elements of the xml-document**: If you now activate the xml-
   elements view in the annotation scheme pane at the right of the window, the view displays
   a treestructure of xml-elements occurring in your document. By selecting an element from
   the tree, you highlight the corresponding text in the editor.

3. Change to **editing mode**: Checking the "Editing Mode" button in the toolbar will enable
   you to modify the guideline document.

4. Process selected text by means of the **MMTx** program: Activate the *semantic collections
   view* to change to *semantic types mode* again. Now you can select some text of the guideline
   in the editor and click the "Run MetaMap" button in the toolbar. The MMTx program
   will tokenize the selected text into sections, sentences, phrases, and concepts, and compute
   syntactic information as well as a list of matching UMLS concepts for each concept.
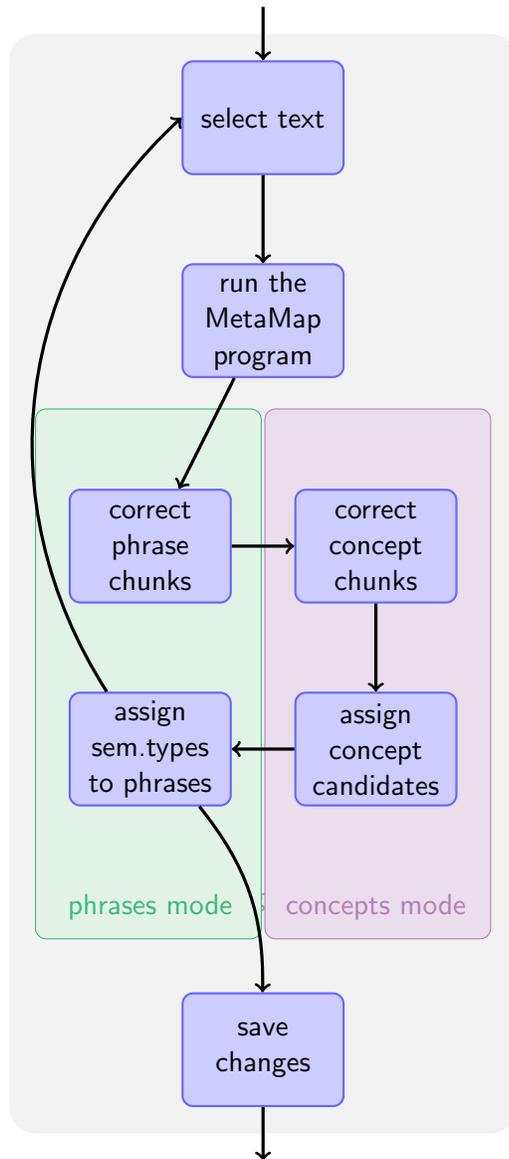
Figure 13: Example of processing a clinical guideline
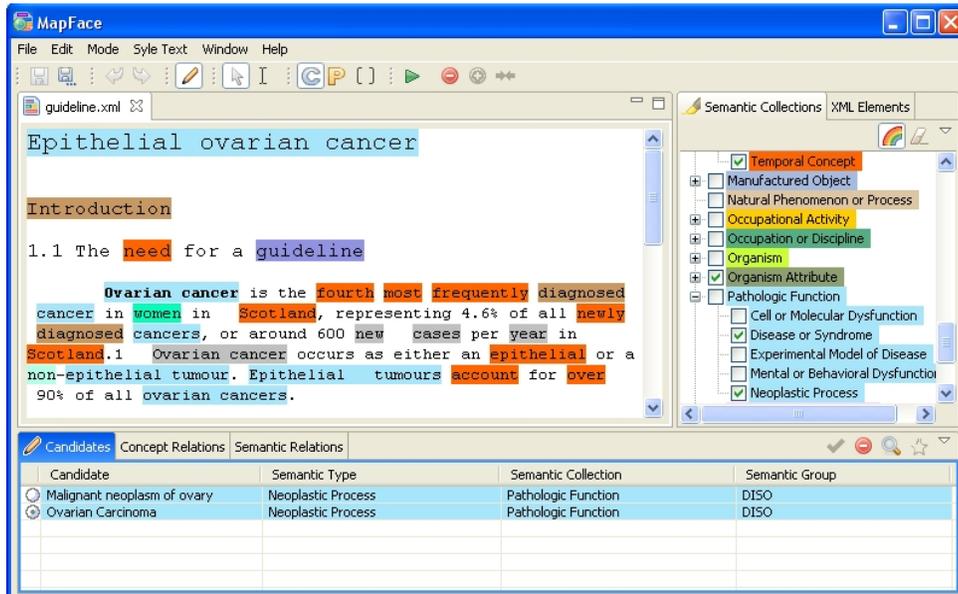
Figure 14: After runnning the MMTx program and highlighting the concept chunks.

5. **Highlight** the assigned semantic types: To make the result of the MMTx program visible, you can click the "Highlight All" button in the *semantic collections view.* All concepts with only one UMLS concept candidate will be highlighted with the color of the semantic collection the candidate refers to. The background of concepts with more than one UMLS concept candidate will be gray.

6. Mark Phrases: clicking the "Mark Phrases" button in the toolbar will **deliminate all phrases** in the processed text **with brackets**, indicating which concepts belong to which phrase.

7. Switch to the **phrases mode**: By checking the "Phrases Mode" button, you can switch to the *phrases mode.* Now the phrases in the processed text will be highlighted.

8. Correct phrase chunks (see Figure 16):

   (a) Remove phrase chunks and then add new chunks:
   - Choose the *arrow cursor* from the toolbar to **select a phrase** by double-clicking its text in the editor.
   - click the "Delete Concept / Phrase" button in the toolbar to **delete the selected phrase**. Repeat these two steps as often as is necessary.
   - Choose the *text selection cursor* to **select** the **text chunk** in the editor you want to create a new phrase chunk from.
   - **Create a new phrase chunk** from the selected text by clicking the "Create Concept/Phrase" button in the toolbar. Again repeat the last two steps as often as is necessary.

   (b) Merge two adjacent phrase chunks into one chunk:

- **Select these two phrases** with the "Text Selection" cursor, and
- **merge these phrases** into one phrase by clicking the "Merge Phrases" button in the toolbar.

Figure 15: Before and after correcting the phrase chunk "in Scotland.1 Ovarian cancer".
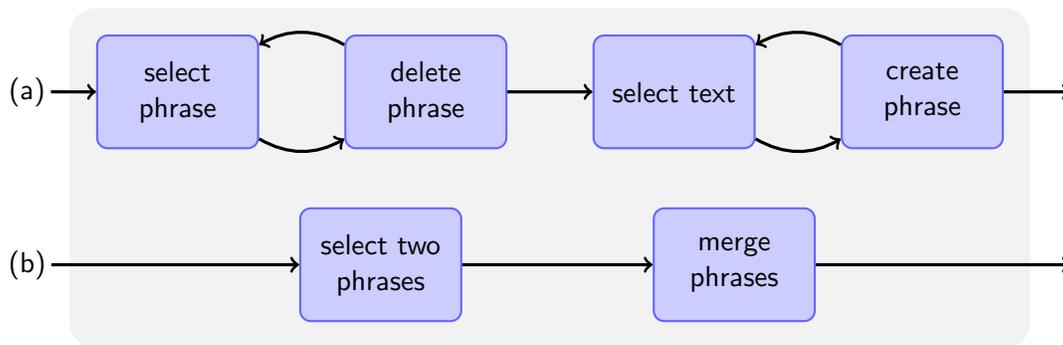
Figure 16: Correcting phrase chunks

9. After doing so for all incorrect phrase chunks, change to the **concepts mode** by clicking the "Concepts Mode" button in the toolbar.

10. Proceed with **correcting concepts chunks** in a similar way as you did with the phrase chunks (see figure 17):

- **Select a concept chunk** in the editor with the "Arrow Cursor".
- **Delete this concept chunk** by clicking the "Delete Concept / Phrase" button. Repeat these two steps as often as is necessary.
- **Select** the **text** you want to create a new concept chunk from with the "Text Selection" cursor, and
- **create a new concept chunk** by clicking the "Create Concept / Phrase" button. Again repeat the last two steps as often as is necessary.
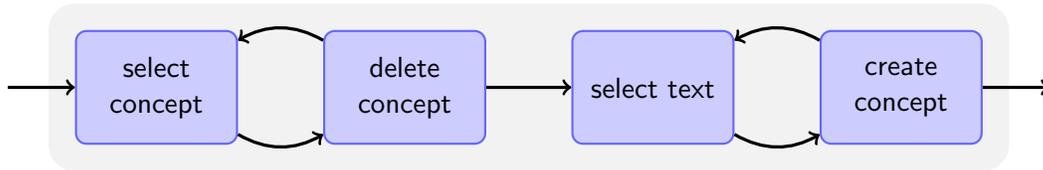
Figure 17: Correcting concept chunks.

11. After correcting the chunks in the text, you can start to assign UMLS concept candidates to the concept chunks (see figure 18)

   - **Select the concept chunk** in the editor, and
   - **select a candidate** in the *candidates view*.
   - Now you can take a look at the **relations of the selected concept candidate** to other concepts orruring in the same section of the text by activating the *candidate relations view*,
   - as well as activating the *semantic relations view* to take a look at the relations of the semantic type of the selected candiate and the semantic types of other concepts in the same section of the text.
   - If you are sure a candidate is not matching the selected concept and you want to keep the list tidy, **remove the candidate from the list** by selecting it and clicking the "Remove Candidate" button.
   - If the correct candidate does not appear in the list, you can **search for additional UMLS concept-candidates** by clicking the "Search for Candidates" button and entering an alternative expression for the selected concept.
   - Finally **assign the matching candidate** to the selected concept chunk, by clicking the "Assign Candidate" button.

12. Switch to the **phrases mode**.

13. Assign semantic types to phrases (see figure 19):

   - **Select a phrase chunk** in the editor.
   - Now you may **reduce the candidates list** to those candidates with most likely semantic types for the selected phrase by clicking the "Show Favorites" button.
   - By selecting a candidate from the list and activating the *semantic relations view* you can take a look at the **relations of the semantic type** of the selected candidate to the semantic types of phrases occuring in the same section of the text.
   - Finally **assign the semantic type** you consider to be correct to the phrase chunk by checking its radio button and clicking "Assign Candidate".
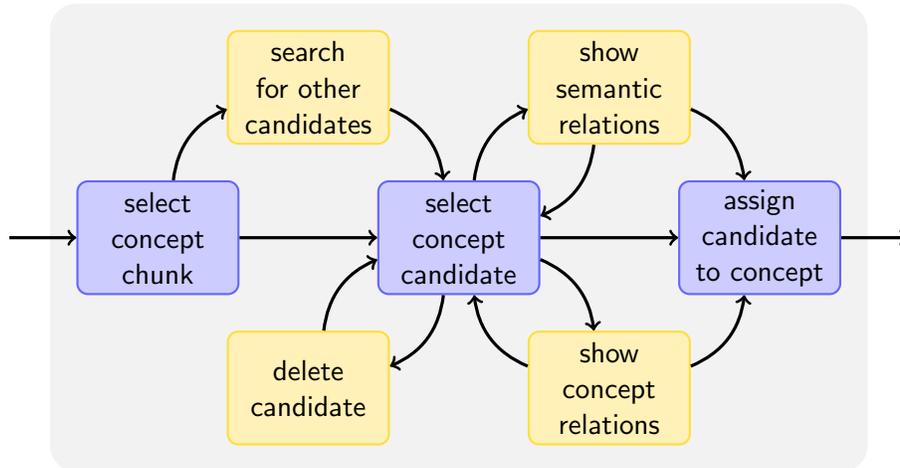
Figure 18: Assigning an UMLS concept candidate to a concept. Orange nodes indicate optional actions.
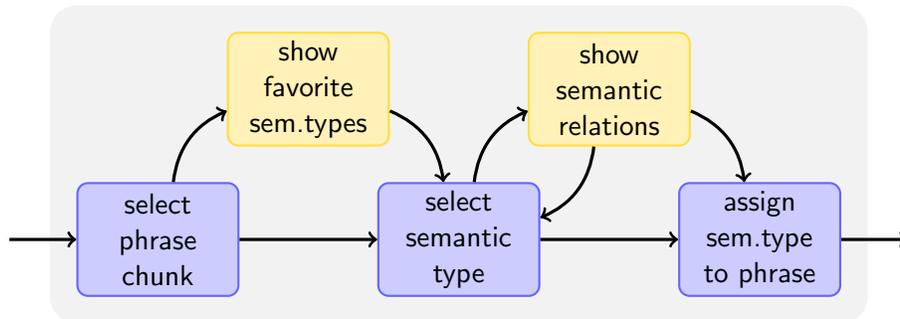
Figure 19: Assigning a semantic type to a phrase. Orange nodes indicate optional actions.

10.2   An Example for Knowledge Engineers

Ths is an example of how a knowledge engineer can use MetaMap to better understand the medical concepts of the guideline.

1. **Open** an xml-document of a guideline that has already been processed by a physician; this is the case if a file with the same name as the xml-document having the extension "idoc" exists. To load the information of both files, they have to be in the save directory.

2. If the *concepts mode* is currently not active, change to the *concepts mode* by clicking the "Concepts Mode" button in the toolbar.

3. **Highlight all concepts** in the text by clicking the "Highlight All" button in the *semantic collections view*. You can undo this by clicking the "No Highlighting" button next to it. In addition you can select single semantic types in the view to highlight only some semantic types of interest.

4. Choose the "Arrow Cursor" from the toolbar to **select a concept chunk** in the editor by double-clicking its text.

5. Now you can see the **information assigned** to this concept in the *candidates view.* By activating the *concept relations view* or the *semantic relations view* and selecting relations from the list, you can see the **relations of this concept** to other concepts in the same section.
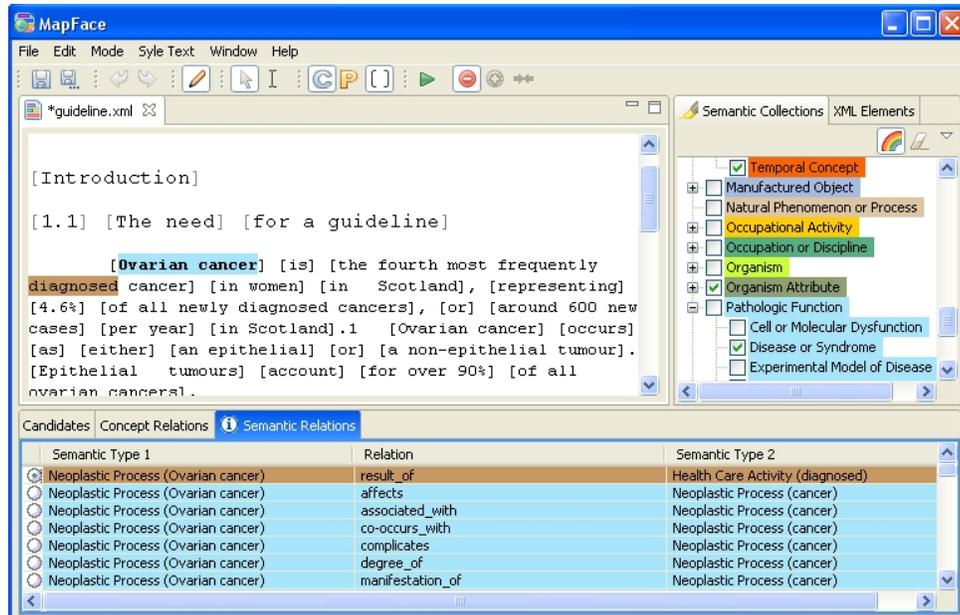
Figure 20: Semantic relations of the UMLS concept "Neoplastic Process" to other UMLS concepts affiliated to concept chunks in the same section of the text.

6. To make it clearer **which concepts belong to which phrases**, you may click the "Mark Phrases" button in the toolbar.

7. Now switch to the **phrases mode** by selecting the "Phrases Mode" button in the toobar.

8. You can now read the **information assigned to phrases**, take a look at the relations of the semantic type of a phrase to the semantic types of phrases in the same section of the text, and highlight phrases according to the color scheme of their semantic type.

References

[1] A. R. Aronson (2001): Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program. In: *Proc. of the AMIA Symposium 2001, pages 17-21.*

[2] A. R. Aronson (2006): MetaMap: Mapping Text to the UMLS Metathesaurus. http://skr.nlm.nih.gov/papers/index.shtml#MetaMap.

[3] P. L. Elkin, Mor Peleg, Ronilda Lacson, Elmer Bernstam, Samson W. Tu, Aziz Boxwala, Robert Greenes, Edward H. Shortliffe (2000): Toward Standardization of Electronic Guideline Representation, in: *MD Computing, 17(6):39–44.*

[4] M. J. Field, K. N. Lohr (ed.) (1990): Clinical Practice Guidelines: Directions for a New Program, National Academies Press, Institute of Medicine, Washington DC. http://www.nap.edu/books/0309043468/html/index.html

[5] K. Kaiser (2007): Medical Terminology Systems. Technical Report, Asgaard-TR-2007, Vienna University of Technology, Institute of Software Technology and Interactive Systems.

[6] The SPECIALIST NLP Tools Web Page. http://lexsrv3.nlm.nih.gov/SPECIALIST/index.html

[7] UMLS®Knowledge Sources, Documentation (2007). http://www.nlm.nih.gov/research/umls/umlsdoc.html