

Tovek Topic Mapper

Version 1.04

Copyright

Information contained in this document may change without notice.

© Copyright TOVEK. All rights reserved.

Tovek Topic Mapper is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

Tovek Topic Mapper is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with Tovek Topic Mapper. If not, see <http://www.gnu.org/licenses/>.

Registered Trademarks

Tovek Tools, Tovek Agent, Tovek Tools Fulltext Plug-in pro Analyst's Notebook, Tovek InfoRating, Tovek Report, Query Editor and Tovek Harvester are registered trademarks of TOVEK.

Analyst's Notebook is a registered trademark of i2 Limited.

Other Trademarks Recognition

Microsoft, Excel, Windows and Windows NT are registered trademarks of Microsoft Corporation.

All other companies and product names are trademarks or registered trademarks of their respective owners.

Tovek Topic Mapper version 1.04
Manual version 6, November 2007

TOVEK

<http://www.tovek.com>

<http://www.tovek.cz>

Contents

Introduction.....	4
Tovek Topic Mapper	4
System Requirements.....	4
Supported Formats	4
Tovek Topic Mapper Window	5
Document view	5
Vocabulary view.....	6
Property view	8
Layouts	9
Vocabularies	11
Creating a new vocabulary	11
Opening and saving vocabularies	13
Adding and modifying sections	14
Adding and modifying concepts	15
Removing concepts and sections	17
Custom concept association types	17
Full-text search	20
Importing and exporting vocabularies	23
Documents	24
Opening and saving documents	24
Importing a new document.....	24
Structure of a document.....	27
Exporting documents in Zthes format	28
Tagging	30
Basic tagging	30
Annotate the tagging associations	31
Remove tagging.....	31
Review tagging	32
Advanced tagging	34
Tagging modes	34
Tagging with multiple concepts versus multiple tagging with one concept	34
Application features	36
Workspaces	36
Application settings.....	36
Common commands.....	38

Introduction

Tovek Topic Mapper

Topic Mapper application has been by Tovek for the Calibrate project to support curriculum experts in mapping their curricula to common vocabularies and so to achieve interoperability of these documents.

Tovek Topic Mapper is platform independent because it is written entirely in Java. The application is compatible with the Topic Map ISO standard (ISO 132 50) and uses the standardized XTM data format for storing and exchanging data but it also supports other formats like zthes and xvd for vocabulary exchange.

In general the application offers two main functionalities:

1. *Mapping of documents*: users are able to identify concepts in a document and to associate them with concepts contained in several topic maps.
2. *Building topic maps of concepts*: users are able to create new topic maps of concepts of an arbitrary form (vocabulary, taxonomy, ontology etc.) and fill them with identified concepts and relations between them. In the further parts of this document these topic maps are called *vocabularies*.

Tovek Topic Mapper designed to support multilingual data, both documents and vocabularies.

System Requirements

Sun Java 1.6.0 or higher

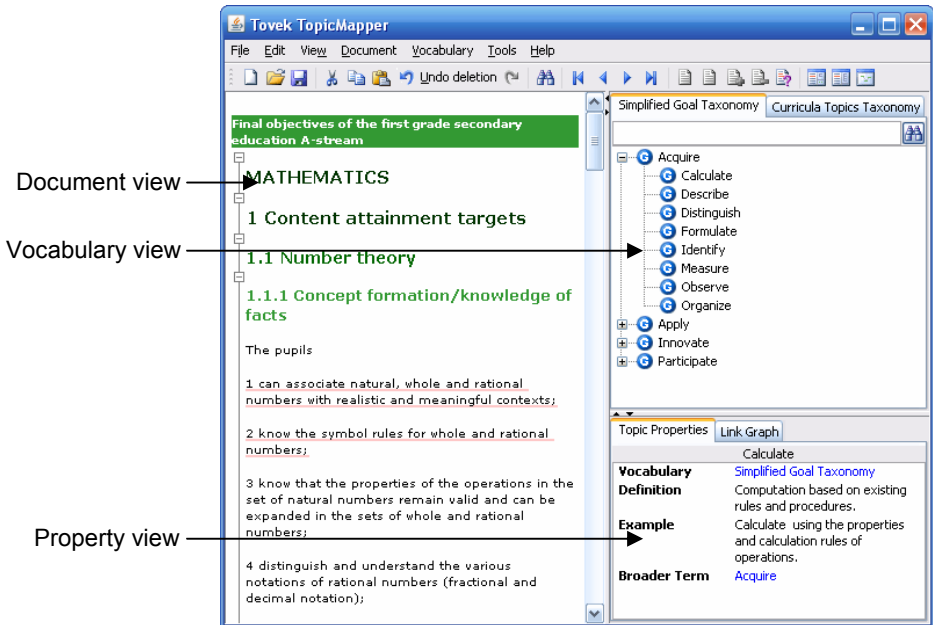
Supported Formats

Documents: XTM, HTML

Vocabularies: XTM, ZTHES, XVD

Tovek Topic Mapper Window

Window of the Tovek Topic Mapper application consists of three main parts:







Document view

The document view is located on the left side of the application window. It shows the text of the currently opened document with preserved formatting. In addition it highlights all already mapped text parts so that it is easy to recognize remaining parts than still have to be processed.


On the left border of the document view there is a handle that shows the specified document structure. Using this handle you can easily collapse or expand whole document sections so that also very long texts can be processed comfortably. To expand/collapse document sections you have to click the displayed plus/minus sign.


Tooltips along this view show overview information about the mapping of the text piece under the mouse cursor.

The document view supports you in selecting pieces of text by automatic recognition of its paragraphs. You can freely select any part of the document for tagging or you can navigate through the text paragraph by paragraph just by clicking the appropriate toolbar buttons:

-  Click the button **First Paragraph** (Ctrl+Shit+9) or run the **Document/First Paragraph** command to navigate to the document first paragraph and to select it.
-  Click the button **Previous Paragraph** (Ctrl+9) or run the **Document/Previous Paragraph** command to navigate to the previous paragraph of the currently selected document part and to select it.
-  Click the button **Next Paragraph** (Ctrl+0) or run the **Document/Next Paragraph** command to navigate to the following paragraph of the currently selected document part and to select it.
-  Click the button **Last Paragraph** (Ctrl+Shit+0) or run the **Document/Last Paragraph** command to navigate to the document last paragraph and to select it.

To select the whole paragraph, which is under the mouse cursor, hold the Ctrl key down and click in the text. Similarly you can select the paragraph where the text cursor is located by running the command **Document/Select Paragraph** or pressing Ctrl+G. The same command can be found in the document context menu.

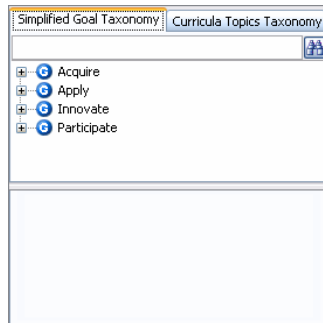
 The document view allows you also to select multiple discontinuous parts of the opened text. To do so select part by part and use the command **Tools/Lock Selection** (Ctrl+L) or **Lock Selection** in the context menu.

 To remove locked selections run the counterpart command **Tools/Unlock Selection** (Ctrl+U) or **Unlock Selection** in the context menu which unlocks all previously locked text selections.

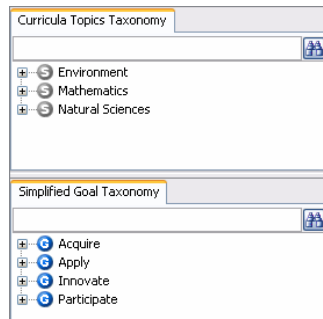
Vocabulary view

The vocabulary view is located in the top right corner of the application window. It displays all opened vocabularies each in separate tab panel. Vocabularies are displayed in a tree form according the hierarchical relation *Broader-Narrower Term*. All terms on one level are sorted alphabetically.

The vocabulary view can be splitted into two tab views so that you don't have to switch between vocabularies that often. To split the view or to undo the splitting run the **View/Toggle Secondary Vocabularies** command. An empty pane appears under the vocabulary tab view.



To move a vocabulary between the two tab views put focus into its view and run either the **View/Move Vocabulary Down** command or **View/Move Vocabulary Up** command. The place of each vocabulary is later on stored within the workspace.



If the full-text search is enabled (default is enabled) there is a text field on the top of the vocabulary view where you can type your query. To run the query press the *Search* button on the right side of the text field or just press *Enter* key (see Full-text search).

This view allows you to create and change vocabularies and to use existing vocabularies for tagging of documents.

Customizing the vocabulary tree view

In the case that you are working with very large vocabulary but you are interested only in a small part, you can customize the vocabulary view

to display only this part. In order to change displayed concepts run one of the following commands:

1. The command **Vocabulary/View/Hide Sub Tree** or the context menu command **View/Hide Sub Tree** hides the selected concepts together with the sub trees rooted at these concepts.
2. The command **Vocabulary/View/Add As Root Node** or the context menu command **View/Add As Root Node** adds all the selected concepts to the top level so that you can start the browsing directly with these concepts.
3. To undo all your changes run the command **Vocabulary/View/Restore Vocabulary Tree** or the context menu command **View/Restore Vocabulary Tree**. These commands restore the vocabulary view to its original state.

The defined layout is NOT stored within the vocabulary but it is saved as parts of the configuration of the current workspace. When you open a workspace you will find vocabularies exactly in the same state as when you stored it, but when you just open a vocabulary, it will appear in its default state.

Note: You do not change the vocabulary data at all by customizing its tree view.

Property view

The property view displays properties of the currently selected item in two different ways – as a text and as a graph. The graphical part always shows the associations of the selected item to other items and the textual part displays details about its occurrences and associations.

Topic Properties

Both of these views allow you to navigate to associated items and to explore the graph of topic associations. In the textual part you can click the titles of associated items and so to navigate to them and to display their properties within the property view. Instead of clicking you can also use the context menu command **Go To**. When the association to the selected item has been reified then you can also navigate to the reifying topic to learn details about the relation. This can be done by running the command **Go To Details** in the context menu.

Link Graph

Browse the link graph to explore relations between vocabulary concepts, concepts and bits of documents.

To move the graph in the window just drag it with mouse while holding down the left mouse button.

To center a graph node within the layout of the graph click it with left mouse button. This will also select the appropriate concept within the vocabulary view or the highlight the corresponding document bit.

To expand a node hold down the *Ctrl* key and click it with left mouse button.

To center a node and remove all other nodes except of its narrowest neighbor nodes hold down the *Shift* key and click it with left mouse button.

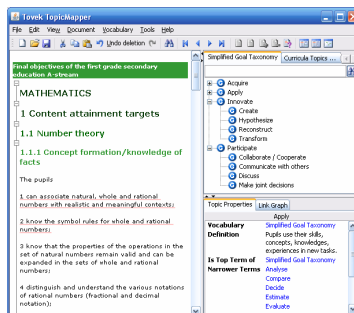
To get more information about a connection move the mouse cursor above it and wait for a tooltip.

Layouts

There are three predefined layouts of the Tovek Topic Mapper window:

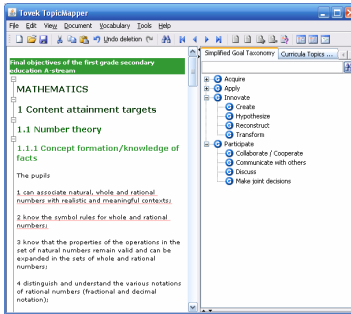


Standard layout shows all described window components; use the **Layout Standard** button or the **View/Layout Standard (Alt+1)** command.

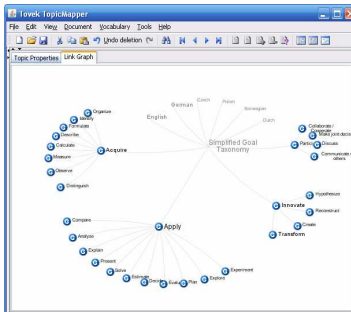




Tagging layout shows all only vocabularies and document views; use the **Layout For Tagging** button or the **View/Layout For Tagging (Alt+2)** command.



Browsing layout shows only graph view; use the **Layout For Browsing** button or the **View/Layout For Browsing (Alt+3)** command.



Vocabularies

Vocabularies are the basic data structures which contain extracted concepts and their relations. They are displayed in the vocabulary view.

Vocabulary in the context of Topic Mapper means a system of concepts associated together with either hierarchical or symmetrical relations. The concepts may be organized into vocabulary sections. A vocabulary section may contain subsections so that there might be a tree structure of vocabulary section defined for more precise grouping of concepts. Vocabulary sections can not be used for tagging of documents.

By default each vocabulary contains following association types that can be used to connect concepts together:

- *Broader-Narrower association* is a hierarchical association that is used to organize concepts in a tree like structure. This association is the basis for the tree view of the vocabulary.
- *Related association* is a symmetrical association for connecting concepts across the tree structure. It is a very simple association type because it does not say how related the concepts are. For semantic richer associations you have to define your own association types that match exactly your needs.
- *Preferred-Deprecated association* is used to deprecate concepts. It is not always possible to remove concepts from a vocabulary especially when they are involved in existing relations to other concepts or to document parts. In this case you should deprecate the concept instead of removing it.

Creating a new vocabulary

To create a new vocabulary run the command **Vocabulary/New Vocabulary....** This will open a dialog for entering metadata about the new vocabulary:

- *Title* is the default name of the vocabulary. It is language independent. This name will be used if there isn't any title for the current user language. It will be displayed as the vocabulary tab title.
- *Authority* is the name of owner of the vocabulary.

- *Version* defines the version number of the vocabulary.
- *PSI Prefix* is the prefix of the public subject indicators that will be generated for each concept when added to the vocabulary. You have the possibility to specify here the whole PSI prefix by typing "http://PSI-Prefix-Domain/PSI-Prefix" or you can skip the PSI-Prefix-Domain. In the later case the default PSI domain "http://psi.tovek.cz/" will be used. If you don't specify any PSI Prefix a random one will be generated for you.
- *Default Language* of the vocabulary is the only one setting that can not be changed later. All concepts stored in the vocabulary need to have a title in this language.
- *Allow Multiple Terms in Tuple Mode* is an additional feature implemented for the Calibrate project that specifies whether you can select one or more concepts from the vocabulary while tagging in so called tuple mode
- *Section and Term Icon*. Each vocabulary can have specific icons for its sections and concepts. You can select any .gif file as the definition of the icons otherwise default icons will be used.
- *Supported Languages*. In this part of the Vocabulary Properties dialog you can select all languages that should be supported by the vocabulary and you can specify its title and description for each of these languages. Later on you will be able to define names and metadata in the scope of these languages for all vocabulary sections and concepts.

Click Ok button to create a new vocabulary or cancel the creation process. You can change most of the above specified properties by opening the vocabulary properties by running the command **Vocabulary/Vocabulary Properties....**

The created vocabulary is empty. It does not contain any section and any concepts. It contains only the predefined concept association types.

Vocabulary Properties

Title:
Simplified Goal Vocabulary

Authority:
Tovek

Version:
Version 1.0

PSI Prefix:
goal_vocabulary/goal_

Default Language:
English (en)

Allow Multiple Terms in Tuple Mode

Icons

Section: Change...

Term: Change...

Supported Languages

Language:
English (en)

Title:
Simplified Goal Vocabulary

Description:
Goal Vocabulary for Calibrate project.

Language Is Supported

Ok Cancel

Opening and saving vocabularies

Once you create a vocabulary or you open existing one you can save it to a file using the command **Vocabulary/Save Vocabulary**. For new vocabularies the destination file must be specified, for already saved vocabularies the original file will be used. To save a vocabulary to a specific destination file you also can use the command **Vocabulary/Save As Vocabulary...**

Modified vocabularies are marked with an asterisk.

To open an existing vocabulary run the command **Vocabulary/Open Vocabulary...** (Ctrl+Shift+V) and select the file containing it.

Adding and modifying sections

Adding a new top section to a vocabulary can be done using the command **Vocabulary/Add Top Section...** (Ctrl+Shift+O) or using the context menu command **Add Top Section....** This command opens a dialog where you are able to specify all section properties.

The only one required input value is the title of the section in the vocabulary default language. If your user language is the same as the vocabulary default language you will have to fill the text field *Title* and the second text field *Translation to Default Language* will be disabled. In the other case the translation field is enabled and you have to enter the default title there.

To specify titles of the section in other supported languages press the button *Other....* It opens a dialog with a language to title table where you can fill in new translations or change existing ones.

By default a PSI for new section is generated automatically but the next section of the dialog allows you to define your PSI for the section topic.

The table below the title fields contains properties of the current edited section. If you are adding a new top section, you will find there the relation *Is Subsection Of* and the vocabulary name. The property is displayed in red color which indicates that it is a new or changed value. You can add new section properties or change existing ones using the context menu or just by clicking the cells of the table.

Add New Section

Title:
Section 1

Translation to Default Language (en):
Other...

User defined PSI:
http://psi.tovek.cz/section_1

Properties:

Is Subsection Of	Action Verbs
Definition	This is a definition of the section 1. en

Ok Cancel

For a section you can specify the following properties:

- Container of the section, it can be either the vocabulary or another section
- Multiple definitions in different language scopes.

Adding a subsection of an existing section is very similar to adding a top section. It can be done by running the command **Vocabulary/Add Subsection...** (Ctrl+O) or the context menu command **Add Subsection....** Both of these commands are enabled only if you have selected exactly one section within the vocabulary view. They open the same dialog as described above and you have to perform the same steps to specify the section.

To modify existing vocabulary section select it in the vocabulary view and run the command **Vocabulary/Modify...** (Ctrl+M) or the context menu command **Modify....**

Adding and modifying concepts

Concepts can be added in the same way as vocabulary sections. To add a new top concept – concept that is directly connected to a vocabulary or a vocabulary section – run the command **Vocabulary/Add Top Term...** (Ctrl+Shift+P) or the context menu command **Add Top Term....** These commands open a dialog which is very similar to the section dialog.

Add New Term

Title:

Translation to Default Language (en):

User defined PSI:

Properties:

Is Top Term of	Action Verbs
Example	Example used to illustrate the Concept 1 en
Definition	Definition of the Concept 1 en

Analogue to the section definition you have to specify the title of the new concept in the vocabulary default language. You can also specify names in other by the vocabulary supported languages by pressing the button *Other...* and filling the language to title table.

In the dialog section below you can specify the PSI of the term topic. To do so, select the checkbox **User defined PSI** and enter a valid URI. If you do not specify a PSI so it will be generated automatically using vocabulary's PSI prefix.

The property table allows you to add or modify concept following properties:

- Container of the concept, it can be either a vocabulary or a vocabulary section
- Multiple definitions of the concept in different language scopes.
- Examples for demonstrating the usage of the term.
- Associations to other concepts - either predefined or custom associations. To add an association to another concept use the sub menu of the context menu item **Add Association To** and select the appropriate association role of the other concept.

To add a sub concept of an existing concept run the command **Vocabulary/Add Subterm...** (Ctrl+P) or the context menu command **Add Subterm....** They open the same dialog as the previous two commands and you have to perform the same steps to define the new concept.

You can also start adding a new concept by dragging a piece of text and dropping it to the appropriate place within the concept tree.

After adding a concept to a vocabulary the tree view has to be reloaded and the collapsed/expanded states of tree nodes get lost. If you are going to introduce more than one concept to the same container and you want to omit finding the container within the tree for each of the concepts, just hold *Ctrl* key down while pressing the *Ok* button of the Add New Term dialog. This way you add the concept to the vocabulary but you prevent reloading the tree view. After you have added all concepts, use the context menu command **View/Reload View** to update the vocabulary tree.

To change an existing concept or its relation to another concept select it in the vocabulary view and run the command **Vocabulary/Modify...** (Ctrl+M) or the context menu command **Modify...**

Removing concepts and sections

Only sections that do not contain any subsections or concepts can be deleted.

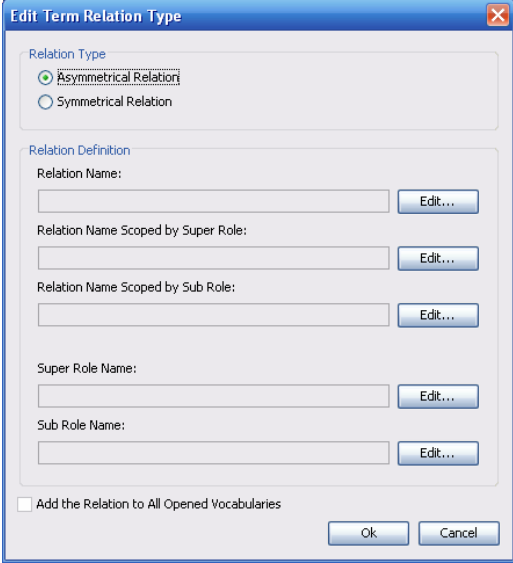
Only those concepts that are not associated with more than one other concept or section can be deleted. But generally concepts should not be deleted once the vocabulary has been used for tagging of documents because this could lead to dead links. Instead of removing concepts you should deprecate them using the *Preferred-Deprecated association*.

To remove a vocabulary section or a concept, select it and run the command **Vocabulary/Remove...** or the context menu command **Remove....**

Custom concept association types

Tovek Topic Mapper gives you the possibility to create your own association types for your vocabularies. You can of course use the predefined associations and you should use the *Broader-Narrower Association* because it is now the only one association that is used to build the tree for the vocabulary view.

But the predefined association types are almost without any semantic information. So for more precise definition why two concepts are related you should define your custom association type.



Edit Term Relation Type

Relation Type

Asymmetrical Relation

Symmetrical Relation

Relation Definition

Relation Name:

Relation Name Scoped by Super Role:

Relation Name Scoped by Sub Role:

Super Role Name:

Sub Role Name:

Add the Relation to All Opened Vocabularies

The application supports two types of associations – Asymmetrical and symmetrical. An association type can be added by running the command **Vocabulary/Add Term Relation Type...** which opens the *Edit Term Relation Type* dialog.

To add a new hierarchical association type select the radio button *Asymmetrical Relation* and fill in the definition:

- *Relation Name*: Default name of the relation can be entered in all supported languages. (Example: Broader-Narrower Term Relation)
- *Relation Name Scoped by Super Role*: Name that is displayed when the associations of a super concept are listed. It can also be specified for all supported languages. (Example: Has Narrower Terms)
- *Relation Name Scoped by Sub Role*: Name that is displayed when the associations of a sub concept are listed. You can enter alternatives for all supported languages. (Example: Has Broader Term)
- *Super Role Name*: Name of the role of the superior concept in all supported language modifications. (Example: Broader Term)

- **Sub Role Name:** Name of the role of the subordinate concept, can also be defined for all supported languages. (Example: Narrower Term)

If you want to add symmetrical association type select the radio button *Symmetrical Relation*. The dialog changes its interface and you can fill in the definition:

- **Relation Name:** Default name of the relation can be entered in all supported languages. (Example: Related Term Relation)
- **Relation Name Scoped by Role:** Name that is displayed when the associations of an involved concept are listed. Can also be supplied for all supported languages. (Example: Related Terms)
- **Role Name:** Name of the role of the involved concept in all supported language modifications. (Example: Related Term)

If you want to introduce the new association type to all currently opened vocabularies, select the check box *Add the Relation to All Opened Vocabularies*. Otherwise the relation will be added only to the last focused vocabulary.

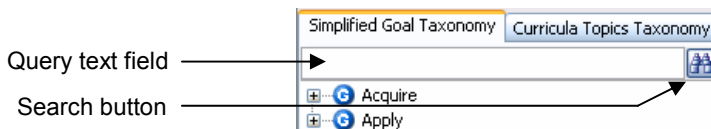
To start editing an association type run the command **Vocabulary/Edit Term Relation Type....** This command lets you first select the association type to edit and then to start editing it in the previously described dialog. You can change any of the entered names but you can not change its type (hierarchical/symmetrical).

Full-text search

Full text search is a feature that allows you to find faster and easier concepts that you want to use for tagging. It is build upon the open source full-text search engine Lucene .

The current Topic Mapper application does not support precomputed indexes so if the full text search is enabled, each vocabulary will be indexed when it is loaded. For very big vocabularies the indexing process can take a long time so in this case it is recommended to switch off this feature.

You can access the search functionality from two different places. The first possibility is simply to enter the query into the search text field that is displayed at the top of each vocabulary tree.



To start a local search press *Enter* key or push the search button. Local search is restricted only to the active vocabulary. It searches all section and concept titles and other textual metadata (definitions, examples), highlights the best matching item and displays its properties within the Property view.

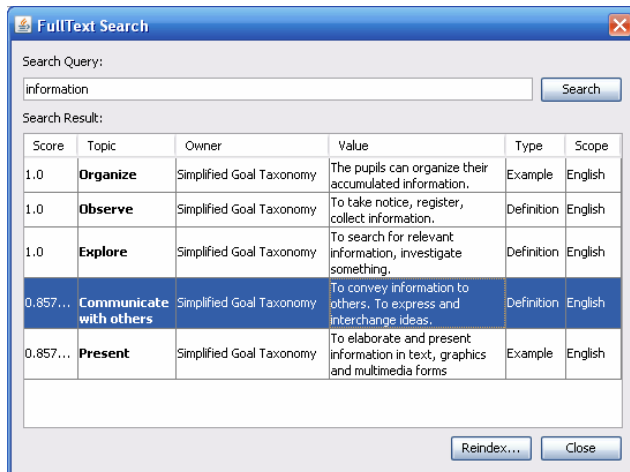


If you need to search through all opened vocabularies you have to use global full-text search. It is accessible through the command **Tools/Search Vocabularies...** or through the toolbar button **Search Vocabularies....** Both of them open a non modal dialog where you can run your queries.

This dialog shows the result of the search in a table containing following columns:

¹ <http://lucene.apache.org/>

- *Score* is a value between 1.0 and 0.0 where 1.0 is the best match.
- *Topic* is the topic whose title or metadata match the query.
- *Owner* is the name of the vocabulary that owns the matching topic.
- *Value* is the topics concrete value that has matched the query. If several topic values match the query, the topic appears multiple times within the table.
- *Type* is the type of the matching value.
- *Scope* is the scope of the matching value, mostly its language.



You can also run the global query directly from the vocabulary view. To do so, type your query into the vocabulary query text field, hold *Shift* key down and press the search button.

The currently used full-text indexes are static and so it is not possible to change them once a vocabulary has been indexed. If you change the vocabulary and you want that the indexes reflect you changes, you have to reindex. Reindexing can be started by pressing the *Reindex...* button of the FullText Search dialog.

Query syntax

Detailed description of the query syntax accepted by the Lucene search engine can be found at <http://lucene.apache.org/java/docs/queryparsersyntax.html>. This section provides only brief introduction.

A query is broken up into terms and operators. There are two types of terms: *Single Terms* and *Phrases*. A Single Term is a single word such as "information". A Phrase is a group of words surrounded by double quotes such as "query syntax".

Multiple terms can be combined together with Boolean operators to form a more complex query.

Lucene supports single and multiple character wildcard searches within single terms (not within phrase queries). To perform a single character wildcard search use the "?" symbol, to perform a multiple character wildcard search use the "*" symbol.

The single character wildcard search looks for terms that match that with the single character replaced. For example, to search for "text" or "test" you can use the search: `te?t`

Multiple character wildcard searches looks for 0 or more characters. For example, to search for test, tests or tester, you can use the search: `test*`. You can also use the wildcard searches in the middle of a term but you can not use them at the beginning.

Lucene supports fuzzy searches based on the Levenshtein Distance, or Edit Distance algorithm. To do a fuzzy search use the tilde, "~", symbol at the end of a Single word Term. For example to search for a term similar in spelling to "roam" use the fuzzy search `roam~`. This search will find terms like foam and roams.

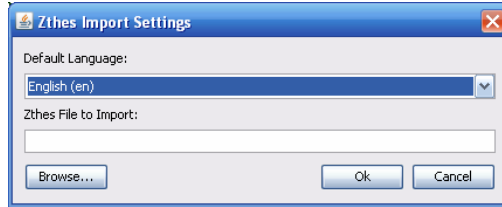
Lucene supports finding words are a within a specific distance away. To do a proximity search use the tilde, "~", symbol at the end of a Phrase. For example to search for a "apache" and "jakarta" within 10 words of each other in a document use the search `jakarta apache~10`.

Boolean operators allow terms to be combined through logic operators. Lucene supports AND, "+", OR, NOT and "-" as Boolean operators (Note: Boolean operators must be ALL CAPS).

Special characters that are part of the query syntax must be escaped. The list of special characters is "+ - && || ! () { } [] ^ " ~ * ? : \".

Importing and exporting vocabularies

Tovek Topic Mapper supports also standard exchange format for thesauri called Zthes².



Vocabularies stored in this format can be imported into the application using the command **Vocabulary/Import Export/Import from Zthes...**, which opens the import dialog.

Here you have to specify the default vocabulary language by selecting one of the supported languages listed in the top combo box. Then the input file must be specified. To start importing the file press the *Ok* button. When the import has been finished a new vocabulary tab appears in the vocabulary view. Now you should check the extracted properties of the new vocabulary.

The application also supports exporting vocabularies in the zthes format. To export a vocabulary make sure that it is active and run the command **Vocabulary/Import Export/Export to Zthes...** and select the destination file.

Note: Both import and export are not necessarily lossless.

² <http://zthes.z3950.org/>

Documents

The current version of the Tovek Topic Mapper allows only one opened document at time. It is the document that you want to tag with concepts from your vocabularies.

Opening and saving documents

Each document that has been processed with Tovek Topic Mapper is stored as Topic Maps within two files. The first file contains document's metadata, definition of its structure and its text with formatting. This file is different for each document. The second file, called *Link File*, can be shared between several documents and it stores the tagging information.

To save a document to a file run the command **Document/Document Save** or **Document/Document SaveAs....** If you are saving a new document or you are running the second command the application asks you to select the destination file. This way you can specify the location of the document but not of the link file. The link file location is given through the documents' properties. Both commands save the document and the link file.

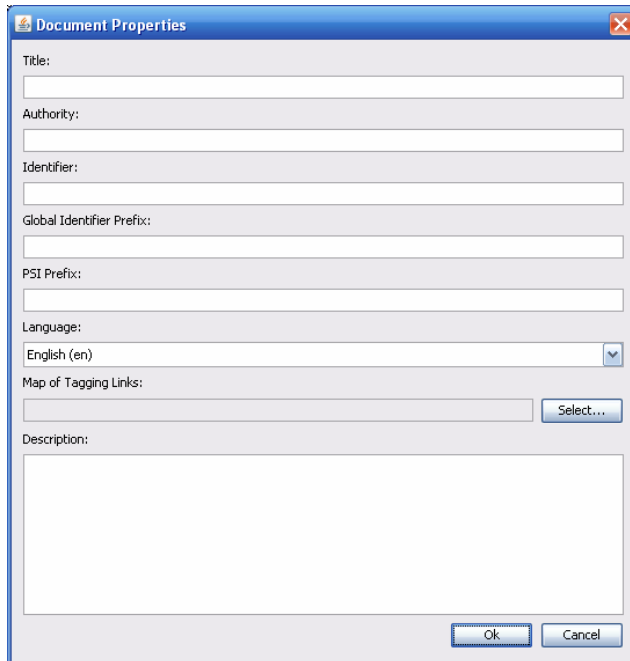
To open a document run the command **Document/Document Open...** (Ctrl+Shift+D) and select the source file. The document and its link file will be loaded into the application. When the link file is missing, the document will be still opened but tagging will be disabled until you specify a correct link file to use.

Importing a new document

Before you can start tagging a document you must first import it to the application. The only supported format now is HTML.

To import an HTML document into the Tovek Topic Mapper run the command **Document/Import Document...** (Ctrl+I) and select the source HTML file.

Once you have selected the file the application opens a *Document Properties* dialog and asks you to fill some metadata about the new document.

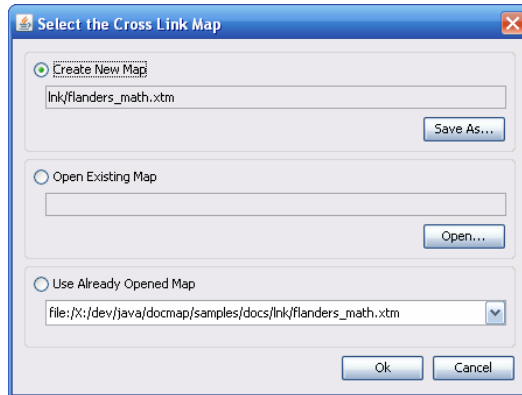


The screenshot shows a 'Document Properties' dialog box with the following fields and controls:

- Title:
- Authority:
- Identifier:
- Global Identifier Prefix:
- PSI Prefix:
- Language:
- Map of Tagging Links: - Description:
-

- *Title* is the first required parameter. In contrast to a vocabulary a document has only one title. It is assumed that it is in the language of the document.
- *Authority* is the name of the owner of the document.
- *Identifier* is the local identifier of the document. This value is especially required during the export of document in zthes format.
- *Global Identifier Prefix* builds together with document's identifier a globally unique identifier for the document.
- *PSI Prefix* is the prefix of the public subject indicators that will be generated for each document section and text bit that arise during tagging sessions. You have the possibility to specify here the whole PSI prefix by typing "http://PSI-Prefix-Domain/PSI-Prefix" or you can skip the PSI-Prefix-Domain. In the later case the default PSI domain "http://psi.tovek.cz/" will be used. If you don't specify any PSI Prefix a random one will be generated for you.

- *Language* of the document has later on impact on creating new vocabulary concepts using parts of the text. You can select one of the supported languages.
- *Map of Tagging Links* is the document's link file. To select a link file press the button *Select...* and use one the three possibilities given in the following dialog:



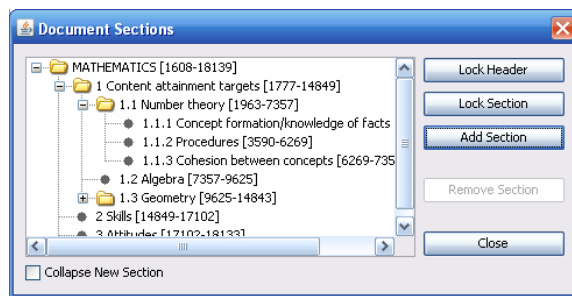
- *Create New Map* is the possibility to create a new empty link file. Use the button *Save As...* to specify the destination file.
 - *Open Existing Map* allows you to share link files between several documents. Press the *Open...* button to select the link file.
 - *Use Already Opened Map* shows you a list of already opened link files where you can select the one that should be used for the document.
- *Description* field allows you to add notes to the document.

Press the *Ok* button to complete the import of the document. You can change the specified properties of the document by running the command **Document/Document Properties...** which opens the same dialog as described above.

Structure of a document

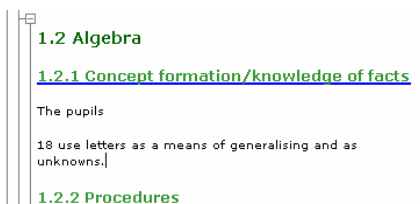
Newly imported document does not have any structure. Although it is not necessary for documents to have a structure it is recommended to define it especially for long documents, because it allows you afterwards to collapse/expand its sections. This way you get a better overview about what part of the document you are currently tagging, you can faster and easier navigate through its text.

Structure of a document can be defined or modified any time by running the command **Document/Document Structure...** (Ctrl+D) or the context menu command **Document Structure...**. Both change the mode of the document view to display the current structure as a tree and open a non modal dialog that allows you to add or remove document sections.

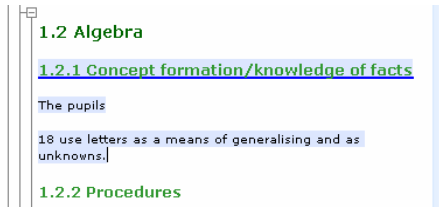


To add a new section follow the next steps:

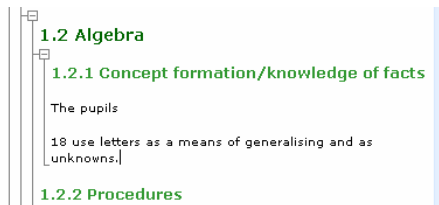
1. Select the header of the section and press the button *Lock Header*. The section will start at the beginning of the selected header and selected text will be displayed as the title of the section. The selected header is underlined.



2. Move the cursor to the end of the section and press the button *Lock Section*. The body of the section will be highlighted.



- Confirm adding of the section by pressing the button *Add Section*. The section will be added and a new leaf of the section tree appears in the document view. If the checkbox *Collapse New Section* is selected the section is in collapsed state initially.



To remove existing sections select them in the section tree of the dialog and press the *Remove Section* button.

To complete the definition of the structure of the document close the dialog by pressing the *Close* button.

Exporting documents in Zthes format

Tovek Topic Mapper allows exporting tagged documents in Zthes format. It is a special feature added for the Calibrate project to achieve seamless interoperability with the Calibrate portal.

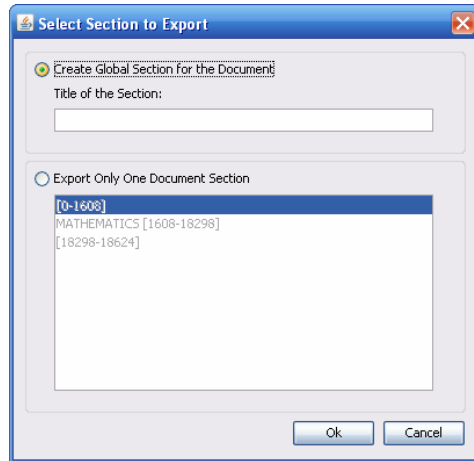
To start the export run the command **Tools/Export To Zthes...** and select the destination file. If the exported document has only one main section then it will be directly exported.

In the other case a dialog appears which lets you select a section of the document to export or to define a dummy top document section:

- To create a new global section, select the first radio button *Create Global Section for the Document* and enter the title of this section. The new section will be created just for the

purpose of the export and as soon as the export will be finished it disappears again.

- To export only one top section of the document, select the radio button *Export Only One Document Section* and select the desired section from the displayed list.



To finish the export press the *Ok* button.

Tagging

Tagging in the sense of the Tovek Topic Mapper means assigning concepts from shared vocabularies to pieces of document.

Basic tagging

To tag a part of a document follow the next steps:

1. Select the part of the document by highlight the text (see Document view).
2. Select concepts that you want to assign to the text.
3. Tag. If you want to assign concepts from only one vocabulary to the selected part of the document you can simply drag them from the vocabulary view and drop them onto the text.

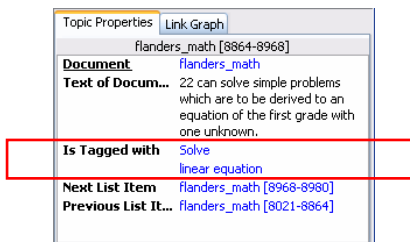


But if you wish to use concepts from more than one vocabulary you have to use the command **Tools/Tag Selection** or the toolbar button **Tag Selection** to do the tagging.

The tagged part of the text will be underlined in order to highlight text parts that have already been processed.

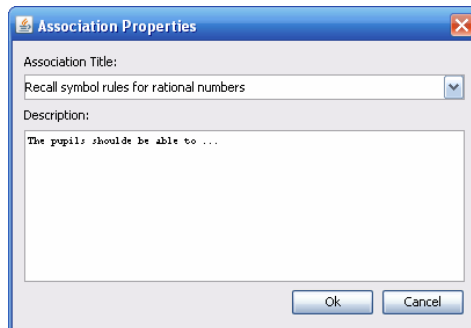


If you want to recall what concepts have been used to tag a given part of the document, select this part and press the toolbar button **Selection Properties...** or run the command **Tools/Selection Properties...**. All the information about the selected document part will be then displayed within the Property view.



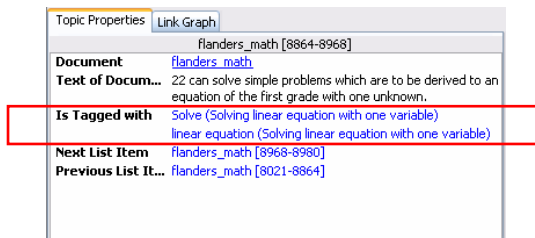
Annotate the tagging associations

If Tovek Topic Mapper is running with default settings it does not ask for annotation of each tagging association. If you want to annotate why you are linking the selected part of the document with selected concepts, hold the *Ctrl* key down and do the tagging as described above. Then the Topic Mapper opens a dialog where you can enter a title of the association and the description.



When you repeatedly use the same set of concepts for tagging the top combo box will be filled with all previously typed titles. You can either select one of these titles or write a new one.

The title that you have entered will be shown in selection properties in brackets behind the tagging concepts.



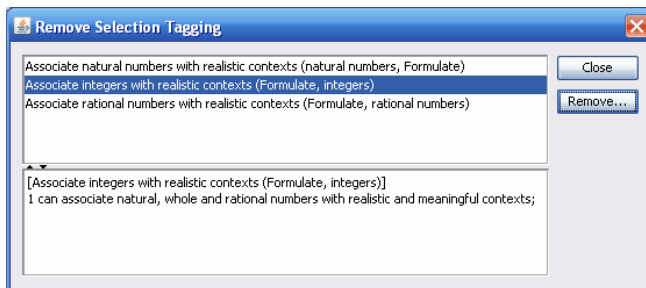
To learn more about the tagging association, right-click it and run the context menu command **Go To Details**.

Remove tagging



To remove an existing tagging association from a document, select the involved document text (or just a part of it) and press the button

Remove Selection Tags... or run the command **Tools/Remove Selection Tags...** (Ctrl+R). A dialog appears containing a list of all tagging associations in which the selected part of the document is involved.



If you select one of these associations then a list of all document parts involved is displayed in the bottom window. To remove the tagging association press the *Remove...* button and confirm the deleting. But be careful because by removing an association you delete its tagging information from *all* involved document parts, not just from the selected one!

To finish removing tagging association close the dialog using the *Close* button.

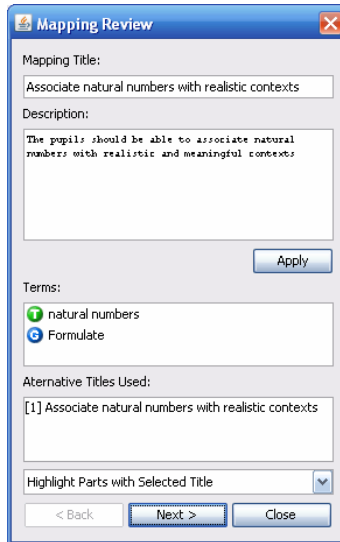
Review tagging

To check the tagging that has been made for the current document and correct it if necessary run the command **Tools/Review Document Mapping...**

A Mapping Review dialog appears. Initially it displays information about the first tagging association:

- *Mapping Title*: If you have entered a title for the current tagging association then it will be displayed here. You are free to change it. To make your changes persistent press the *Apply* button.
- *Description*: Displays your description of the tagging association. Also this text can be modified and is stored by pressing the *Apply* button.
- *Terms*: List of involved concepts.

- **Alternative Titles Used:** If the same set of concept is involved in more than one tagging association then this windows might be of interest. It displays all titles assigned to these associations together with the count for the case that a title has been used several times. This feature can be used to consolidate the association titles.



To review for which parts of the document are involved in tagging associations with the same set of concepts, use the combo box below the *Alternative Titles Used* window:

- **Highlight All Parts:** If this value is selected then all document part involved in tagging association with the same set of concepts will be highlighted.
- **Highlight Parts with Selected Title:** If this value is selected you can click titles listed in the *Alternative Titles Used* to highlight the appropriate parts of the document.
- **No Highlights:** Nothing will be highlighted

Use the *< Previous* and the *Next >* button to navigate through the tagging associations. Complete the tagging review by clicking the *Close* button.

Advanced tagging

Tagging modes

For the Calibrate project the Tovek Topic Mapper application supports so called *Tuple Mode* of tagging.

In this mode the user is forced to select a tuple of concepts from at least two different vocabularies to be able to add a new tagging association.

A valid tuple is actually an n-ary set of concepts which contains according to the type of the underlying vocabularies only one or unlimited count of concepts.

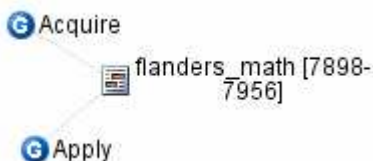
The type of a vocabulary is defined in vocabulary properties through the setting *Allow Multiple Terms in Tuple Mode* (see Creating a new vocabulary). If this setting is enabled (checkbox selected) it is possible to choose more than one concept from the vocabulary to build a tuple. In the opposite case, the tuples can contain only one concept from the vocabulary.

To switch between the tuple and the normal tagging mode change the application setting *Allow Tagging with Tuples Only* (see Application settings)

Tagging with multiple concepts versus multiple tagging with one concept

Assume the scenario where one piece of a document should be tags with two different concepts. This can be done in two different ways:

1. After selecting the text part, it is tagged with the first concept and then it is tagged with the second one. Tagging the text this way leads to two new tagging associations:



2. The second possibility is to mark the two concepts and tag only once. This way only one association is created. The semantics of such tagging different from the previous one. It means that the pair of the two concepts describes the text, not the concepts on their own:



Similar scenario can be found when several text parts are tagged (see Document view on how to select multiple document parts). The way how text parts should be handled during tagging can be changed by the application settings by changing the *Tag Separate Document Parts Separately* property.

If the property is disabled for each selected distinct part of the document a new tagging association will be created. In the opposite case only one tagging association is created. This association involves all selected document parts.

Combined with the two ways of tagging four different results can be achieved by tagging multiple document parts with multiple concepts!

Application features

Workspaces

Workspace contains information about currently opened vocabularies and document, about the state of the views etc. You can save this configuration into a workspace file. This file can be loaded again and the application state will be restored so that you can start your work at same point where you have interrupted it last time.



To store the workspace run the command **File/Save Workspace** (Ctrl+S) or press the **Save Workspace** toolbar button. To change the location of the workspace file use the command **File/SaveAs Workspace....**



Saved workspace can be loaded by running the command **File/Open Workspace...** (Ctrl+O) or by pressing the **Open Workspace** toolbar button.



The command **File/New Workspace** of the toolbar button **New Workspace** will close all opened vocabularies and documents and prepare the application to start new work.

Application settings

Run the command **Tools/Settings...** to open a dialog for changing the application settings. This dialog is divided into four sections.

The first section is the *Language Settings* section. Here you can select the user language for the application. This setting is important when the vocabularies and tagging details are displayed. If a vocabulary supports your language than all concepts that have a translation for your language will be displayed correctly. All other concepts will be displayed in the vocabulary default language and according to the Appearance settings marked with a different icon. Concepts from vocabularies that do not support the user language will be displayed in vocabulary default language. To change the user language use the combo box containing a list of all by the application supported languages.

The next section is the *Appearance* section which contains following settings:

- *Expand section by default:* Each workspace contains data about which document section has been expanded or collapsed. If this data is missing that the document view behaves according to this setting.
- *Highlight terms and sections with missing translation to user language:* If this property is enabled then all concepts and vocabulary section that do not have translation into the selected user language will be displayed with a special yellow icon so that they can be easily recognized and the translation of the whole vocabulary completed.
- *Show term approval state:* This feature is not supported in the current version of the Tovek Topic Mapper.

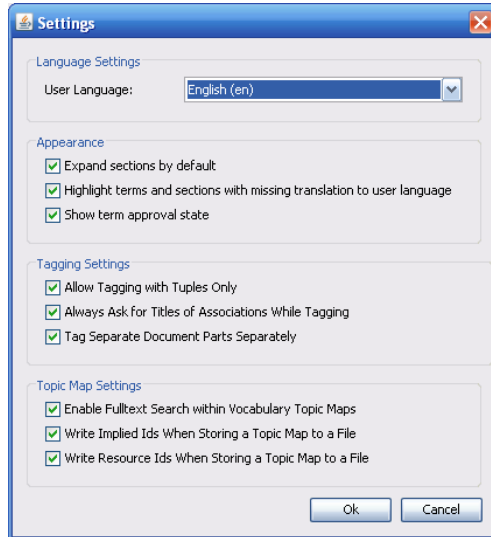
The third section is the *Tagging Section* that defines properties for tagging:

- *Allow Tagging with Tuples Only:* Enable this setting to change to the tuple tagging mode. (see Tagging modes)
- *Always Ask for Titles of Associations While Tagging:* Using this setting you can change the default behavior of the application during tagging. The default value is disabled. If you enable it the application will always display the dialog for annotating the tagging association when you tag a part of the document, except when you hold down the *Ctrl* key. (see Annotate the tagging associations)
- *Tag Separate Document Parts Separately:* is a property that change the tagging behavior described in the section Tagging with multiple concepts versus multiple tagging with one concept.

The last section of the dialog contains settings for handling of Topic Maps:

- *Enable Fulltext Search within Vocabulary Topic Maps:* If you want to use full-text search for the vocabularies then enable this feature (see Full-text search).
- *Write Implied Ids When Storing a Topic Map to a File:* Default value is enabled. You can disable this feature to reduce the size of the Topic Map files but it should be done carefully otherwise the linking information can be lost.

- *Write Resource Ids When Storing Topic Map to a File*: Default value is enabled. You can disable this feature to reduce the size of the Topic Map files but it should be done carefully otherwise the linking information can be lost.



Common commands

Tovek Topic Mapper application has limited support for the common edit commands which are located in the **Edit**:



Cut (Ctrl+X)



Copy (Ctrl+C)



Paste (Ctrl+V)



Undo (Ctrl+Z)



Redo (Ctrl+Y)

Note: These commands have only limited functionality now but later releases will handle this issue.

Known Issues

- Defining sections of a document does not work properly if the html source code of the document contains simple tags like `
`