AutoMap User's Guide

CASOS Technical Report

Kathleen M. Carley, Jana Diesner and Matt De Reno[1]

October 2006

CMU-ISRI-06-114

Carnegie Mellon University
School of Computer Science
Institute for Software Research International (ISRI)
Center for Computational Analysis of Social and Organizational Systems (CASOS)

Abstract

AutoMap is software for computer-assisted Network Text Analysis (NTA). NTA encodes the links among words in a text and constructs a network of the linked words. AutoMap subsumes classical Content Analysis by analyzing the existence, frequencies, and covariance of terms and themes.

[1] This work was supported in part by the Office of Naval Research under Contract No. N00014-06-1-0772, ONR, and N00014-06-10921, by the National Science Foundation IGERT in CASOS, the Air Force Office of Sponsored Research with a MURI with George Mason University under Grant No. 600322GRGMASON, and the Army Research Lab under Grant No. DAAD19-01-2-0009. Additional support was provided by the center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the National Science Foundation, the Office of Naval Research, the Air Force Office of Sponsored Research, the Army Research Lab or the U.S.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
OCT 2006		**00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
AutoMap User's Guide	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Carnegie Mellon University,Computer Science Department,Pittsburgh,PA,15213**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**110**	

Keywords: Semantic Network Analysis, Dynamic Network Analysis, Mental Models, Social Networks, AutoMap

Table of Contents

Automap: An Overview

1. Network Text Analysis (NTA)
2. Semantic Network Analysis
3. Social Network Analysis (SNA)
4. Dynamic Network Analysis (DNA)

Automap Graphical User Interface (GUI)

Text Examples

Load Input

1. Load a single text
2. Load a set of texts

Text Analysis Utilities

1. Browse through texts
   1.1 Example for Browse Menu
2. Create and refresh Concept List
   2.1 Example for Concept List
3. Create and refresh Union Concept List
   3.1 Save Union Concept List
   3.2 Example for Union Concept List

Text Pre-Processing

1. Introduction to Text Pre-Processing in AutoMap
2. Hierarchy of Pre-Processing Techniques
3. NLP (Natural Language Processing) Utilities
   3.1 Named-Entity Recognition
   3.1.1 Example for Named-Entity Recognition
   3.2 Symbol RemovalN-Gram Identification: Bi-Grams
4. Stemming
   4.1 Stemming Example
5. Delete List
   5.1 Open a Delete List

   5.1.1 Small predefined Delete List
   5.1.2 Extensive predefined Delete List

   5.2 Create a Delete List
   5.3 Edit a Delete List

5.4 Apply a Delete Lis
5.4.1 Direct Adjacency
5.4.2 Rhetorical Adjacency

5.5 Un-apply Delete List
5.6 Save a Delete List
5.7 Save text(s) after application of Delete List
5.8 Example for the Application of a Delete List

1. Generalization Thesaurus
   6.1 Open a Generalization Thesaurus
   6.2 Create a Generalization Thesaurus
   6.3 Edit a Generalization Thesaurus
   6.4 Apply a Generalization Thesaurus

   6.4.1 Thesaurus content only
   6.4.1.1 Direct Adjacency
   6.4.1.2 Rhetorical Adjacency

   6.5 Un-Apply a Generalization Thesaurus
   6.6 Save a Generalization Thesaurus
   6.7 Save text(s) after application of Generalization Thesaurus
   6.8 Example for building and applying a Generalization Thesaurus

2. Meta-Matrix Thesaurus
   7.1 Open a Meta-Matrix Thesaurus
   7.2 Edit a Meta-Matrix Thesaurus
   7.3 Build a Meta-Matrix Thesaurus
   7.4 Apply a Meta-Matrix Thesaurus

   7.4.1 Thesaurus content only
   7.4.1.1 Direct Adjacency
   7.4.1.2 Rhetorical Adjacency

   7.5 Un-Apply a Meta-Matrix Thesaurus
   7.6 Save a Meta-Matrix Thesaurus
   7.7 Save text(s) after application of Meta-Matrix Thesaurus
   7.8 Example for editing and applying a Meta-Matrix Thesaurus

3. Sub-Matrix Selection
   8.1 Select Sub-Matrix Categories
   8.2 Create or Modify Sub-Matrix Selection
   8.3 Apply Sub-Matrix Selection

   8.3.1 Direct Adjacency
   8.3.2 Rhetorical Adjacency

8.4 Un-Apply a Sub-Matrix Selection
8.5 Save Sub-Matrix Selection
8.6 Save text(s) after Sub-Matrix Selection
8.7 Example for Sub-Matrix Selection

Statement Formation Choices

1. Analysis Settings
   1.2 Default Analysis Settings

Output Options

1. Term Distribution Lists and Matrices
   1.1 Types and content of Term Distribution Lists and Matrices
   1.2 Example for Term Distribution List and Matrices
   1.3 Additional Output Options
   1.4 Examples for additional Output Formats
2. Save Non-Identified Concepts
3. Additional Output Formats
   3.1 DL for UCINET
   3.2 DyNetML
   3.3 Examples for Additional Output Formats
4. Network Analytic Measures

Analyses

1. Semantic Network Analysis
   1.1 Example for Map Analysis
2. Meta Matrix Text Analysis
   2.1 Example for Meta Matrix Text Analysis
3. Sub Matrix Text Analysis
   3.1 Example for Sub Matrix Text Analysis

Outputs

1. Map
   1.1 Example for Map Output
2. Statistics
   2.1 Example for Statistic Output
3. Additional Outputs

AutoMap Usage "Snapshots"

1. Split Input Text Files
2. Compare Maps
3. Merge DyNetML Files

4. Convert Network Data Formats
5. Edit Network Data
6. Visualize Maps

Clear

Exit

Further Information
How to cite AutoMap
Questions, Bugs, and Comments
References

AutoMap: An Overview

AutoMap is a software tool to analyze text using the method of *Network Text Analysis*. It performs a specific type of Network Text Analysis called *Semantic Network Analysis*. Semantic analysis extracts and analyzes links among words to model an authors "mental map" as a network of links. Additionally, Automap supports *Content Analysis*.

Coding in AutoMap is computer-assisted; the software applies a set of *coding rules* specified by the user in order to code the texts as networks of concepts. Coding texts as maps focuses the user on investigating meaning among texts by finding relationships among words and themes.

The coding rules in AutoMap involve text *pre-processing* and *statement formation*, which together form the *coding scheme*. Text pre-processing condenses data into concepts, which capture the features of the texts relevant to the user. *Statement formation rules* determine how to link concepts into statements.

Listed below are the steps a user would follow (in typical order) to use AutoMap:

1. **Pre-Process** texts.
2. Perform **Semantic Network Analysis** on texts.
3. Run **MetaMatrix Text Analysis** and **Sub Matrix Text Analysis** (Both techniques are sub-types of Map Analysis).
4. **Compare Maps** generated with AutoMap.
5. Compute **network analytic measures** per texts and words.

Network Text Analysis (NTA)

NTA theory is based on the assumption that language and knowledge can be modeled as networks of words and relations. Network Text Analysis encodes links among words to construct a network of linkages. Specifically, Network Text Analysis analyzes the *existence*, *frequencies*, and *covariance* of terms and themes, thus subsuming classical Content Analysis.

Semantic Network Analysis

In map analysis, a **concept** is a single idea, or *ideational kernel,* represented by one or more words. Concepts are equivalent to nodes in Social Network Analysis (SNA). The link between two concepts is referred to as a **statement**, which corresponds with an edge in SNA. The relation between two concepts can differ in strength, directionality, and type. The union of all statements per texts forms a semantic **map**. Maps are equivalent to networks.

Social Network Analysis (SNA)

Social Network Analysis is a scientific area focused on the study of relations, often defined as social networks. In its basic form, a social network is a network where the nodes are people and the relations (also called links or ties) are a form of connection such as friendship. Social Network Analysis takes graph theoretic ideas and applies them to the social world. The term "social network" was first coined in 1954 by J. A. Barnes (see: Class and Committees in a Norwegian Island Parish). Social network analysis is also called network analysis, structural analysis, and the study of human relations. SNA is often referred to as the science of "connecting the dots."

Today, the term Social Network Analysis (or SNA) is used to refer to the analysis of any network such that all the nodes are of one type (e.g., all people, or all roles, or all organizations), or at most two types (e.g., people and the groups they belong to). The metrics and tools in this area, since they are based on the mathematics of graph theory, are applicable regardless of the type of nodes in the network or the reason for the connections.

For most researchers, the nodes are actors. As such, a network can be a cell of terrorists, employees of global company or simply a group of friends. However, nodes are not limited to actors. A series of computers that interact with each other or a group of interconnected libraries can comprise a network also.

Where to find out more on SNA?

- *Scott, John, 2000, Social Networks, Sage (2nd edition)*
- *Wasserman, S. & K. Faust, 1994, Social Network Analysis: Methods and Applications.*

Dynamic Network Analysis

Dynamic Network Analysis (DNA) is an emergent scientific field that brings together traditional social network analysis (SNA), link analysis (LA) and multi-agent systems (MAS). There are two aspects of this field. The first is the statistical analysis of DNA data. The second is the utilization of simulation to address issues of network dynamics. DNA networks vary from traditional social networks in that are larger dynamic multi-mode, multi-plex networks, and may contain varying levels of uncertainty.

DNA statistical tools are generally optimized for large-scale networks and admit the analysis of multiple networks simultaneously in which, there are multiple types of nodes (multi-node) and multiple types of links (multi-plex).In contrast, SNA statistical tools focus on single or at most two mode data and facilitate the analysis of only one type of link at a time.

DNA statistical tools tend to provide more measures to the user, because they have measures that use data drawn from multiple networks simultaneously. From a computer simulation perspective, nodes in DNA are like atoms in quantum theory, nodes can be, though need not be, treated as probabilistic. Whereas nodes in a traditional SNA model are static, nodes in a DNA model have the ability to learn. Properties change over time; nodes can adapt: A company's employees can learn new skills and increase their value to the network; Or, kill one terrorist and three more are forced to improvise. Change propagates from one node to the next and so on. DNA adds the critical element of a network's evolution and considers the circumstances under which change is likely to occur.

Illustrative problems that people in the DNA area work on -

1. Developing metrics and statistics to assess and identify change within and across networks.
2. Developing and validating simulations to study network change, evolution, adaptation, decay...
3. Developing and validating formal models of network generation and evolution.
4. Developing and testing theory of network change, evolution, adaptation, decay...
5. Developing techniques to visualize network change overall or at the node or group level.
6. Developing statistical techniques to see whether differences observed over time in networks are due to simply different samples from a distribution of links and nodes or changes over time in the underlying distribution of links and nodes.
7. Developing control processes for networks over time.
8. Developing algorithms to change distributions of links in networks over time.

9. Developing algorithms to track groups in networks over time.
10. Developing tools to extract or locate networks from various data sources such as texts.
11. Developing statistically valid measurements on networks over time.
12. Examining the robustness of network metrics under various types of missing data.
13. Empirical studies of multi-mode multi-link multi-time period networks.
14. Examining networks as probabilistic time-variant phenomena.
15. Forecasting change in existing networks Identifying trails through time given a sequence of networks. Identifying changes in node criticality given a sequence of networks anything else related to multi-mode multi-link multi-time period networks.

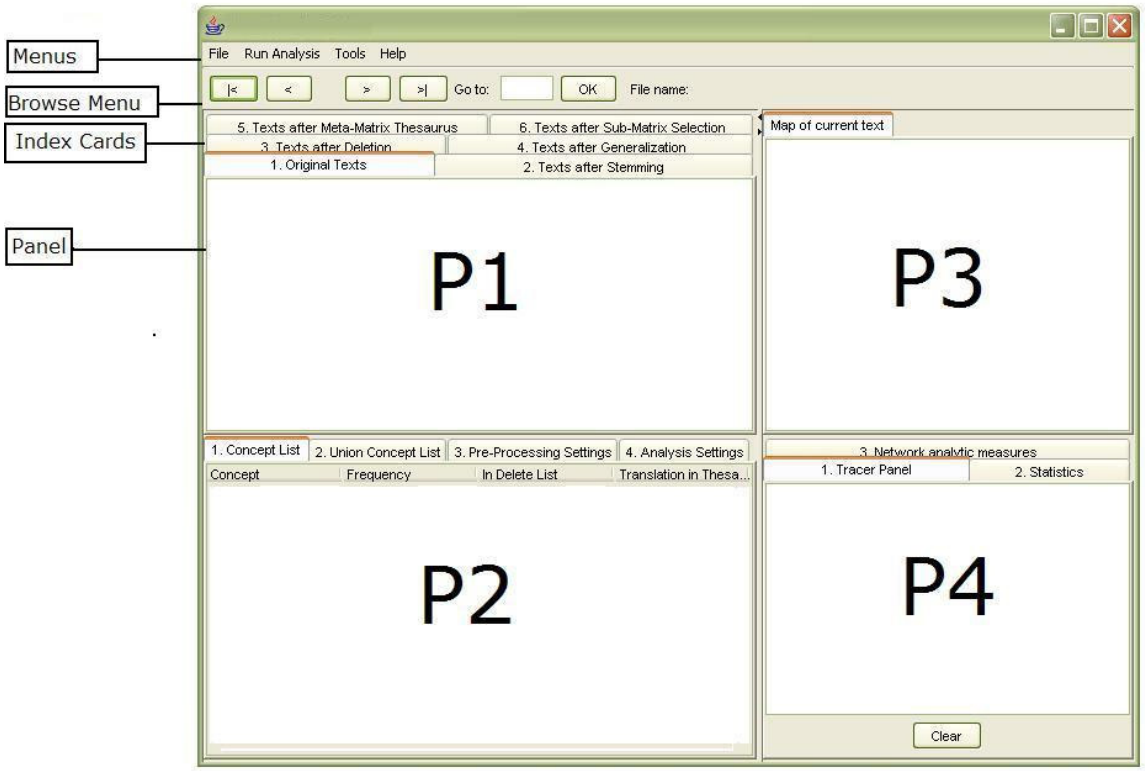
Automap Graphical User Interface

AutoMap's graphical user interface (GUI) is divided into four primary quadrants (or panels): they will be referred to as **P1** (top left)**, P2** (bottom left)**, P3** (top right) and **P4** (bottom right).

The drop-down menu bar provides access to various analysis tools and utilities. The browse menu bar allows you to quickly navigate between loaded texts.

Index cards, or "tabs," provide a tabular interface allowing you to navigate each panel respectively. The GUI reads any changes dynamically from the XML file. To do that, the user needs to refresh the tool.

The screen shot below highlights primary features of the Automap GUI and where to find them:



It is important to note that **P2** *can* be edited. The other panels (**P1, P3, P4**) *cannot* be edited. Information displayed on **P1 to P4 always relates to each other**. The Text Browse Menu relates to all panels at the same time.

Window sizes do not have an upper threshold. AutoMap will automatically set window size to largest text size upon user's request. This is a new button on the Analysis Settings panel. This enables text set specific maximum window sizes, which also enhances efficiency.

The **Action Tracer Panel** in P4 will log preprocessing utilities applied to your text. This is a handy way to keep track of changes and actions relating to your text.

Tool Tips

In the various pre-processing panels, such as utilities, "tool tips," provide more information on certain routines. Tool tips become visible when you slide he mouse over that particular tool.

Examples

This user's guide provides illustrative examples for all AutoMap functions. The sample texts below will be used throughout this guide.

Tip! it is suggested you follow along using the examples below. To do so, simply copy and paste the texts below into wordpad or notepad and save as a .txt file in the same folder.

Our Text Example

```
Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs.
Brown was unhappy with the dog. She yelled at it saying
"You impossible dog!" But the dog kept eating the flowers
and weeds. She asked Mr. Cray to stop the dog. He
couldn't.

Mrs. Brown planted roses and weeded the garden. The silly
dog % dug up the roses looking for a vole on June 12,
1880. Weeding was no longer needed.

Prof. Darren, Mrs. Brown & Mr. Cray met the next day to
concoct a plan. John Darren and Mrs. Brown put up a
scarecrow. She thought it would scare the dog. Mr. Cray
put up a fence. Problem solved. Then Mrs. Brown planted
lotus, carnations, daffodils, and roses.
```

Our Text Example II

```
Mr. Cray's brown dog stopped eating the lotus blossom at
12 pm the next day. Mrs. Brown was now happy with the
dog. She said "You good dog!" The dog no longer ate the
flowers and weeds. Mr. Cray was pleased too.

Mrs. Brown watered the roses and fertilized the garden on
June 13, 1880.

Prof. Darren, Mrs. Brown & Mr. Cray met over dinner and
discussed how the plan had worked. John Darren and Mrs.
Brown would take down the scarecrow the following week.
She thought it was too scary for the dog. Mr. Cray
painted his fence. Then Mrs. Brown watered lotus,
carnations, daffodils, and roses.
```

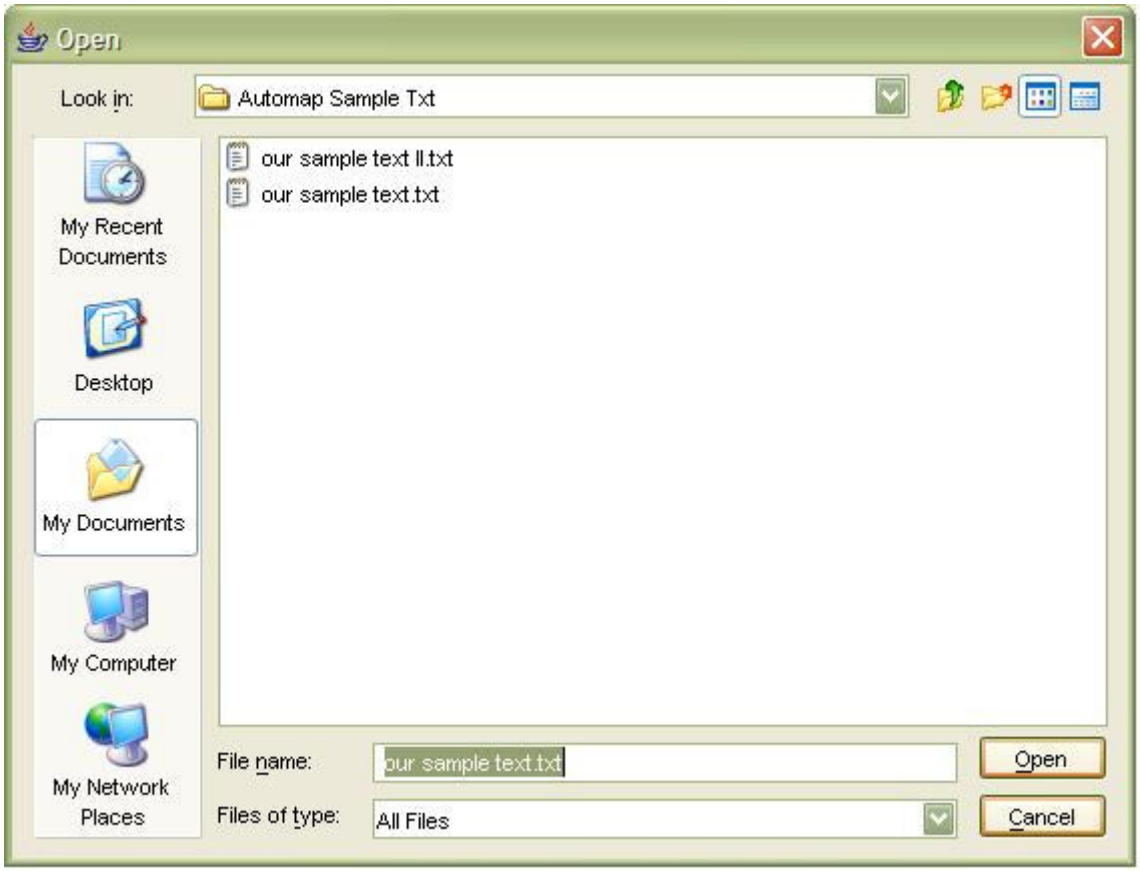
Load Input

1. To open a single text

To open a single text file into Automap, proceed as follows from the Automap menu bar:

File Open > Open single file

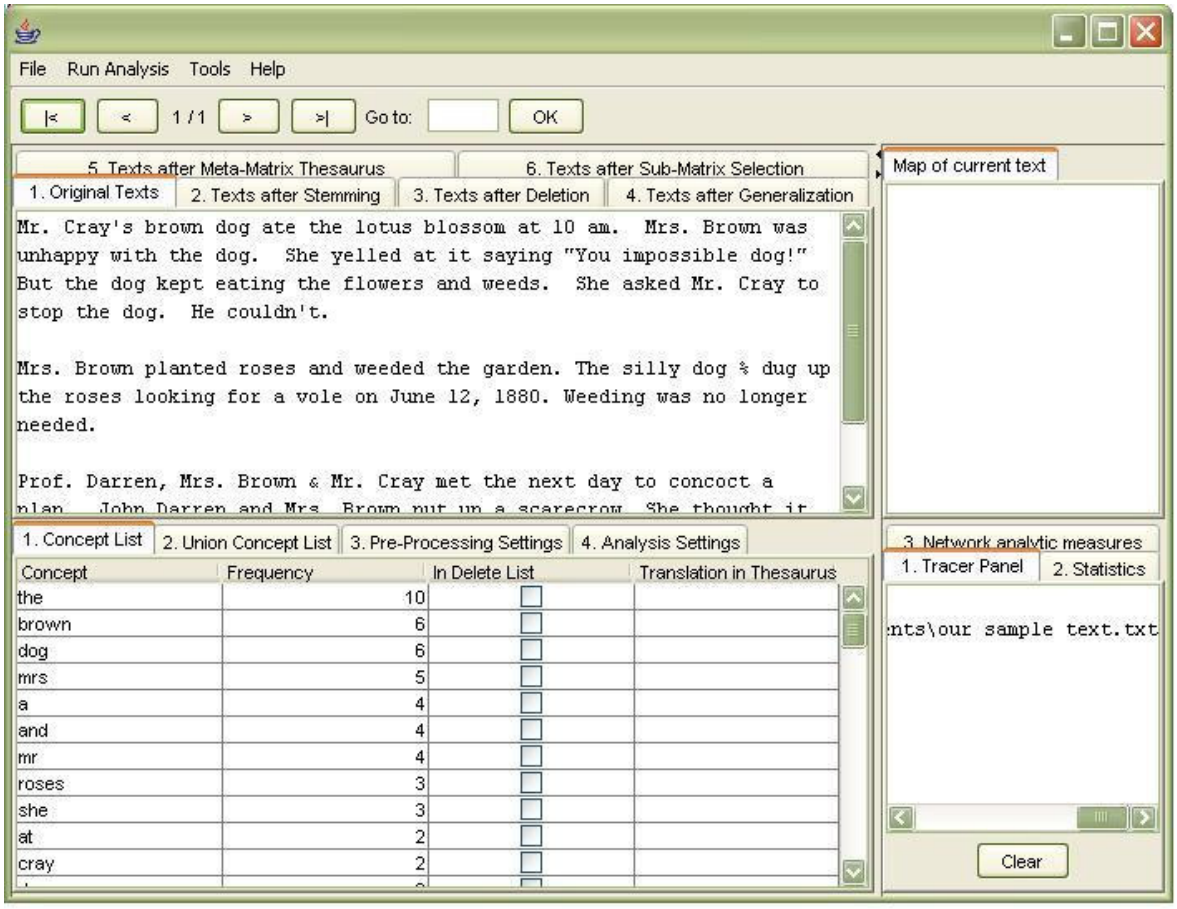
A file chooser will pop up (screen shot below).



Double click on the file that you wish to analyze > Select the Open button.

The text will be displayed in **P1** on tab **No. 1. Original Text**.

The loading of .TXT files (caps) is now enabled.

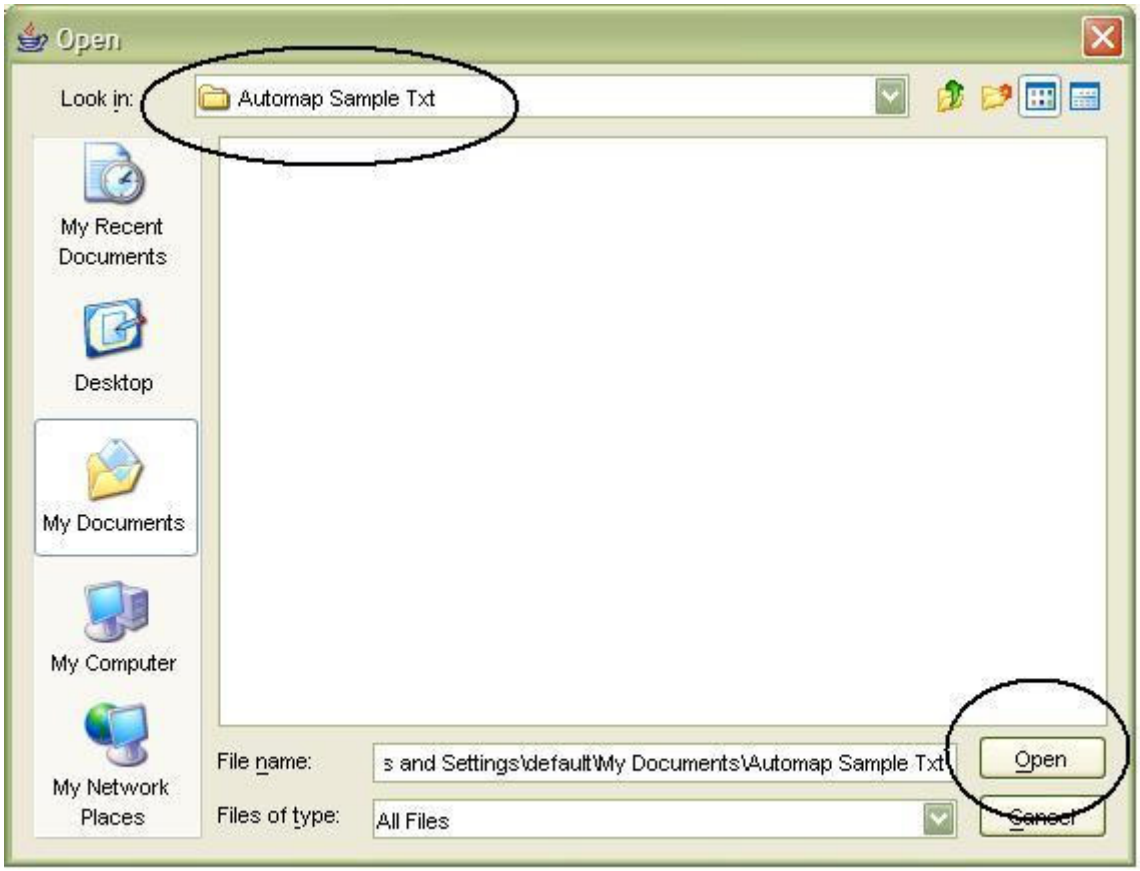


2. To open a set of texts

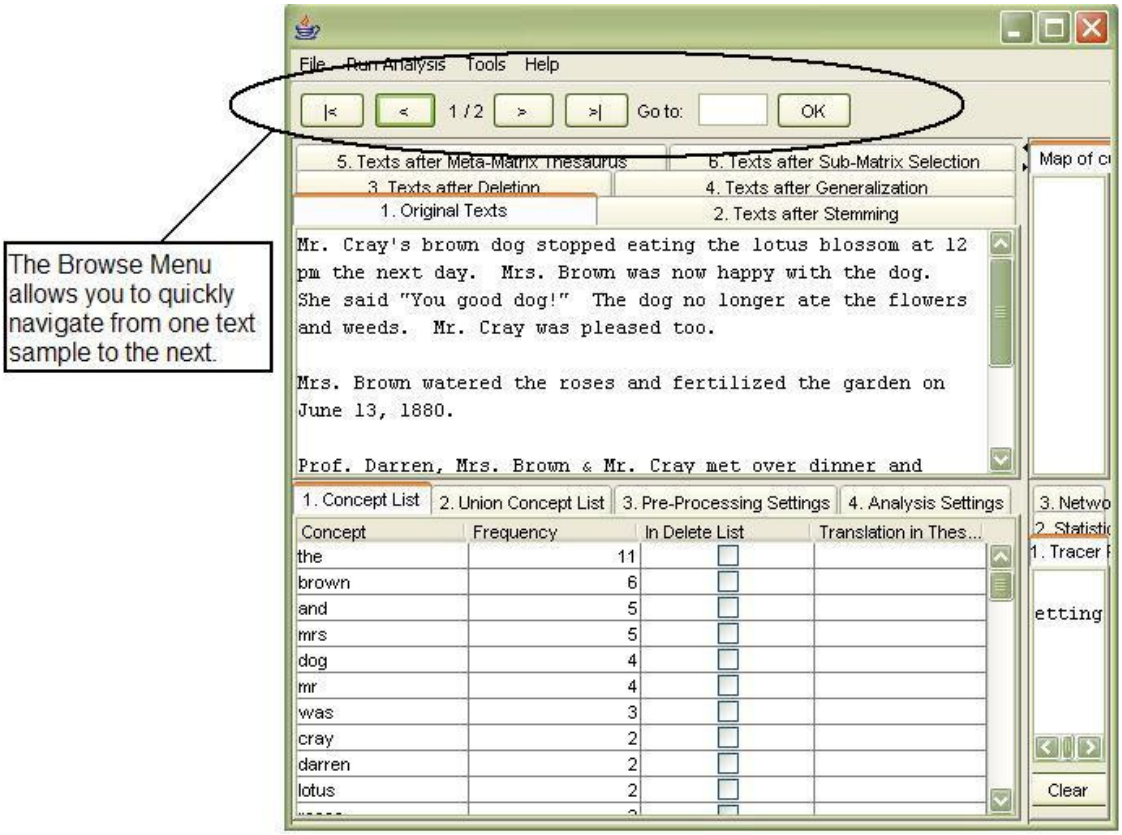
Should you wish to analyze multiple texts at the same time, they must be stored in one folder. To do so, proceed as follows from the Automap menu bar:

File menu > **Open multiple files**

After you select "Open multiple files" a folder chooser will pop up. Again, be sure that correct folder is selected in the folder chooser. The black ellipses in the screen shot below highlight where you should be looking in the file chooser:



Choose the folder that contains the texts you wish to analyze and single click on it. The folder will be highlighted. Do not double click on the folder. Select the **Open** button (see above screen shot). The first text will be displayed in panel **P1**, tab **no. 1. Original Text** . You can browse through the texts by using the Browse Menu**.**



Text Analysis Utilities

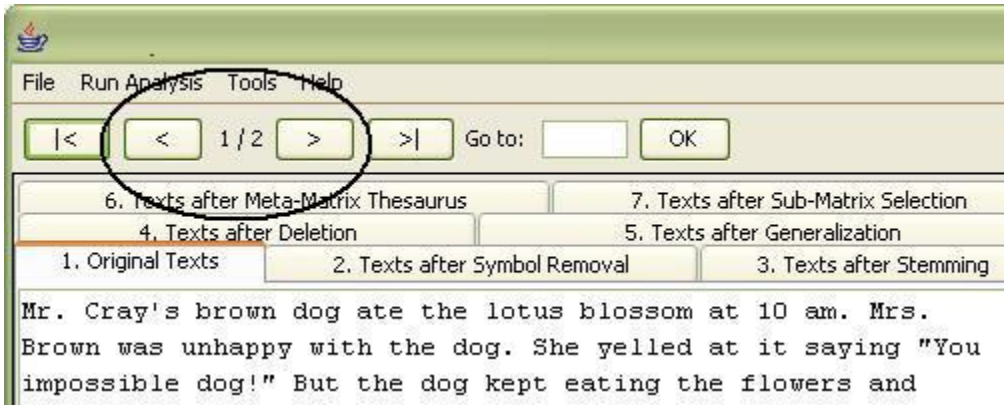
Automap contains text analysis utilities to help you in the pre-processing and data analysis of your text examples. Take time to become familiar with them as they allow you to quickly work with your text examples.

1. Browse through texts
This function enables you to quickly jump from text example to another. All panels are synchronized in the Browse Menu. A series of screen shots below the instruction ad emphasis on how to access text Automap's text browsing features.

How to use the **Browse Menu**:

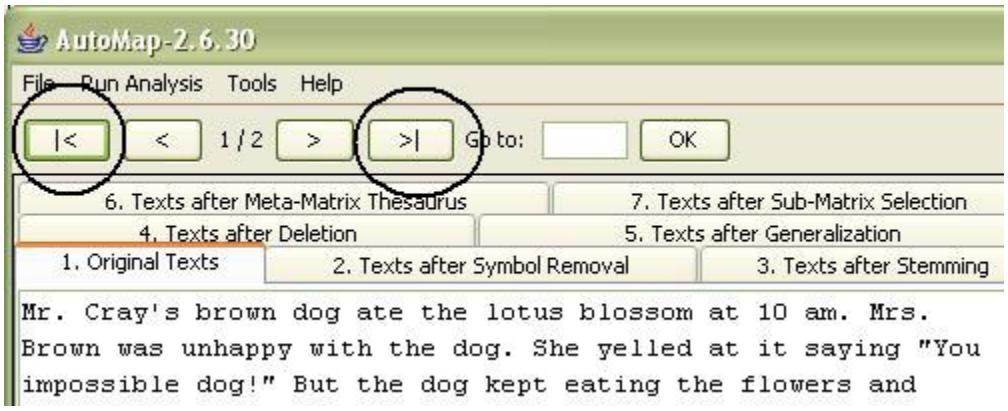
To go backward or forward text by text:

Click the ">" button or the "<" button in the browse menu bar.



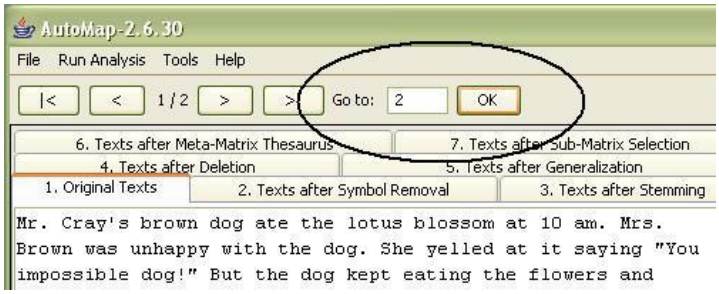
To go to first or last text in the text set:

Click the ">>" button or the" <<" button.



To go to a specific text:

Enter the text number in text field right next to the Go to command and hit OK.



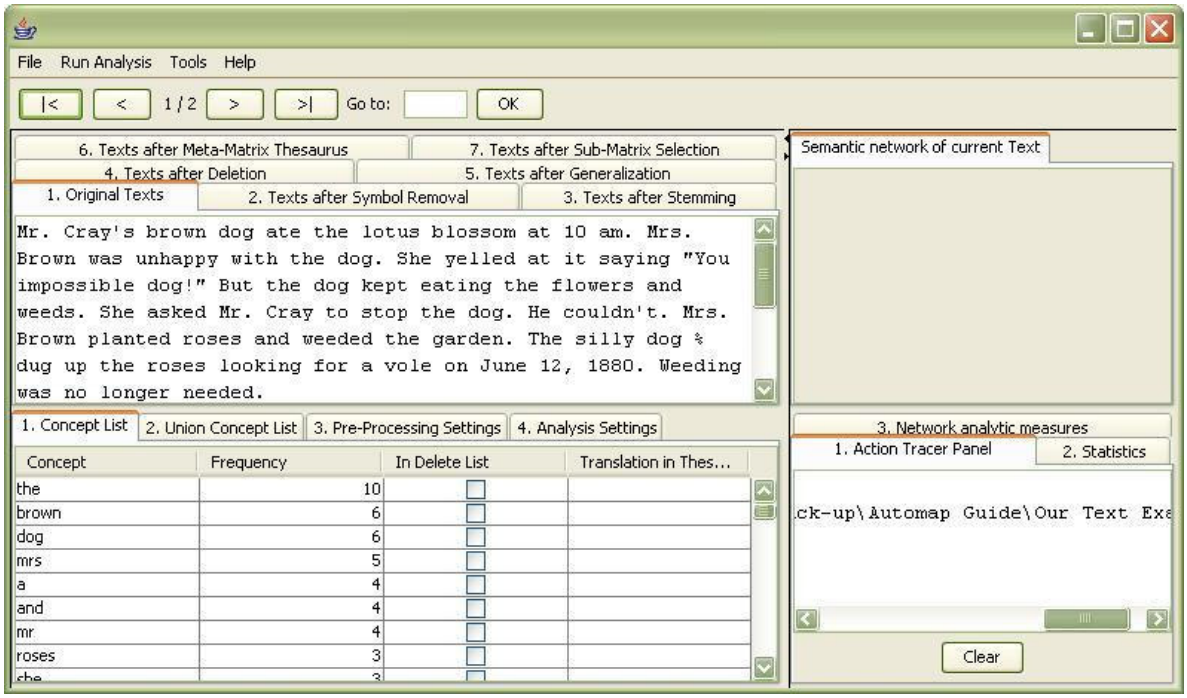
The name of the currently selected text is displayed on the Browse menu.

These files can be browsed:

- All input texts (**P1, No. 1. Input Text** index card).

- All texts after each stage of pre-processing (all index cards of **P1**).
- Concept lists per text (original input text or texts after each stage of pre-processing).
- Map (**P3**) and Statistics (**P4**) outputs that relate to the text currently displayed on the index cards of **P1**.

1.1 Example for Browse Menu



The example shows a part of the *Our Text 1.txt* in panel **P1** tab **no. 1 Original Texts**.

The browse menu tells you several important facts:

- The data set opened contains two texts.
- The first of the loaded texts is displayed.
- The filename of the displayed text.

2. Concept List

The Concept List is displayed in panel **P2** tab **No. 1 Concept List**. The Concept List is created automatically once a text or a set of texts are loaded or modified in Automap.

The concept list tells you several important facts about your text:

- The number of concepts found in the text displayed in **P1** tab **No. 1 Original Text.**

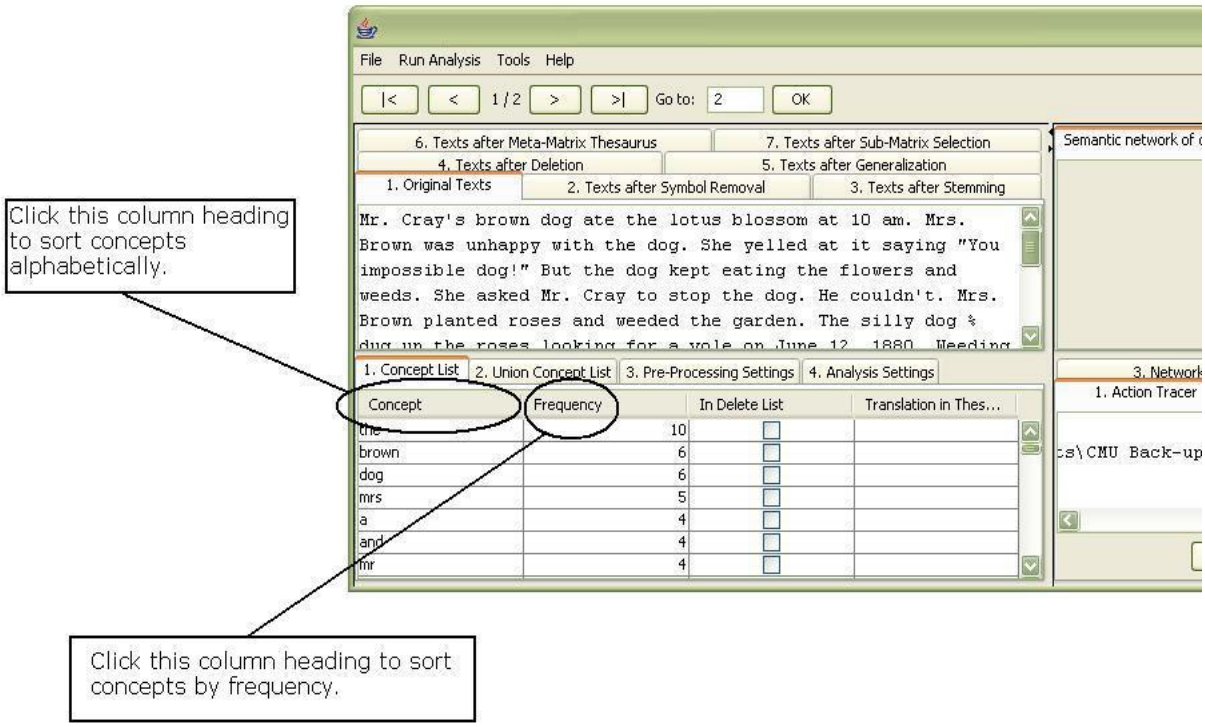
- Related frequencies of those concepts.
- Cumulated unique concepts and total concepts contained in the data set.

Tip! The number of unique concepts considers each concept only once, whereas the number of total concepts considers repetitions of concepts.

By default, the Concept List is sorted by decreasing frequency of concepts. To sort the list alphabetically, click on the first-column header **Concept**. In order to resort the list, click on the header of the second-column header **Frequency.**

2.1 Example for Concept List

The example below shows a part of the Concept List for the text displayed in the browse menu. The Table is ordered by Frequency. The concept list contains more entries than the interface can display:



3. Create and refresh Union Concept List

The **Union Concept List**, found in panel **P2**, differs from the *Concept List* (tab no. 1) in that it considers concepts across *all* texts loaded in Automap, rather than one single text file. There are several key pieces of information the Union Concept List tells you:

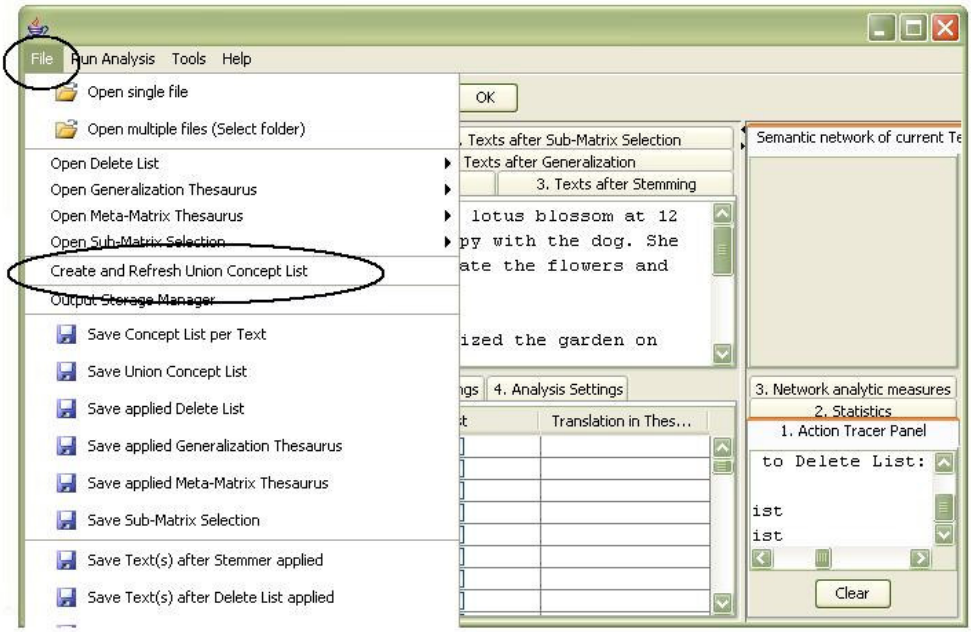
- Concepts contained in all loaded text sets.
- Related, cumulative frequencies of concepts in all text sets.
- Cumulative unique and total concepts.

Union Concept List results are displayed on tab **No. 2 Union Concept List** in **P2**. However, you must first refresh the union concept list from the file menu, before viewing your results on the **o. 2 Union Concept** tab. The union concept list can be refreshed after each step of pre-processing in order to visualize the impact of pre-processing operations on the union of concepts.

To refresh the Union Concept, from the drop-down menu bar:

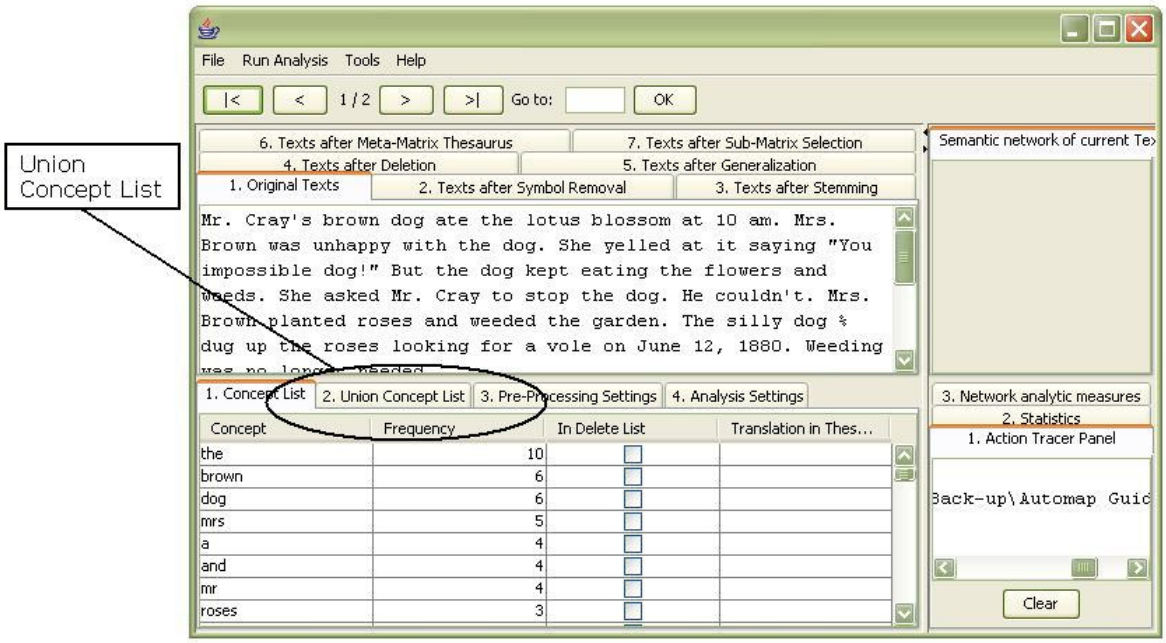
File menu > Refresh Union Concept List.

The ellipse in the screen shot below shows were to access the **Refresh Union Concept List** from the drop-down menu bar:

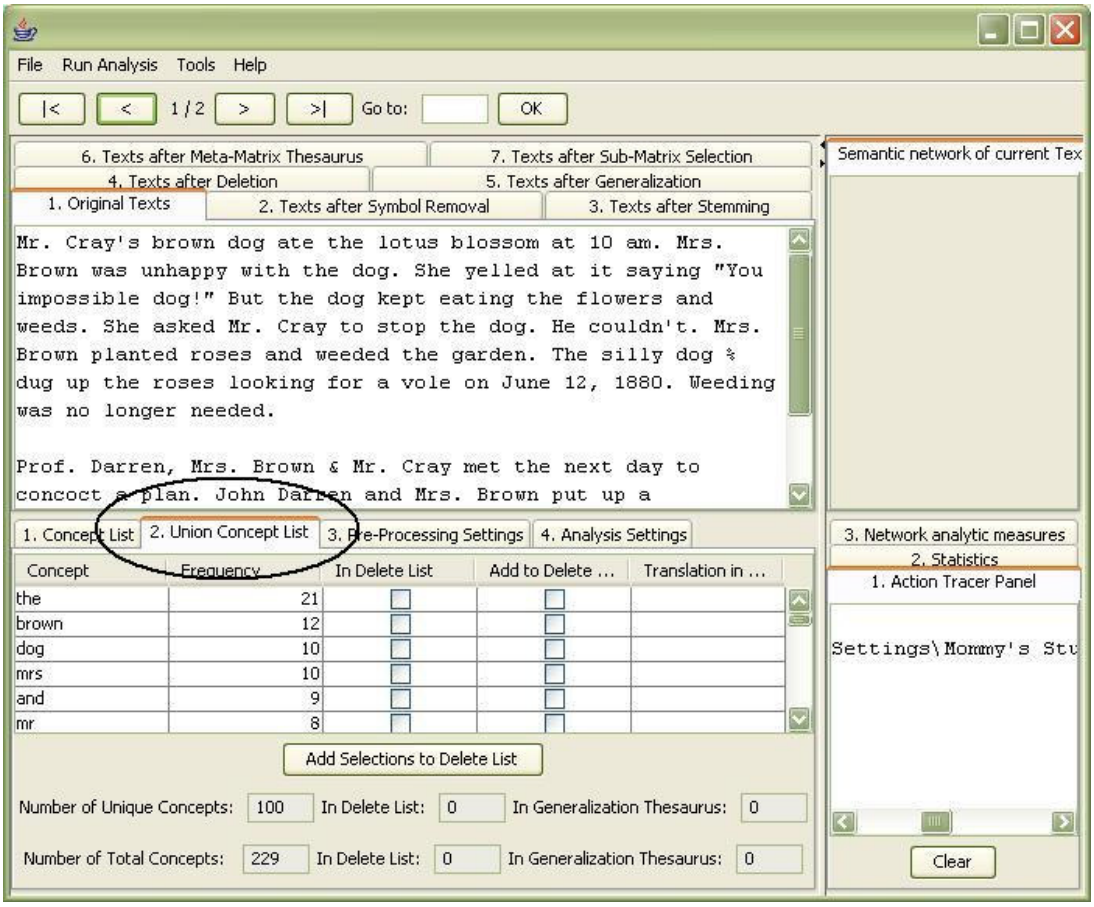


The call out box in the screen shot below, shows where to locate the **Union Concept** tab, which will contain the results of the Refresh Union Concept List

analysis:



Your results will be displayed in tab **No 2. Union Concept List**. The black ellipse in the screen shot below highlights where to find this tab in the Automap GUI.



By default, the list is sorted by decreasing frequency of concepts. In order to sort the list alphabetically click on the first-column header **Concept**. To re-sort the list, click on the second-header column **Frequency.**

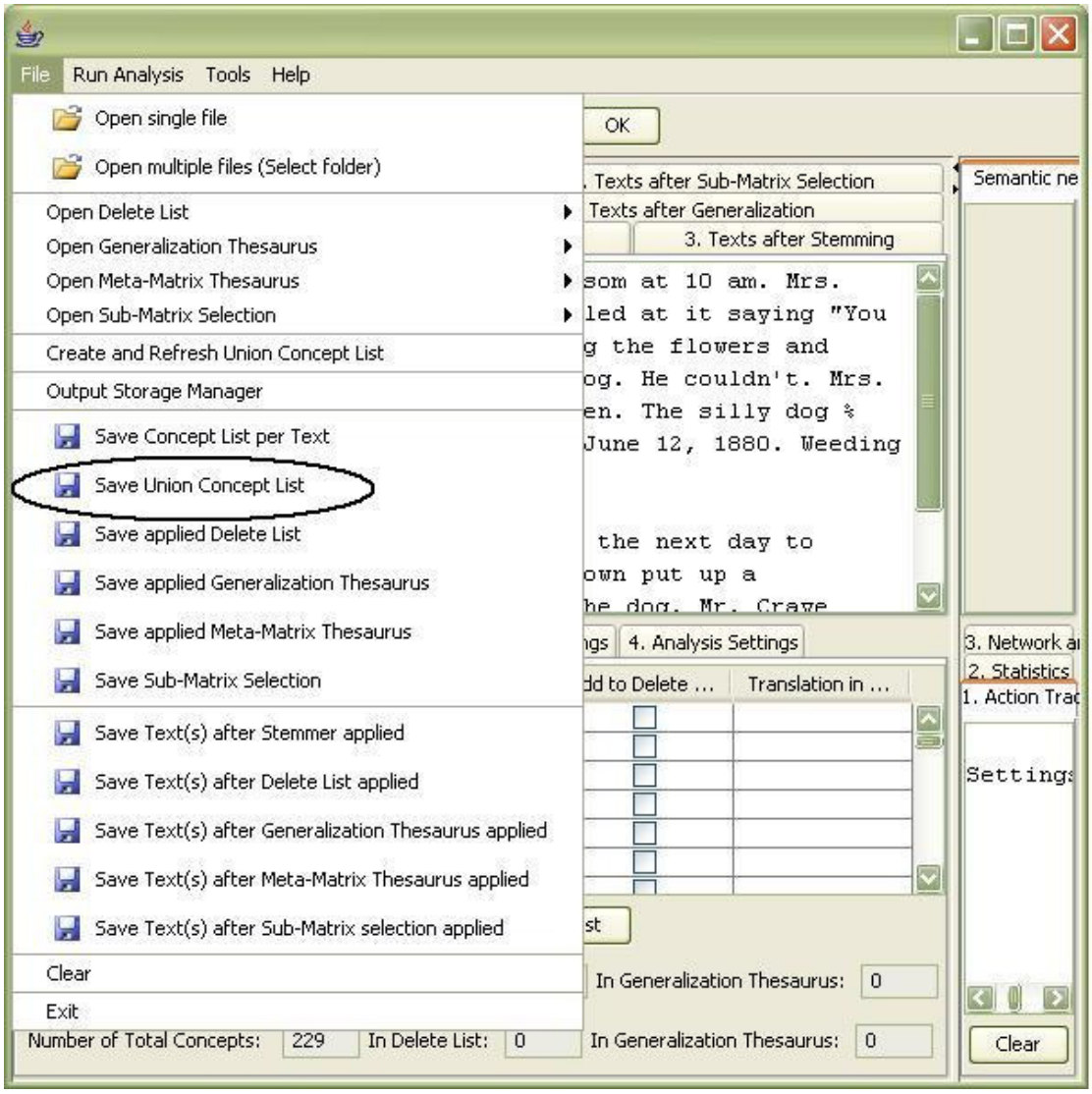
Note: The number of unique concepts considers each concept only once, whereas the number of total concepts also considers repetitions of concepts.

3.1 Save Union Concept List

To save a Union Concept List follow these steps:

File menu > Save Union Concept List.

A file chooser will pop up. The black ellipse below highlights how to save a Union Concept List from the drop-down menu bar.



3.2 Example for Union Concept List

Let us walk through an example of creating a Union Concept List working with our text examples from above. They are restated below for you convenience:

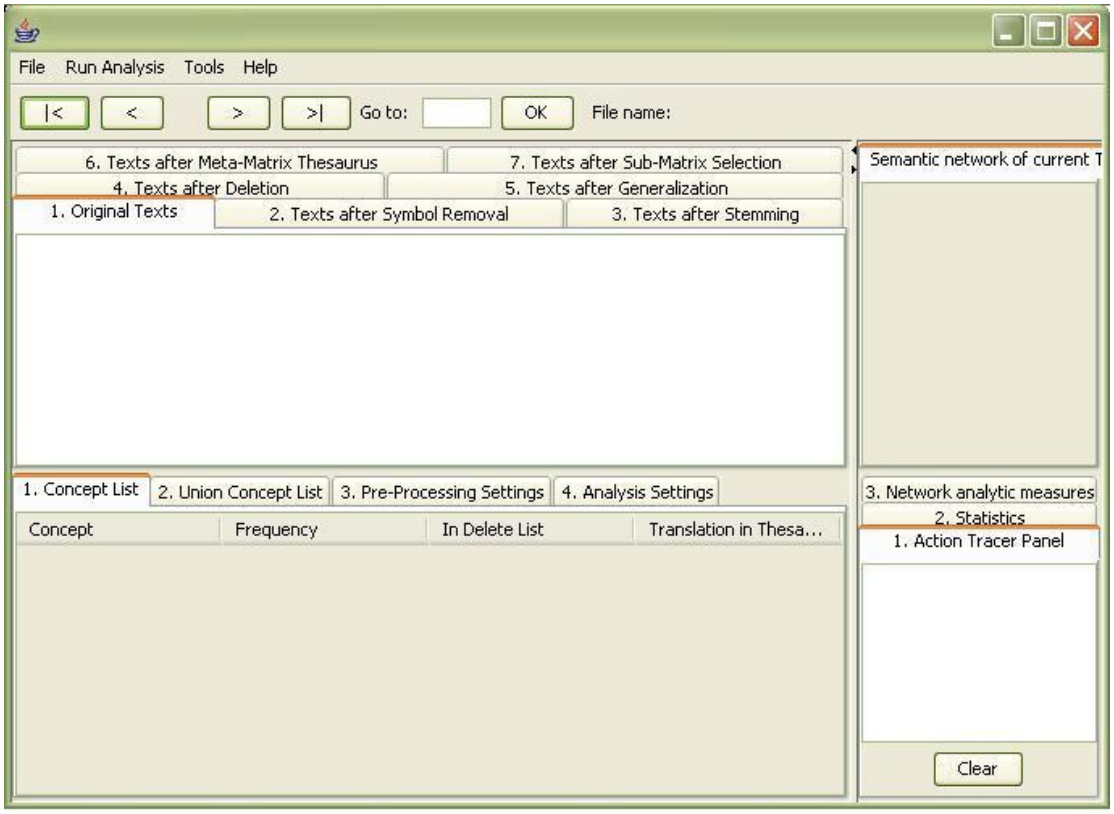
```
1) Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was
unhappy with the dog. She yelled at it saying "You impossible dog!" But
the dog kept eating the flowers and weeds. She asked Mr. Cray to stop
the dog. He couldn't. Mrs. Brown planted roses and weeded the garden.
The silly dog % dug up the roses looking for a vole on June 12, 1880.
Weeding was no longer needed.

Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan.
John Darren and Mrs. Brown put up a scarecrow. She thought it would
scare the dog. Mr. Craye putup a fence. Problem solved. Then Mrs. Brown
planted lotus, carnations, daffodils, and roses.
```

2) Mr. Cray's brown dog stopped eating the lotus blossom at 12 pm the next day. Mrs. Brown was now happy with the dog. She said "You good dog!" The dog no longer ate the flowers and weeds. Mr. Cray was pleased too.

Mrs. Brown watered the roses and fertilized the garden on June 13, 1880.

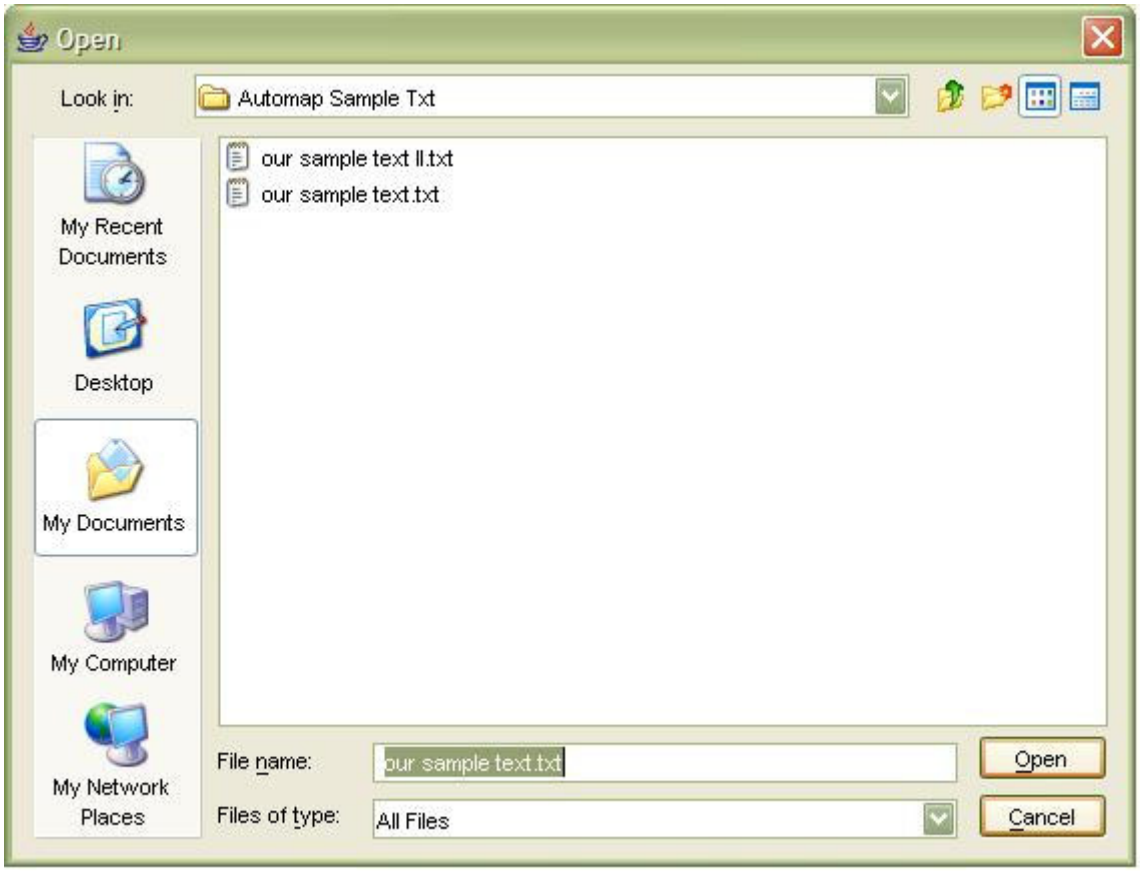
Prof. Darren, Mrs. Brown & Mr. Cray met over dinner and discussed how the plan had worked. John Darren and Mrs. Brown would take down the scarecrow the following week. She thought it was too scarry for the dog. Mr. Craye painted his fence. Then Mrs. Brown watered lotus, carnations, daffodils, and roses.

The first step is to load in your multiple texts from which we will create and save a Union Concept List. An empty AutoMap GUI is shown below before any text is loaded into it. This is what you will see when you first run AutoMap:



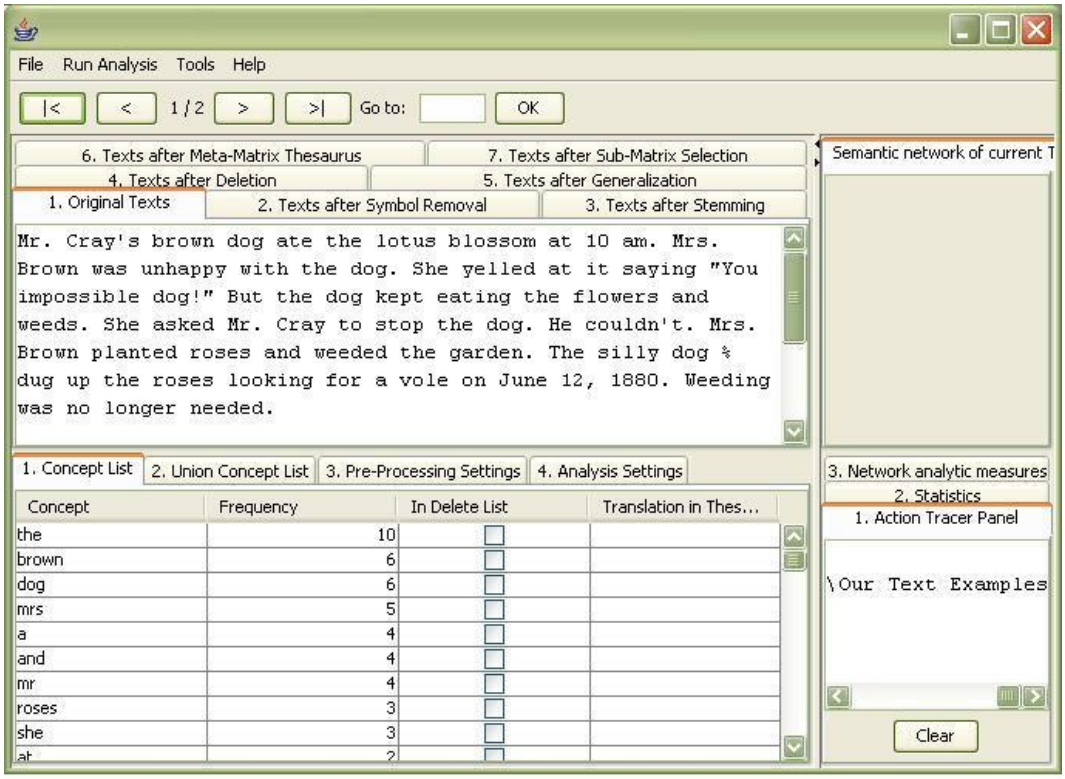
From the drop-down menu:

File > Open Multiple Files > (select location of folder on applicable drive containing your text examples)



Select "Open"

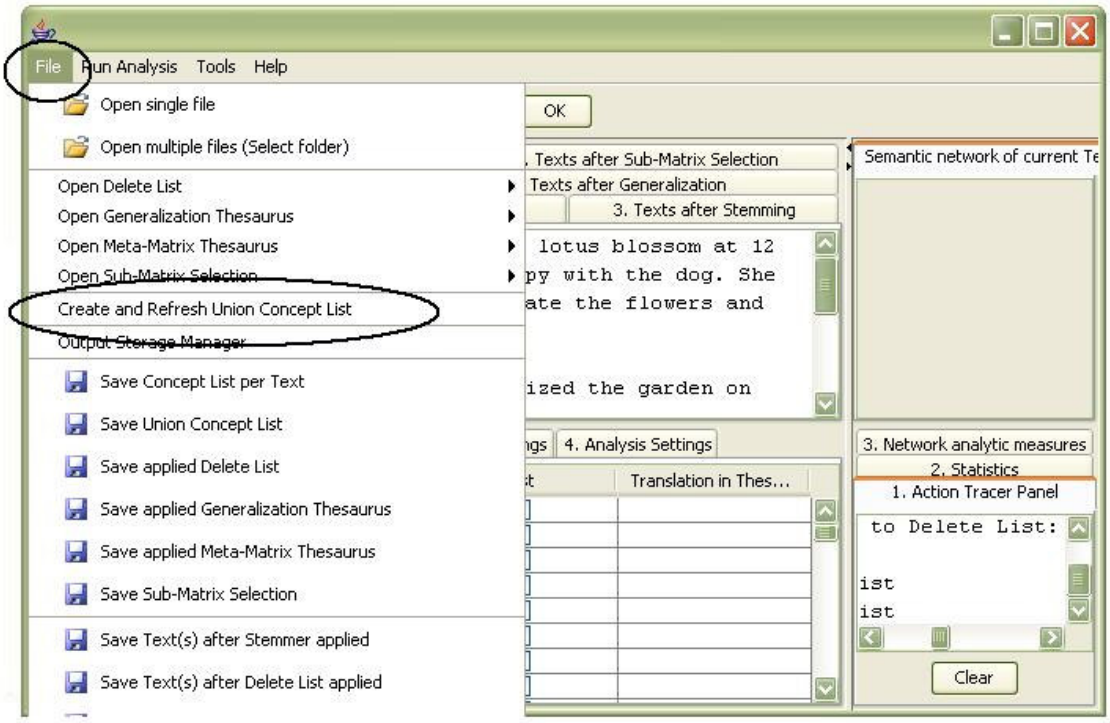
The screen shot below displays the Automap GUI after our text examples have been loaded into it.



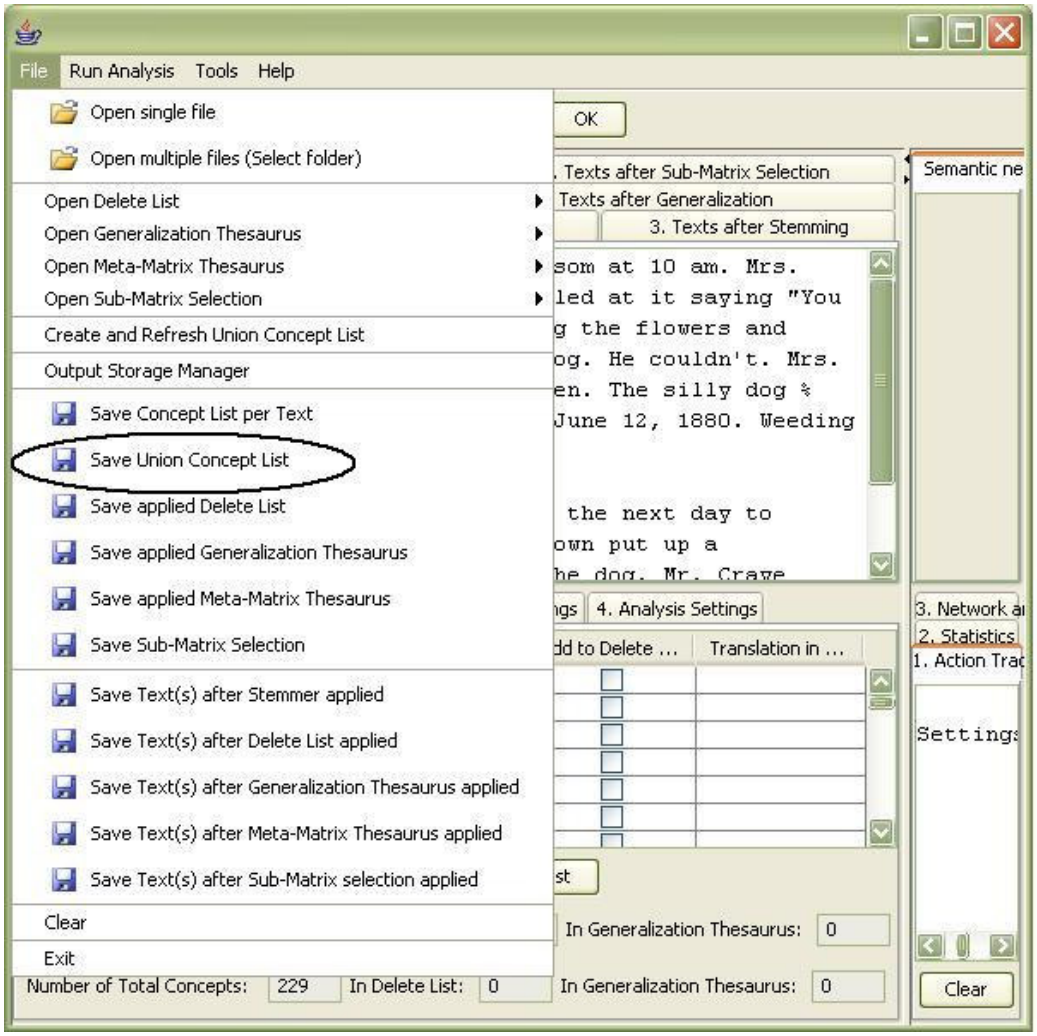
To run the Create and Refresh Union Concept Utility based on multiple texts:

File > Create and Refresh Union Concept List

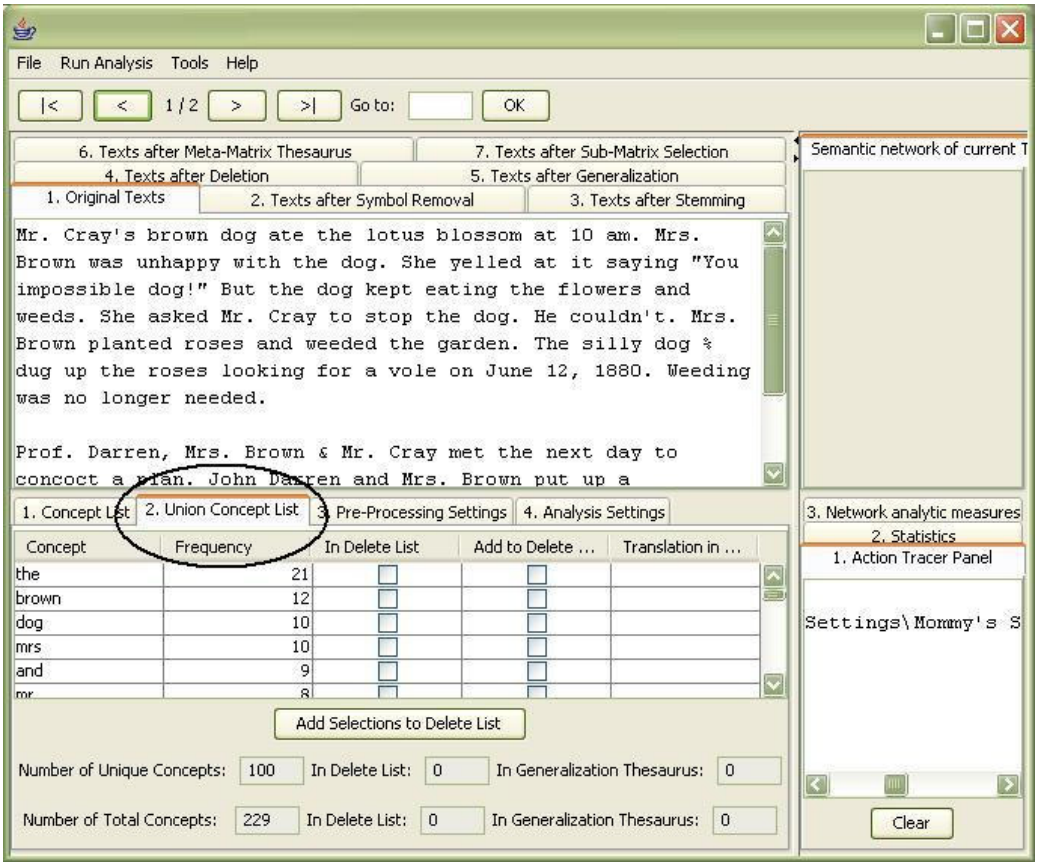
The following series of screen shots present a step-by-step on how to create an refresh a union concept list:



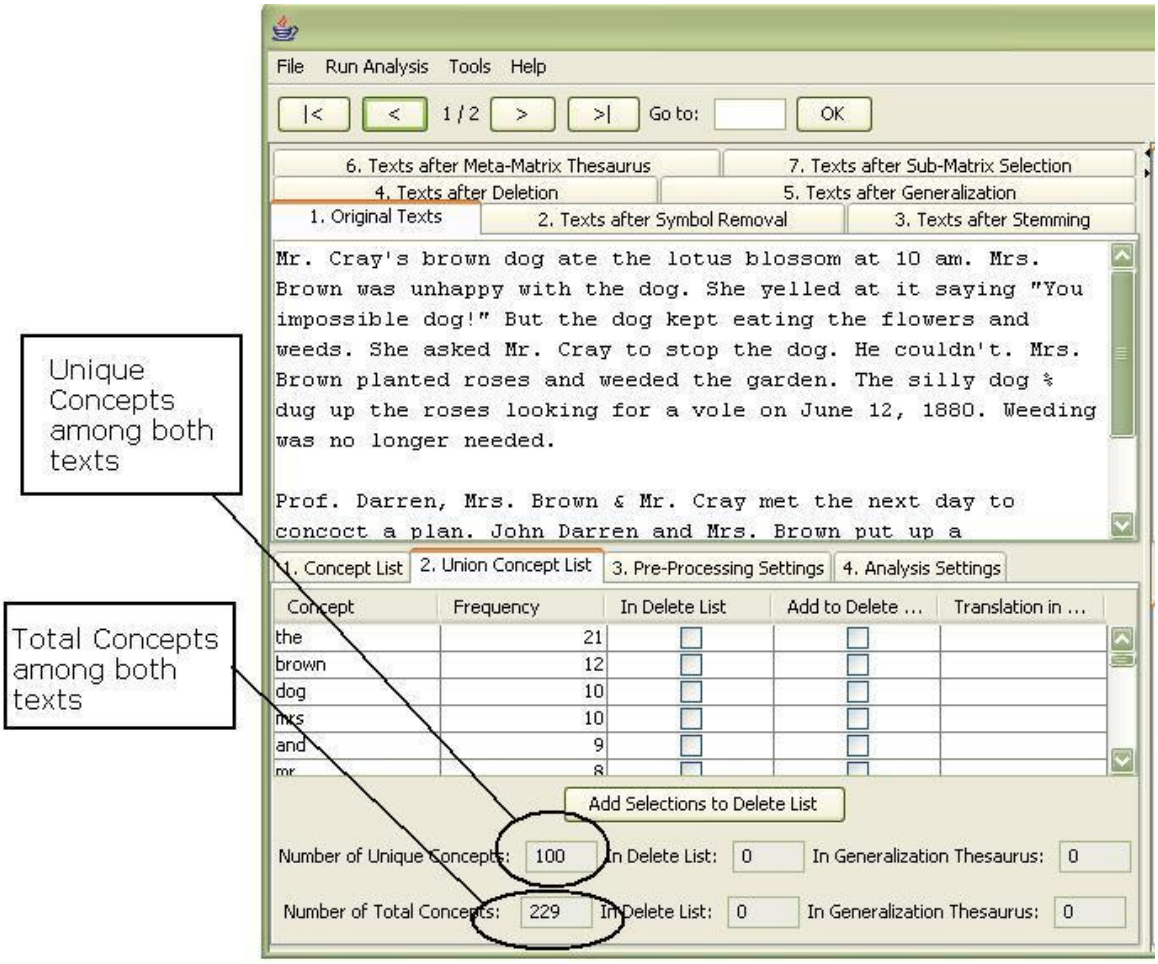
Select "Save Union Concept List" from the drop-down menu:



The result is shown in the screen shot below:



The table is ordered by *Frequency*. The concept list contains more entries than the interface can display.
The displayed Union Concept List indicates there are 100 unique concepts and 229 total concepts in the data set.



Text Pre-Processing

1. Introduction to Text Pre-Processing in AutoMap

Pre-processing reduces the data to terms relevant to you.

Tip! All pre-processing techniques in AutoMap are optional.

There are some points to consider before you begin Pre-Processing:

1. Namely, Map Analysis can be run without any prior data pre-processing.
2. Meta Matrix Text Analysis and Sub Matrix Text Analysis *require* pre-processing.

Pre-processing is *semi-automated* and *iterative* and involves several key processes:

- **Named-Entity Recognition**

Named-Entity Recognition is an Automap feature that allows you to retrieve proper names (e.g. names of people, organizations, places), numerals, and abbreviations from texts (Magnini, Negri, Prevete & Tanev, 2002). The AutoMap Named-Entity Recognition functionality detects:

- Single words that are capitalized.
  - Example: Copenhagen.
- Adjacent words that are capitalized.
  - Example: The New York City Police Department.
- A string of adjacent words that are capitalized, but can be intervened by one non-capitalized word. The first and the last word in this string are capitalized.
  - Example: Canadian Department of National Defense.

- **Stemming**

  Stemming detects inflections and derivations of concepts in order to convert each concept into the related morpheme (Jurafsky & Martin, p.83, 654). AutoMap offers 2 stemmers:

  - The **Porter Stemmer** (Porter, 1980). This stemmer uses the Porter Stemming algorithm. Additionally, it converts irregular verbs into the verb's infinitive.
    - Example: The phrase "Hospitals switched to using emergency generators," will be stemmed to "Hospit switch to be us emerg gener."
  - KSTEM or KROVETZ stemmer (Krovetz, 1995). An inflection and derivation-based stemmer. The KSTEM or KROVETZ stemmer can be customized.
- **Collocation Identification**

  A word's collocates are words appearing next to or near to it.

Tip! Collocations occuring with high frequency are powerful indicators of a pattern of meaning in a text.

Collocations are helpful to construct *thesauri* in AutoMap. AutoMap can identify collocations of size 2 (Bigrams) as shown below:

- mobile phones
- shopping centre

- **Deletion**

  Deletion removes non-content bearing conjunctions and articles from texts (Carley, 1993). Non-content bearing concepts to be deleted from the texts are denoted in a Delete List. When applying a Delete List, AutoMap searches the text(s) for concepts specified in the Delete List and delete matches from the text(s). Example:

- Original input text: The New York City Police Department said a number of people were trapped in elevators for awhile.
- Entries in the Delete List: the, a, of, were, in, for, awhile.
- Text after deletion: New York City Police Department said number people trapped elevators.

- **Thesauri**

A thesaurus associates concepts with more abstract concepts. When applying a thesaurus, AutoMap searches the text set for the *text-level concepts* denoted in the thesaurus and translates matches into the corresponding concept. The terminology of a thesaurus depends on the content and the subject of the data set (Burkart, 1997: 163; Zuell & Alexa, 2001: 313).

Generalization Thesaurus.

A generalization thesaurus typically is a two-columned collection that associates text-level concepts with higher-level concepts. The text-level concepts represent the content of a data set, and the higher-level concepts represent the text-level concepts in a generalized way (Burkart, 1997; Klein 1997: 256; Popping & Roberts 1997: 382).

- Example: Copenhagen will be associated with the higher-level concept City.
- Related type of Analysis: Map Analysis.

Meta-Matrix Thesaurus

A Meta-Matrix Thesaurus associates text-level concepts with meta-matrix categories. Since one concept might need to be translated into several meta-matrix categories, a meta-matrix thesaurus can consist of more than two columns. For example, the concept “commander” corresponds with the categories agent and knowledge.

- Example: City will be associated with and translated into Location.
- Related type of Analysis: Meta-Matrix Analysis.

For the *meta-matrix thesaurus*, column headers start with concept knowledge. The order AND naming of column headers of the *meta-matrix thesaurus* can be changed in the XML file.

Sub-Matrix Selection. The Sub-Matrix Selection denotes which Meta-Matrix Categories should be retranslated into concepts used as input for the meta-matrix thesaurus.

- Example: Location will be selected and translated into Copenhagen, Oskarshamn and Ringhals, if those concepts were associated with the Meta-Matrix category "Location" in the Meta-Matrix Thesaurus.
- Related type of Analysis: Sub-Matrix Analysis.

2. Hierarchy of Pre-Processing Techniques

If you apply a pre-processing technique of a lower order prior to a technique of higher order, the pre-processing will be maintained through all following procedures of higher order. You can un-apply each technique after applying it, if needed.

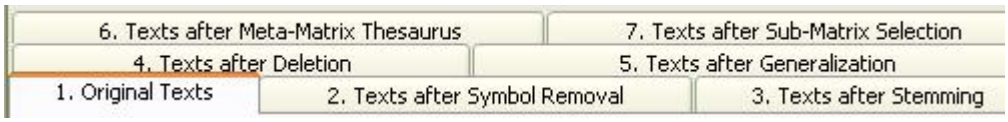
Tip! All pre-processing techniques are optional.

If you wish to apply multiple pre-processing techniques, do this in the following order:

1. Named-Entity Recognition: This is an utility that does not impact the data. Can be used before any type of analysis is run. Can be used before or after Stemming.
2. Collocation/ bigram Identification: This is an utility that does not impact the data. Can be used before any type of analysis is run.
3. Stemming: Can be used before any type of analysis is run. Can be used before or after Named-Entity Recognition.
4. Deletion: Can be used before any type of analysis is run.
5. Thesauri:
    1. Generalization Thesaurus
       Can be applied before Semantic Network Analysis is run. Can be applied before Meta-Matrix Thesaurus is applied.
    2. Meta-Matrix Thesaurus
       Has to be applied if Meta-Matrix Analysis should be run.
    3. Sub-Matrix Selection
       Can only be performed if Meta-Matrix Thesaurus was applied. Has to be applied if Sub-Matrix Analysis should be run.

The numbering of the index card tabs on **P1** and **P2** reflect this hierarchy in order to make the sequence of the pre-processing steps more intuitive.

Numbering of index card tabs on P1:



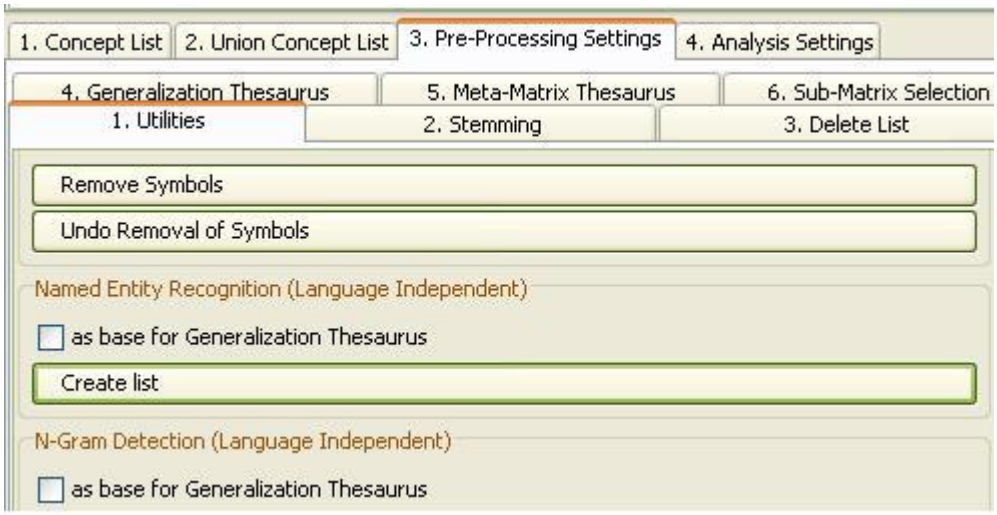
Numbering of index card tabs on P2:

3. NLP Utilities

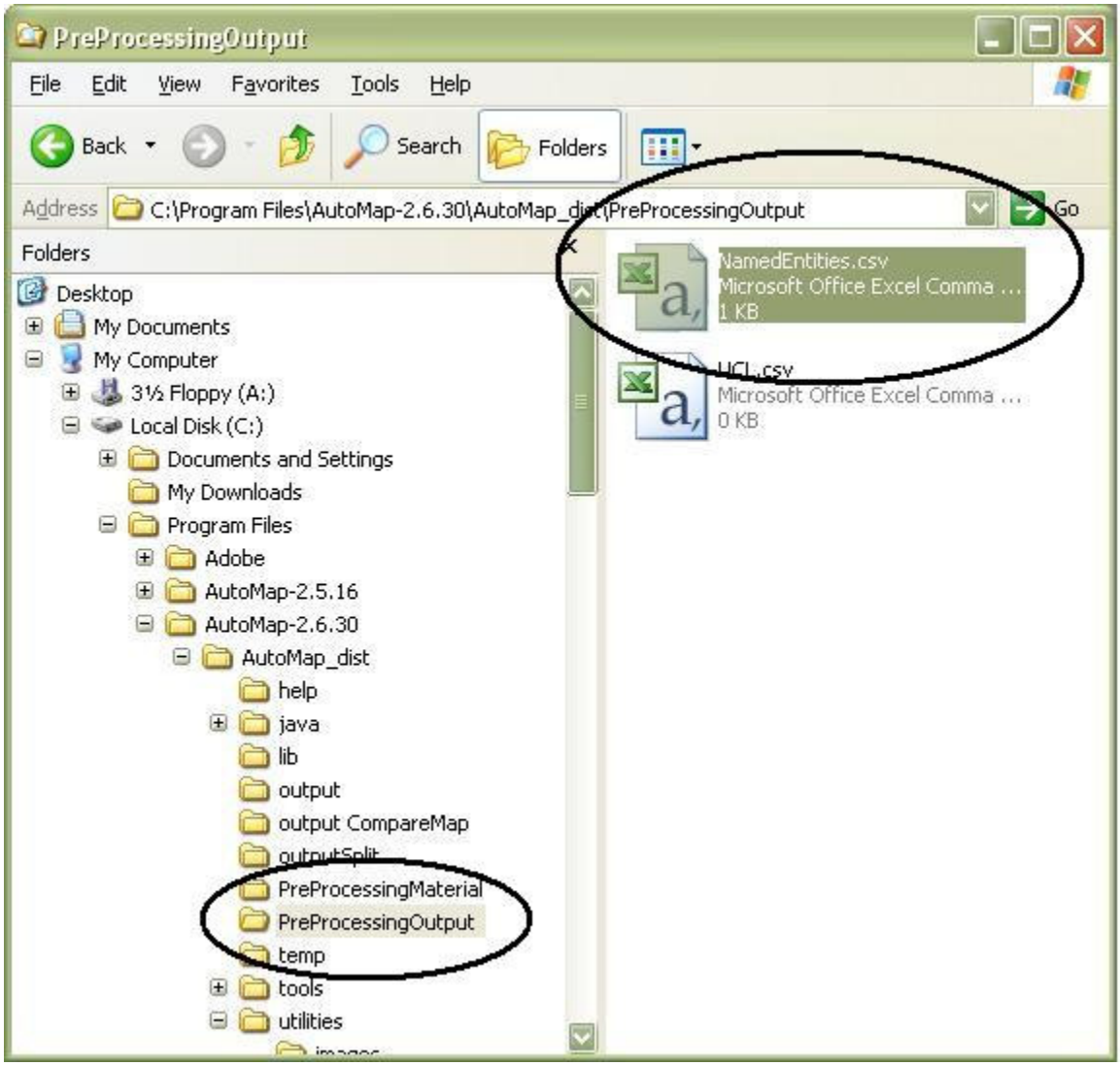
3.1 Named-Entity Recognition

More information about Named-Entity Recognition in AutoMap.
To create a list of all Name-Entities that are contained in the data set opened, go to **Utilities (tab no. 3)** in **P2** and click the **Create and save Named Entities List** button in the Named-Entities Field. The resulting list will be automatically saved under *NamedEntities.csv* in the root directory of AutoMap.

The Named-Entity Recognition interface:



The black ellipses in the screen shot below below highlight where to find the *NamedEntities.csv* file in your root directory.



3.1.1 Example for Named-Entity Recognition

Resulting *NamedEntities.csv file* after the following Small Predefined Delete List was applied with rhetorical adjacency:

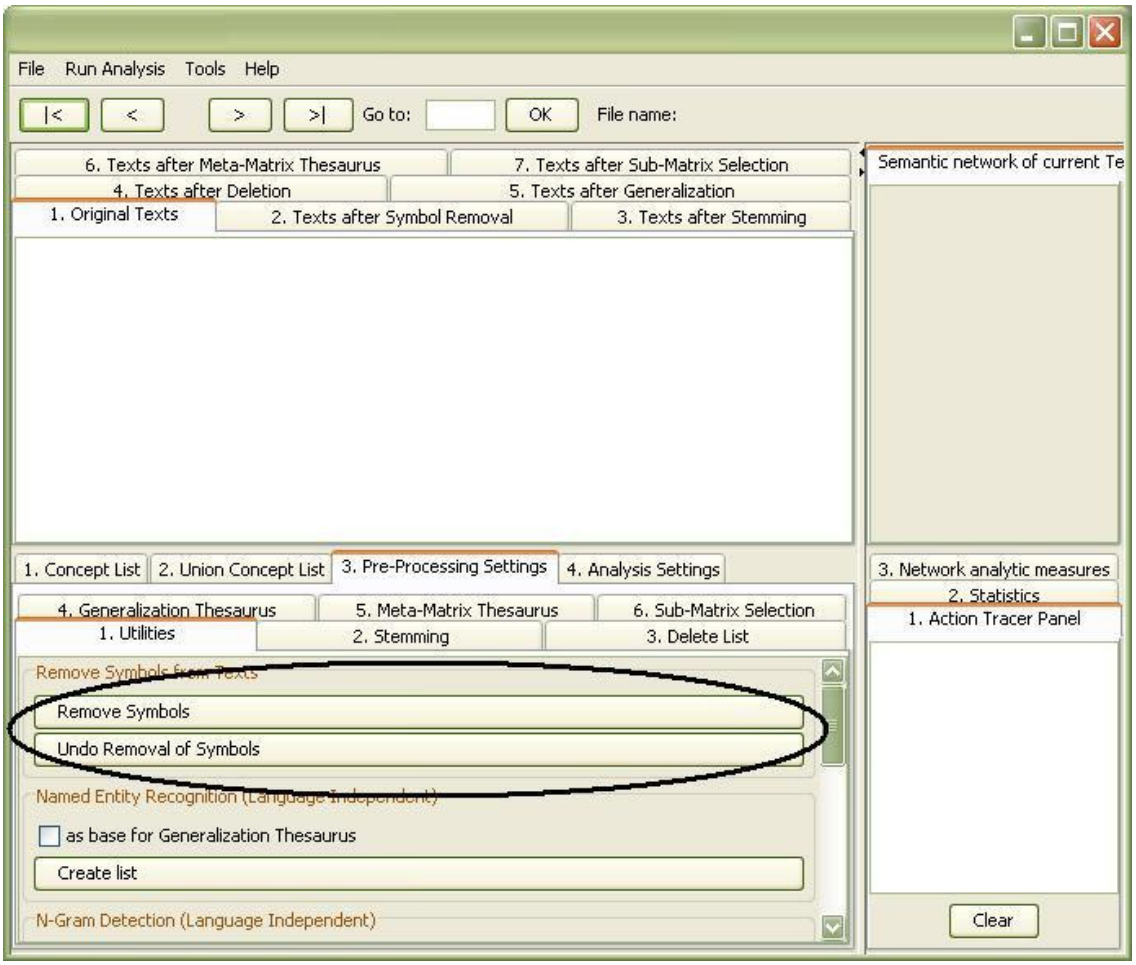
	John Darren and Mrs. Brown
a	Mr. Cray
an	He couldn't. Mrs. Brown
some	Prof. Darren Mrs. Brown & Mr.
many	Cray
this	13-Jun
that	But
these	Mr. Cray's
those	Mr. Craye
the	Problem solved. Then Mrs. Brown
all	Then Mrs. Brown
one	12-Jun
every	Weeding
	She
	She asked Mr. Cray
	Mrs. Brown

Redundant concepts can be converted to one word by *stemming*. Concepts not relevant to the user can be eliminated by *deletion*.

3.2 Symbol Removal

In the lower left panel **(P2)** you will find an option under **tab no. 3 Pre-Processing Utilities** labeled on **tab no. 1** as **Symbol Removal**. This routine

removes or strips off all characters that are neither a letter nor a number. It maintains sentence marks. It converts question marks and exclamation marks into sentence marks. This helps replace the *delchar* option on the delete list in a more user-friendly fashion. The overall purpose of this routine is to do a very thorough cleaning of the data in a fully automated, easy to use fashion. This routine can be unapplied by using the “un-apply” button, which is located close the “apply” button.



3.21 N-gram Identification: Bigrams

More information about Collocation Identification in AutoMap.
To create a list of all bigrams that are contained in the data set opened, go to the **Utilities, tab no. 3, in P2** and select the **Create Bigram (Correlation) List** button in the **N-gram Detection**. The resulting list will be automatically saved under *CorrelationList.csv* in the root directory of AutoMap.

4. Stemming

More information about Stemming.

To stem a text (or text set), go to the **Stemming (tab no. 4)** in **P2**.

Porter Stemmer:
To apply the Porter Stemmer select the **Apply** button next to Porter Stemmer and stemming of irregular verbs for English. The stemmed text(s) will be displayed on the **tab no. 2 Stemmed Text** in **P1**.

To unstem the texts, go to **tab no. 4. Stemming** in **P2** and select the **Un-Apply** button. **Tab no. 2. Stemmed Text** in **P1** will be cleared.

Krovetz Stemmer:

For the Krovetz stemmer, several customization options are offered:

1. **Decide whether capitalized words should be stemmed or not.**

   Use radio buttons in the interface to make your selection. By default, capitalized words are stemmed.

2. **Define words to be modified by the stemmer.**

   These words are collected in a *protection list*, named *selfdefined_protected_concepts.txt,* stored in the AutoMap root directory under *utilities\KStem*. To avoid stemming certain words put them in this list, one word per line, without any line delimiter.

3. **Define specific stems for certain words**.

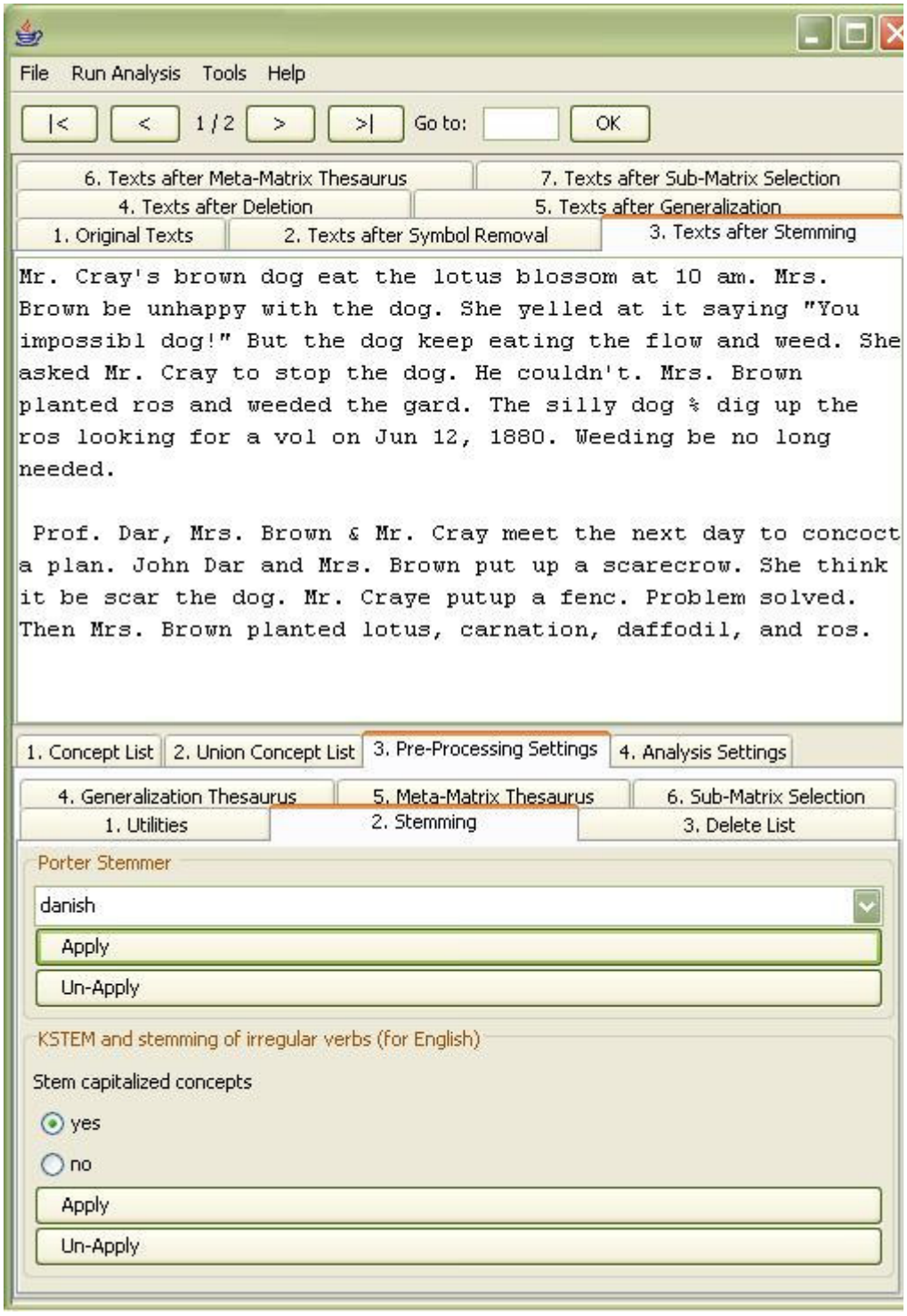
   These words are collected in a list of pairwise associations, named *selfdefined_pairs.tx*t, which are stored in the AutoMap root directory under *utilities\KStem*. To stem a certain word into a pre-defined term, put the pair (first word / pre-defined stem) in the list, one pair per line, without any line delimiter. The *selfdefined_pairs.txt list* that comes along with AutoMap contains already such pairs, which handle the correct stemming or irregular verbs in English.

To apply the Krovetz stemmer:
Select the **Apply** button next to K-stem. The stemmed text(s) will be displayed on **tab no. 2. Stemmed Text** index card in **P1**.

To unstem the texts, go to the **tab no. 4 Stemming** in **P2** and select the **Un-Apply** button. The **tab no. 2 Stemmed Text** in **P1** will be cleared.

4.1 Example for Stemming (Porter)

Stemmed text in P1 and interface of Stemming index card on P2:



5. Deletion
The Delete List is *not* case sensitive.
More information about Deletion in AutoMap.

You can use the predefined Delete Lists that AutoMap offers or create your own Delete List. All lists can be edited.

5.1 Open a Delete List

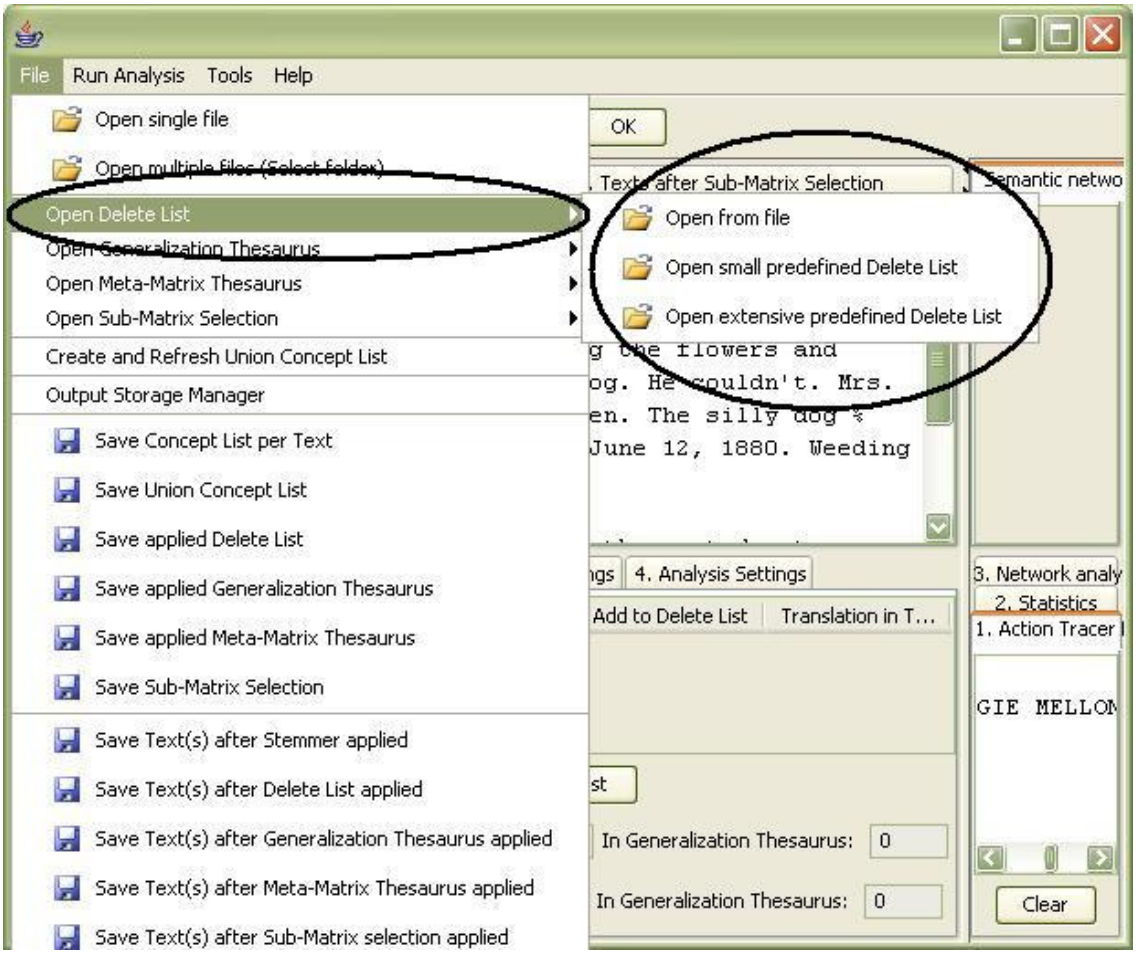
Click the **File menu,** select *Open Delete List* and choose one of the following options:

Open from file. A file chooser will appear. Select a delete list and hit the **Open** button.

Open small predefined Delete List. AutoMap's predefined small delete list will be opened.

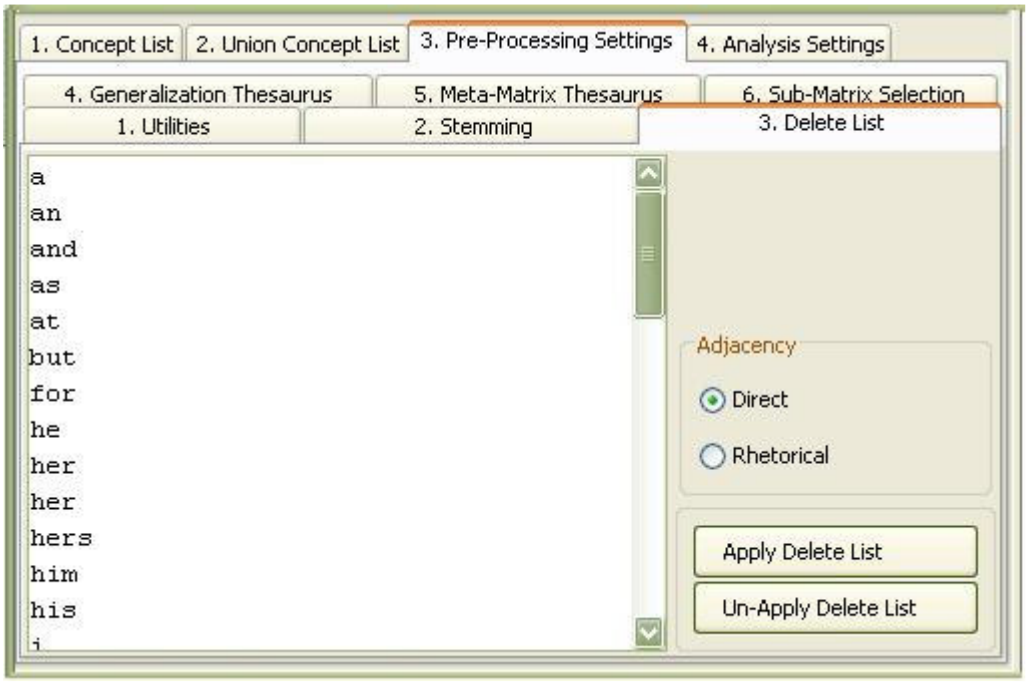
Open extensive predefined Delete List. AutoMap's extensive small delete list will be opened.

The black ellipses in the screen shot below shows where to access the **Delete List** utility:



The Delete List will be displayed in **P2, tab no. 5. Delete List index card**:

Interface of the Delete List index card:



The Delete List can be edited.

5.1.1 Small predefined Delete List
The Delete List is compiled of words that occur most frequently in English: *a, an, and, some, many, this, that, these, those, the, all, one, every.*

The Small Delete List can be edited.

5.1.2 Extensive predefined Delete List
An Extensive Delete List is based on words occurring most frequently in English: *a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself*. As the name indicates, the Extensive Delete list contains more words than the Small Predefined Delete List.

The Extensive Delete List can be edited.

5.2 Create a Delete List
There are two ways to create a Delete List:

1. Within AutoMap:
Go to the **Delete List** index card. The general structure of a Delete List is one single *concept* per line. Add concepts by typing one concept per line. Hit enter after entering a concept. Avoid empty lines. See the example for more information.

2. Outside of AutoMap:
Use a text editor to create a Delete List. Please consider these instructions to create a Delete List:

1. The general structure of a Delete List is one single Concept per line.
2. Avoid empty lines.
3. The Delete List is *not* case sensitive.
4. Save the List.
5. Open the Delete List in AutoMap.
6. You can edit the Delete List in AutoMap if you wish.

5.3. Edit a Delete List
On the **Delete List** index card you can:

- Add concepts: Type one concept per line. Hit enter after entering a concept.
- Modify concepts: Go the the line and retype concept.
- Drop concepts: Mark the concept and hit the delete key.

5.4 Apply a Delete List
If you wish to apply a Delete List and a Thesaurus we recommend first applying a Delete List and then a Thesaurus. Next, follow these steps:

1. Before applying a delete list, an adjacency option can be chosen on the **Delete List** index card. Adjacency can be either direct (default) or rhetorical. If the user does not change the adjacency option, AutoMap uses direct adjacency for deletion and analysis.
2. To delete the concepts specified in the Delete List from all texts loaded click the **Apply Delete List** button on the **Delete Concepts** Index card.
3. See the pre-processed texts in **P1, Delete List (tab no. 3)**.

When applying a Delete List AutoMap does three things:

1. Search the text(s) for concepts specified in the Delete List.
2. Delete matches from the text(s).
3. Display the resulting text(s) in **P1, Delete List (tab no. 3)**.
   If direct adjacency was chosen, concepts specified in the delete list are simply deleted from texts and concepts left and right to deleted concepts will appear adjacent to each other in terms of visualization and statement formation.

If rhetorical adjacency was chosen placeholders (xxx) are inserted where a concept was deleted. The placeholders retain original distances of maintained concepts for purposes of visualization and analysis.

To apply multiple delete lists load the first one in, apply it, then load in the next, apply it, and so on.

5.4.1 Direct Adjacency

If direct adjacency is chosen, concepts in the text that match concepts specified in the delete list will be deleted from texts. As a result concepts left and right of a deleted concept move together and will be treated as directly adjacent to each other for visualization and analysis.

To apply direct adjacency check the radio button in the **Delete List** index card. Then apply the delete list.
If the user does not change the adjacency option, AutoMap uses direct adjacency for deletion and analysis.

5.4.2 Rhetorical adjacency

If rhetorical adjacency is chosen *placeholders* "xxx" are inserted where a concept was deleted. The placeholders retain the original distances of the maintained concepts visually for analysis.

To apply direct adjacency check the button on the **Delete List** tab. Then apply the delete list.
If the user does not change the adjacency option, AutoMap uses direct adjacency for deletion and analysis.

5.5 Un-Apply a Delete List

To un-apply a Delete List that was applied to the data, in **P2** go to the **Delete List (tab no. 5)** index card and select the **Un-Apply** button. The **tab no. 3 Delete List** index card on **P1** will be cleared.

5.6 Save an applied Delete List

To save a Delete List that you have applied to the data, click the **File menu,** select **Save Delete List as**. A file chooser will pop up.

5.7 Save text(s) after application of Delete List

To save the text(s) after the application of the Delete List, click the **File menu**, select **Save Text(s) after Delete List applied.** All texts are automatically saved in a folder called "preprocessed" in the root directory of AutoMap. The filename will be *"after_DL_NameOfYourText.txt"*.

5.8 Examples for the application of a Delete List

Input text	Tool used	Setting	Resulting text
Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would	AutoMap's extensive Delete List: *a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself*	Direct Adjacency (default)	Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy with dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses.
		Rhetorical Adjacency	Mr. Cray's brown dog ate xxx lotus blossom xxx 10 am. Mrs. Brown was unhappy with xxx dog. xxx yelled xxx xxx saying "You impossible dog!" xxx xxx dog kept eating xxx flowers xxx weeds. xxx asked Mr. Cray xxx stop xxx dog. xxx couldn't. Mrs. Brown planted roses xxx weeded xxx garden. xxx silly dog % dug up xxx roses looking xxx xxx vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray

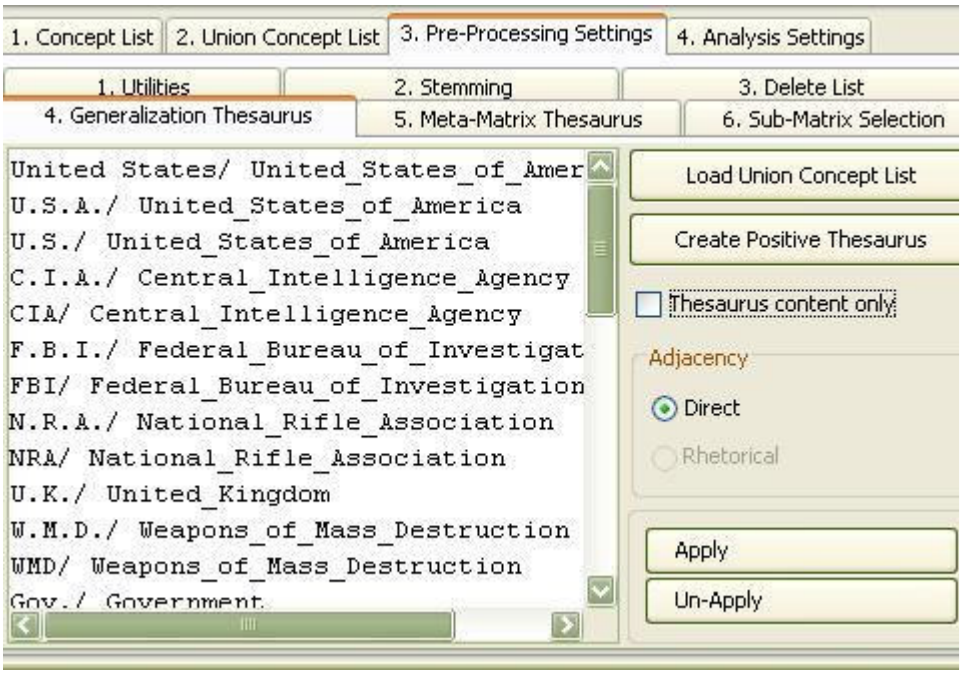
scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.			scare xxx dog. Mr. Craye put up xxx fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, xxx roses.
	The following concepts were added to the Delete List. Then the Delete List was applied again. *from, in, what, was, with*	Direct Adjacency (default)	Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown unhappy dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12, 1880. Weeding no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses.
		Rhetorical Adjacency	Mr. Cray's brown dog ate xxx lotus blossom xxx 10 am. Mrs. Brown xxx unhappy xxx xxx dog. xxx yelled xxx xxx saying "You impossible dog!" xxx xxx dog kept eating xxx flowers xxx weeds. xxx asked Mr. Cray xxx stop xxx dog. xxx couldn't. Mrs. Brown planted roses xxx weeded xxx garden. xxx silly dog % dug up xxx roses looking xxx xxx vole on June 12, 1880. Weeding xxx no longer needed. Prof. Darren,

			Mrs. Brown & Mr. Cray met xxx next day xxx concoct xxx plan. John Darren xxx Mrs. Brown put up xxx scarecrow. xxx thought xxx xxx scare xxx dog. Mr. Craye put up xxx fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, xxx roses.

6. Generalization Thesaurus
More information about thesauri.
The Generalization Thesaurus is **not** case sensitive.

6.1. Open a Generalization Thesaurus
Click the **File menu,** select Open Generalization Thesaurus. A file chooser will pop up. Double click the thesaurus you wish to wish to open or single click the thesaurus and then hit the **Open** button. The thesaurus will be displayed on **P2, 6. Generalization Thesaurus** index card.

Generalization Thesaurus Interface index card:



6.2. Create a Generalization Thesaurus
There are two ways to create a Thesaurus:

1. Within AutoMap:

Go to **P2, (tab no. 6) Generalization Thesaurus** (see also the interface of the Generalization Thesaurus index card for an example).
Use the Text Area on this Index card.
Build and edit a thesaurus.

AutoMap supports users in building a generalization thesaurus by loading the *union of concepts* from the highest level of pre-processing applied into the **Generalization Thesaurus** field. This is found on the **Generalization Thesaurus** index card and can be used upon demand.

Follow these steps to load the union concept list into the **Generalization Thesaurus** field:

1. Create or refresh the Union Concept List.
2. Hit the **Load Union Concept List** button on the **Generalization Thesaurus** index card.

This concept list loaded into AutoMap can be refined by applying *Named-Entity Recognition* and *Deletion* prior to Generalization.

Here is an example for multi-step pre-processing:

To further illustrate multi-step pre-processing techniques, copy the text passages below, then save as a TXT file as "Our Text I.txt" and "Our Text II.txt " respectivley. Load these examples into Automap to follow along.

Input texts:

Our Text I.txt

```
Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was
unhappy with the dog. She yelled at it saying "You impossible dog!" But
the dog kept eating the flowers and weeds. She asked Mr. Cray to stop
the dog. He couldn't. Mrs. Brown planted roses and weeded the garden.
The silly dog % dug up the roses looking for a vole on June 12, 1880.
Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met
the next day to concoct a plan. John Darren and Mrs. Brown put up a
scarecrow. She thought it would scare the dog. Mr. Craye put up a
fence. Problem solved. Then Mrs. Brown planted lotus, carnations,
daffodils, and roses.
```

Our Text II.txt:

```
Mr. Cray's brown dog stopped eating the lotus blossom at 12 pm the next
day. Mrs. Brown was now happy with the dog. She said "You good dog!"
The dog no longer ate the flowers and weeds. Mr. Cray was pleased too.
Mrs. Brown watered the roses and fertilized the garden on June 13,
1880. Prof. Darren, Mrs. Brown & Mr. Cray met over dinner and discussed
how the plan had worked. John Darren and Mrs. Brown would take down the
```

scarecrow the following week. She thought it was too scary for the dog. Mr. Craye painted his fence. Then Mrs. Brown watered lotus, carnations, daffodils, and roses.

Pre-processing technique applied	Entries	Result
1. Create Named-Entity List.		John Darren and Mrs. Brown Mr. Cray He couldn't. Mrs. Brown Prof. Darren Mrs. Brown & Mr. Cray 13-Jun But Mr. Cray's Mr. Craye Problem solved. Then Mrs. Brown Then Mrs. Brown 12-Jun Weeding She She asked Mr. Cray Mrs. Brown
2. Create Named-Entity List and use it to build a generalization thesaurus.	Mr. Cray's/Mr. Craye Mr. Craye/Mr. Cray's	
3. Add further words that belong together to the generalization thesaurus.	Prof. Darren/Prof_Darren	
4. Apply generalization thesaurus (no thesaurus content only).		mr. craye brown dog ate the lotus blossom at 10 am. mrs. brown was unhappy with the dog. she yelled at it saying "you impossible dog!" but the dog kept eating the flowers and weeds. she asked mr. cray to stop the dog. he couldn't. mrs. brown planted roses and weeded the garden. the silly dog % dug up the roses looking for a vole on june 12, 1880. weeding was no longer needed. mrs. brown & mr. cray, mrs. brown & mr. cray met the next day to concoct a plan. john darren and mrs. brown put up a scarecrow. she thought it would

		scare the dog. mr. cray's put up a fence. problem solved. brown planted lotus, carnations, daffodils, and roses.
3. Deletion (rhetorical adjacency).	a an some many this that these those the all one every	Mr. Cray's brown dog ate xxx lotus blossom at 10 am. Mrs. Brown was unhappy with xxx dog. She yelled at it saying "You impossible dog!" But xxx dog kept eating xxx flowers and weeds. She asked Mr. Cray to stop xxx dog. He couldn't. Mrs. Brown planted roses and weeded xxx garden. xxx silly dog % dug up xxx roses looking for xxx vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met xxx next day to concoct xxx plan. John Darren and Mrs. Brown put up xxx scarecrow. She thought it would scare xxx dog. Mr. Craye put up xxx fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.
4. Create Union Concept List.		(Coincides with Union Concept List loaded into AutoMap, see cell below)

5. Load Union Concept List into AutoMap.		1880 asked blossom brown but concoct craye daffodils dinner discussed dog dog!" down dug fence for his how it looking lotus on over plan pm problem said saying scary solved stopped take up watered weeded weeding weeds with worked yelled

Outside of AutoMap: Use a text editor to create a Thesaurus.
Build and edit a thesaurus.
Save the Thesaurus.
Open the Thesaurus in AutoMap.
You can edit the Thesaurus in AutoMap if you wish.

6.3 Edit a Generalization Thesaurus

You can add, change or drop the lines of a thesaurus on **P2, (tab no. 6) Generalization Thesaurus**.

The general structure of a Thesaurus follows the five points below (see also the interface of the Generalization Thesaurus index card for an example):

1. Every line contains *Concept/ Key Concept* or in other words *Old Word/ New Word*.
2. A *Concept* can be one or more words.
3. A *Key Concept* is one word.
4. Be sure to separate the words by a slash.
5. The Thesaurus is **not** case sensitive.

6.4 Apply a Generalization Thesaurus

If you wish to apply a Delete List and a Generalization Thesaurus please be sure to use the Delete List first and then the Thesaurus. Then go through the following process:

1. Decide if you want to use the *Thesaurus content only* option or not. If you do not select the Thesaurus content only option this setting will not be applied.
2. If you select the Thesaurus content only option you can choose an *adjacency option*.
3. Adjacency can be either direct (default) or rhetorical.
4. To apply your Generalization Thesaurus with the settings you have specified click the **Apply Thesaurus** button on the **Generalization Thesaurus** Index card. AutoMap uses the entries in the Thesaurus to search the text(s) for concepts. If a match is found it will be translated into a key concept. Again, the Thesaurus is *not* case sensitive.
5. See the pre-processed texts on the **P1, (tab no. 4) Generalization Thesaurus.**
   If the Thesaurus content only option and Direct Adjacency were chosen only key concepts would be displayed and considered for analysis. If the Thesaurus content only option and Rhetorical Adjacency were chosen key concepts and their original distances, which are symbolized by place holders (xxx), are displayed and considered for analysis.

6.4.1 Thesaurus content only

If the Thesaurus content only option is chosen AutoMap performs the following steps:

1. Search the text(s) for concepts specified in the thesaurus.
2. Translate matches into key concepts.
3. Maintain only key concepts in the pre-processed texts. The rest of the input text is dropped and will not be considered for further pre-processing or analysis. The original distances of the key concepts will not be maintained. However, punctuation marks like the end of sentences and paragraphs are maintained and considered for analysis.
4. As a result, all key concepts in the resulting text appear directly adjacent to each other.

To select the Thesaurus content only option check the **Thesaurus content only** item on **P2, (tab no. 6.) Generalization Thesaurus** index card. You can now choose to either use direct or rhetorical adjacency for the application of the Generalization Thesaurus. Then apply the Generalization Thesaurus. To switch from not using the Thesaurus content only option uncheck the Thesaurus content only item on P2, **(tab no. 6.) Generalization Thesaurus** and apply the Generalization Thesaurus again.

If the Thesaurus content only option is NOT chosen AutoMap performs the following steps:

1. Search the text(s) for concepts specified in the thesaurus.
2. Translate matches into key concepts.
3. Keep the rest of the text as it is. This means, all other concepts in the text that did not match concepts specified in the thesaurus will not be affected in any way. Original distances of both unaffected concepts and key concepts will be maintained. This rule does not apply if a concept consisting of more than one word was translated into a key concept.

The Thesaurus content only item on **P2, (tab no. 6) Generalization Thesaurus** index card by default is not checked because AutoMap does not apply the Thesaurus content only option. To switch to using the Thesaurus content only option check the Thesaurus content only item on **P2, (tab no. 6) Generalization Thesaurus** and then apply the Generalization Thesaurus again.

6.4.1.1 Direct Adjacency

Direct adjacency means that original distances of concepts that represent the key concepts will neither be visualized nor considered for analysis.

To choose the direct adjacency click the **Direct** button in the **Adjacency** field on **P2, (tab no. 6) Generalization Thesaurus** index card. Then apply the

Generalization Thesaurus. If the user does not change the adjacency option, AutoMap uses direct adjacency for generalization and analysis.

6.4.1.2 Rhetorical Adjacency

Rhetorical adjacency means that the original distance of key concepts will be considered for the analysis. Original distances of concepts that represent the key concepts will be visually symbolized by placeholders (xxx) and considered for analysis. Rhetorical adjacency can only be applied if the Thesaurus content only option was not chosen.

To choose the rhetorical adjacency click the **Rhetorical** button in the **Adjacency** field on **P2, Generalization Thesaurus (tab no. 6)**. Then apply the Generalization Thesaurus.
If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis.

6.5 Un-Apply a Generalization Thesaurus

To un-apply a Generalization Thesaurus that was applied to the data, go to **P2, Generalization Thesaurus (tab no. 6)** and hit the **Un-Apply** button. The **Generalization Thesaurus (Tab no. 4)** on **P1** will be cleared.

6.6 Save an applied Generalization Thesaurus

To save a Generalization Thesaurus that you have applied to the data**,** click the **File menu,** select **Save Generalization Thesaurus As (** a file chooser will pop up).

6.7 Save text(s) after application of Generalization Thesaurus

To save the text(s) after the application of the Generalization Thesaurus, click the **File menu**, select **Save Text(s) after Generalization Thesaurus applied.** All texts are automatically saved in a folder called "preprocessed" in the root directory of AutoMap. The filename will be after_general_thes_NameOfYourText.txt".

6.8 Example for the building and applying a Generalization Thesaurus

Applying a thesaurus to text that was not pre-processed

Input text	Tool used	Setting	Resulting text
Mr. Cray's brown dog ate the	Mr Craye/Mr_Cray Mrs Browns/Mrs_Brown John	Not Selected – Thesaurus	mr. cray's brown dog eating the lotus blossom at 10 am. mrs. brown was unhappy with

lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence.	Darren/Prof_John_Darren Prof Darren/Prof_John_Darren yelled/yelling ate/eating	content only (default)	the dog. she yelling at it saying "you impossible dog!" but the dog kept eating the flowers and weeds. she asked mr. cray to stop the dog. he couldn't. mrs. brown planted roses and weeded the garden. the silly dog % dug up the roses looking for a vole on june 12, 1880. weeding was no longer needed. prof. darren, mrs. brown & mr. cray met the next day to concoct a plan. prof_john_darren and mrs. brown put up a scarecrow. she thought it would scare the dog. mr. craye put up a fence. problem solved. then mrs. brown planted lotus, carnations, daffodils, and roses.
		Thesaurus content only, Direct Adjacency (default if Thesaurus content only is chosen)	. eating... yelling.....,.. .,... prof_john_darren.....,,,.
		Selected - Thesaurus content only, Rhetorical Adjacency	xxx. xxx xxx xxx eating xxx xxx xxx xxx xxx xxx. xxx. xxx xxx xxx xxx xxx xxx. xxx yelling xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx. xxx xxx xxx. xxx xxx xxx xxx xxx. xxx xxx. xxx. xxx xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx. xxx, xxx. xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx. prof_john_darren xxx xxx. xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx.

Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.			xxx. xxx xxx xxx. xxx xxx. xxx xxx. xxx xxx xxx, xxx, xxx, xxx xxx.

Applying a thesaurus to text that was pre-processed with a Delete List, direct adjacency:

Input text	Tool used	Setting	Resulting text
Text after application of customized extensive delete list, direct adjacency: Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy with dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses	Thesaurus (same as above)	Not selected - Thesaurus content only (default)	mr. cray's brown dog eating lotus blossom 10 am. mrs. brown was unhappy with dog. yelling saying "you impossible dog!" dog kept eating flowers weeds. asked mr. cray stop dog. couldn't. mrs. brown planted roses weeded garden. silly dog % dug up roses looking vole on june 12, 1880. weeding was no longer needed. prof. darren, mrs. brown & mr. cray met next day concoct plan. prof_john_darren mrs. brown put up scarecrow. thought scare dog. mr. craye put up fence. problem solved. then mrs. brown planted lotus, carnations, daffodils, roses.
		Selected - Thesaurus content only, Direct Adjacency (default if Thesaurus content only is chosen)	. eating... yelling.....,.. .,... prof_john_darren.....,,,.
		Thesaurus content only, Rhetorical Adjacency	xxx. xxx xxx xxx eating xxx xxx xxx xxx. xxx. xxx xxx xxx xxx xxx. yelling xxx

looking vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses.			xxx xxx xxx xxx xxx xxx xxx xxx. xxx xxx. xxx xxx xxx. xxx. xxx. xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx. xxx, xxx. xxx xxx. xxx xxx xxx xxx xxx xxx. prof_john_darren xxx. xxx xxx xxx xxx. xxx xxx xxx. xxx. xxx xxx xxx. xxx xxx. xxx xxx. xxx xxx xxx, xxx, xxx, xxx.

Applying a thesaurus to text that was pre-processed with a Delete List, rhetorical adjacency:

Input text	Tool used	Setting	Resulting text
Text after application of Customized extensive delete list, rhetorical adjacency:	Thesaurus (same as above)	Not Selected - Thesaurus content only (default)	mr. cray's brown dog eating xxx lotus blossom xxx 10 am. mrs. brown was unhappy with xxx dog. xxx yelling xxx xxx saying "you impossible dog!" xxx xxx dog kept eating xxx flowers xxx weeds. xxx asked mr. cray xxx stop xxx dog. xxx couldn't. mrs. brown planted roses xxx weeded

Mr. Cray's brown dog ate xxx lotus blossom xxx 10 am. Mrs. Brown was unhappy with xxx dog. xxx yelled xxx xxx saying "You impossible dog!" xxx xxx dog kept eating xxx flowers xxx weeds. xxx asked Mr. Cray xxx stop xxx dog. xxx couldn't. Mrs. Brown planted roses xxx weeded xxx garden. xxx silly dog % dug up xxx roses looking xxx xxx vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met xxx next day xxx concoct xxx plan. John Darren xxx Mrs. Brown put up xxx scarecrow. xxx thought			xxx garden. xxx silly dog % dug up xxx roses looking xxx xxx vole on june 12, 1880. weeding was no longer needed. prof. darren, mrs. brown & mr. cray met xxx next day xxx concoct xxx plan. prof_john_darren xxx mrs. brown put up xxx scarecrow. xxx thought xxx xxx scare xxx dog. mr. craye put up xxx fence. problem solved. then mrs. brown planted lotus, carnations, daffodils, xxx roses.
		Thesaurus content only, Direct Adjacency (default if Thesaurus content only is chosen)	. eating... yelling.....,.. .,... prof_john_darren.....,,,.
		Thesaurus content only, Rhetorical Adjacency	xxx. xxx xxx xxx eating xxx xxx xxx xxx xxx xxx. xxx. xxx xxx xxx xxx xxx xxx. xxx yelling xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx. xxx xxx xxx. xxx xxx xxx xxx xxx. xxx xxx. xxx. xxx xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx. xxx, xxx. xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx. prof_john_darren xxx xxx. xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx. xxx. xxx xxx xxx xxx. xxx xxx. xxx xxx. xxx xxx xxx, xxx, xxx, xxx xxx.

xxx xxx scare xxx dog. Mr. Craye put up xxx fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, xxx roses..			

7. Meta-Matrix Thesaurus
More information about Meta-Matrix Thesaurus.
A Meta-Matrix Thesaurus has to be applied if Meta-Matrix Analysis should be performed.
A Meta-Matrix Thesaurus associates concepts with meta-matrix categories:

- Agent
- Knowledge
- Resource
- Task/Event
- Organization
- Location
- Action
- Role
- Attribute
- Any user-defined category (as many as the user defines)

When applying a Meta-Matrix Thesaurus, AutoMap searches the text(s) for the entries specified in the Meta-Matrix Thesaurus and translates matches into related Meta-Matrix categories. If you also want to apply a Delete List or/ and a Generalization Thesaurus you will need to apply these pre-processing tools before the Meta-Matrix Thesaurus.The Meta-Matrix Thesaurus is **not** case sensitive.

You might also see the meta matrix model as implemented in AutoMap to better understand the meta-matrix.

7.1 Open a Meta-Matrix Thesaurus

Click the **File menu,** select **Open Meta-Matrix Thesaurus** and choose **Open from highest level of pre-processing.**

The union of concepts after the highest level of pre-processing applied so far will be displayed in alphabetical order on **P2, Meta-Matrix Thesaurus (tab no. 7)** index card.

Note: The Meta-Matrix Thesaurus can be edited.

If you have a pre-defined Meta-Matrix Thesaurus available that matches (some of) the concepts contained in the loaded Meta-Matrix Thesaurus you can open this file.

To do so, click the **File menu,** select **Open Meta-Matrix Thesaurus** and choose **Open from file.**
If a concept contained in the pre-defined file matches a concept in the currently opened Meta-Matrix Thesaurus the meta-matrix categories assigned to this concept in the pre-defined file will be automatically assigned to the concept in the currently opened Meta-Matrix Thesaurus.

Note: The pre-assigned Meta-Matrix Thesaurus can be edited.

7.2 Edit a Meta-Matrix Thesaurus

To each concept that appears in the Concept column of the Meta-Matrix Thesaurus you can assign special categories:

- One or multiple meta-matrix categories to a concept by checking these categories. To unselect an assignment, uncheck the meta-matrix categories.

- Self-defined meta-matrix categories (e.g. location for Copenhagen). To do this, enter the category you wish to define in the last column (User-defined) of the **Meta-Matrix Thesaurus.** You can define as many categories as you want to. You can assign a concept to multiple self-defined categories. In this case, you will need a single row for each assignment. Hit enter when you have finished a self-defined row. To unselect an assignment, uncheck the meta-matrix categories.

- No meta-matrix category to a concept. In this case, this concept will not be considered for meta-matrix pre-processing and Meta-Matrix Analysis.

7.3 Build a Meta-Matrix Thesaurus

You can build a Meta-Matrix Thesaurus outside of AutoMap by using a text editor. Please consider these instructions:

1. Every line contains *(Key) Concept/ Meta-Matrix category/ Meta-Matrix category/*.... You can assign a (key) concept to as many Meta-Matrix

categories as you want to. If a concept that appears in the Meta-Matrix Thesaurus written by you does not appear in the Concepts column of the Meta-Matrix Thesaurus this will not result in a conflict. AutoMap automatically searches for concepts contained in the Concepts column of the Meta-Matrix Thesaurus and when it finds a concept that also appears in your manually built Thesaurus it assigns the categories that you assigned to this concept to the concept on the index card.

2. Avoid empty lines.
3. Make sure to separate the words by a slash.
4. The Meta-Matrix Thesaurus is case **not** sensitive.
5. Save the Meta-Matrix Thesaurus.
6. Open the Meta-Matrix Thesaurus in AutoMap.
7. You can edit the Meta-Matrix Thesaurus in AutoMap if you wish.

7.4 Apply a Meta-Matrix Thesaurus

Meta-matrix pre-processing is a higher level of pre-processing than the application of a Delete List and a Generalization Thesaurus. Thus, if you also want to apply a Delete List or/ and a Generalization Thesaurus you will need to apply these pre-processing tools before the Meta-Matrix Thesaurus.

Follow this process:

1. Decide if you want to use the Thesaurus content only option or not.
   If you do not select the Thesaurus content only option this setting will not be applied.
2. If you select the Thesaurus content only option you can choose an adjacency option.
3. Adjacency can be either direct (default) or rhetorical.
   If you do not change the adjacency option, no adjacency option will be applied.
4. To apply your Meta-Matrix Thesaurus with the settings you have specified click the **Apply Thesaurus** button on the **Meta-Matrix Thesaurus** Index card.
   AutoMap uses the entries in the Thesaurus to search the text(s) for concepts. If a match is found it will be translated into a Meta-Matrix category. The Thesaurus is case insensitive.
5. See the pre-processed texts on the **P1, Meta-Matrix Thesaurus (tab no. 5).**
   Meta-Matrix categories are confined by tags (e.g., <agent>). If multiple Meta-Matrix categories were assigned to a concept these categories appear tagged (e.g., if police was assigned to agent and resource, than the translated text would look like this: <agent, resource>).

7.4.1 Thesaurus content only

If the Thesaurus content only option is chosen AutoMap does the following:

1. Search the text(s) for concepts specified in the Concept column of the Meta-Matrix Thesaurus.
2. Translate matches into related meta-matrix categories.
3. Maintain only meta-matrix categories in the pre-processed texts. The rest of the text is dropped and will not be considered for further pre-processing or analysis. Punctuation marks like the end of sentences and paragraphs are maintained and considered for analysis.

To select the Thesaurus content only option check the **Thesaurus content only** item on **P2, Meta-Matrix Thesaurus (tab no. 7)**. You can now choose to either direct (default) or rhetorical adjacency. Then apply the **Meta-Matrix Thesaurus**. In order to switch from using the Thesaurus content only option uncheck the Thesaurus content only item and apply the Thesaurus again.

If the Thesaurus content only option is NOT chosen AutoMap does the following:

1. Search the text(s) for concepts specified in the thesaurus.
2. Translate matches into meta-matrix categories.
3. Keep the rest of the text as it is. This means, all other concepts in the text that did not match concepts specified in the thesaurus will not be affected at all. Original distances of both unaffected concepts and meta-matrix categories will be maintained. This rule does not apply if a concept consisting of more than one word is translated into a meta-matrix category.

AutoMap by default does not select the Thesaurus content only option. Therefore, the Thesaurus content only item on **P2, Meta-Matrix Thesaurus (tab no. 7)** by default is not checked. Just apply the Meta-Matrix Thesaurus. In order to switch to using the Thesaurus content only option check the Thesaurus content only item on **P2, Meta-Matrix Thesaurus (tab no. 7)** index card and then apply the Thesaurus again.

7.4.1.1 Direct Adjacency

Direct adjacency means that original distances of concepts that represent the key concepts will neither be visualized nor considered for analysis.

To choose the direct adjacency click the **Direct** button in the **Adjacency** field on **P2, Meta-Matrix Thesaurus** index card **(tab no. 7)**. Then apply the Meta-Matrix Thesaurus. If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis. As a result, only meta-matrix categories are displayed on **P1, Meta-Matrix Thesaurus (tab no. 5)** will be considered for analysis. All meta-matrix tags in the resulting text appear directly adjacent to each other.

7.4.1.2 Rhetorical Adjacency

Rhetorical adjacency can only be applied if the Thesaurus content only option was not chosen.
To choose the rhetorical adjacency click the **Rhetorical** button in the **Adjacency** field on **P2, (tab no. 7) Meta-Matrix Thesaurus**. Then apply the Meta-Matrix Thesaurus. If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis. As a result, the meta-matrix tags and the rest of the text are displayed on **P1, (tab no. 5) Meta-Matrix Thesaurus** and will be considered for analysis. Original distances of meta-matrix tags that represent the key concepts will be visually symbolized by placeholders (xxx) and considered for analysis.

7.5 Un-Apply a Meta-Matrix Thesaurus

To un-apply a Meta-Matrix Thesaurus that was applied to the data, go to **P2, (tab. no. 7) Meta-Matrix Thesaurus** and hit the **Un-Apply** button. The **tab no. 5 Meta-Matrix Thesaurus** on **P1** will be cleared.

7.6 Save an applied Meta-Matrix Thesaurus

If you wish to save a Meta-Matrix Thesaurus you first need to apply it.
To save the Thesaurus, click the **File menu,** select **Save Meta-Matrix Thesaurus as**. A file chooser will pop up.

7.7 Save text(s) after application of Meta-Matrix Thesaurus

To save the text(s) after the application of the **Meta-Matrix Thesaurus**, click the **File menu**, select **Save Text(s) after Meta-Matrix Thesaurus applied.** All texts are automatically saved in a folder called "preprocessed" in the root directory of AutoMap. The filename will be after_MMCatThes_NameOfYourText.txt".

7.8 Example for editing and applying a Meta-Matrix Thesaurus

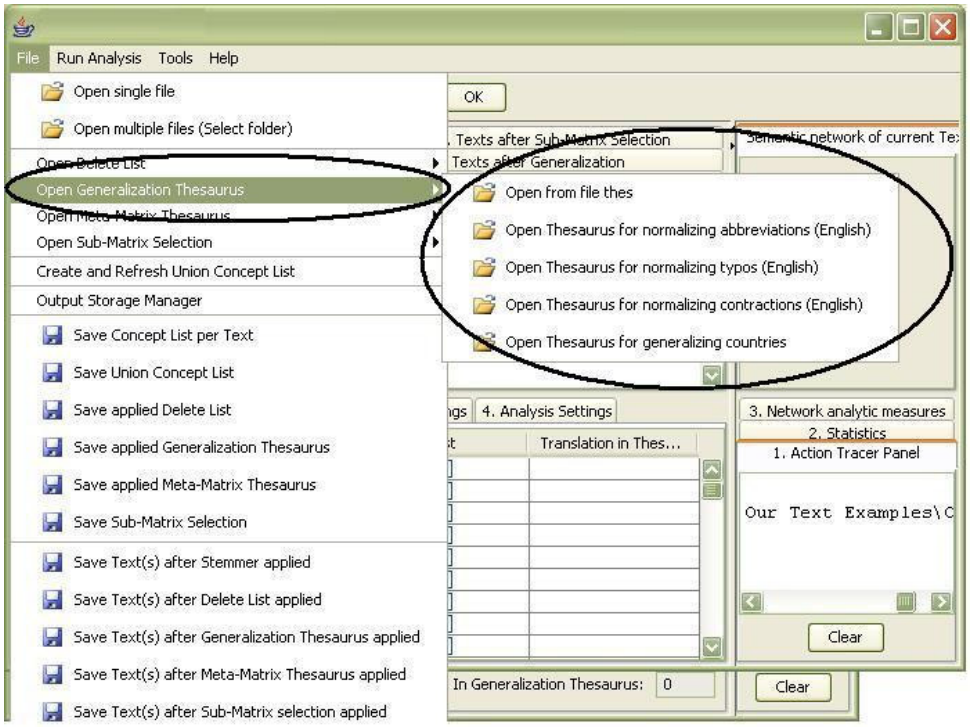
An Extract from the **Our Text I.txt** was used as input:

```
Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was
unhappy with the dog. She yelled at it saying "You impossible dog!" But
the dog kept eating the flowers and weeds. She asked Mr. Cray to stop
the dog. He couldn't. Mrs. Brown planted roses and weeded the garden.
The silly dog % dug up the roses looking for a vole on June 12, 1880.
Weeding was no longer needed.

Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan.
John Darren and Mrs. Brown put up a scarecrow. She thought it would
scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs.
Brown planted lotus, carnations, daffodils, and roses.
```

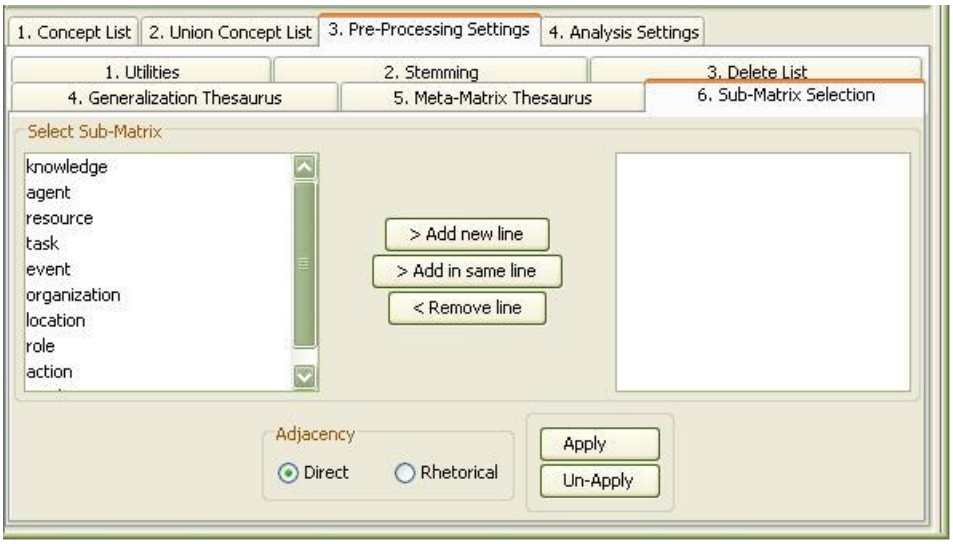
The customized extensive Delete List was applied to this text. The resulting text looks like this:

```
Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy
with dog. yelled saying "You impossible dog!" dog kept eating flowers
weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses
weeded garden. silly dog % dug up roses looking vole on June 12, 1880.
Weeding was no longer needed.

Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John
Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up
fence. Problem solved. Then Mrs. Brown planted lotus, carnations,
daffodils, roses.
```

Then, open the Meta-Matrix Thesaurus by clicking the **File menu,** selecting **Open Meta-Matrix Thesaurus**, and choosing **Open from highest level of pre-processing.** The black ellipse in the screen shot below underscore how to open a Meta-Matrix Thesaurus from the file menu.



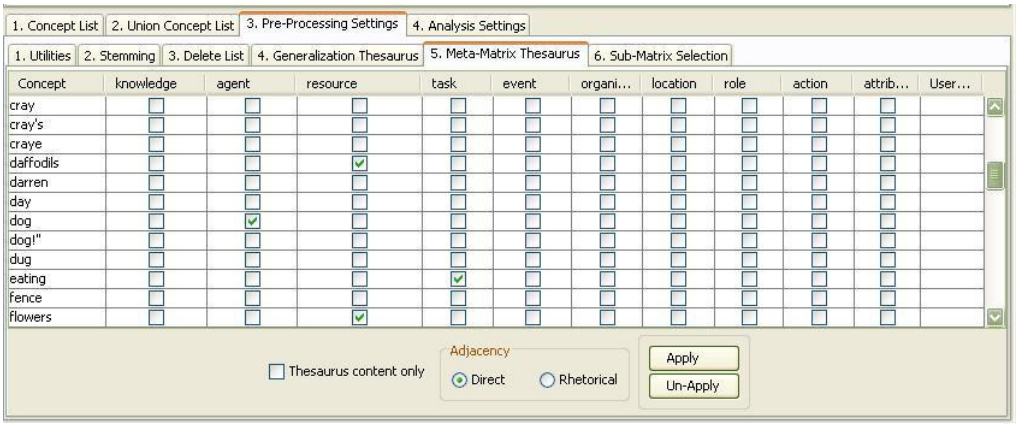
The union of concepts from the highest level of pre-processing will be displayed in alphabetical order on **P2, tab no. 3 - Preprocessing Settings, tab no. 5 - Meta-Matrix Thesaurus**. As we did not pre-process the text the original input sentence is used for input for the Meta-Matrix Thesaurus.



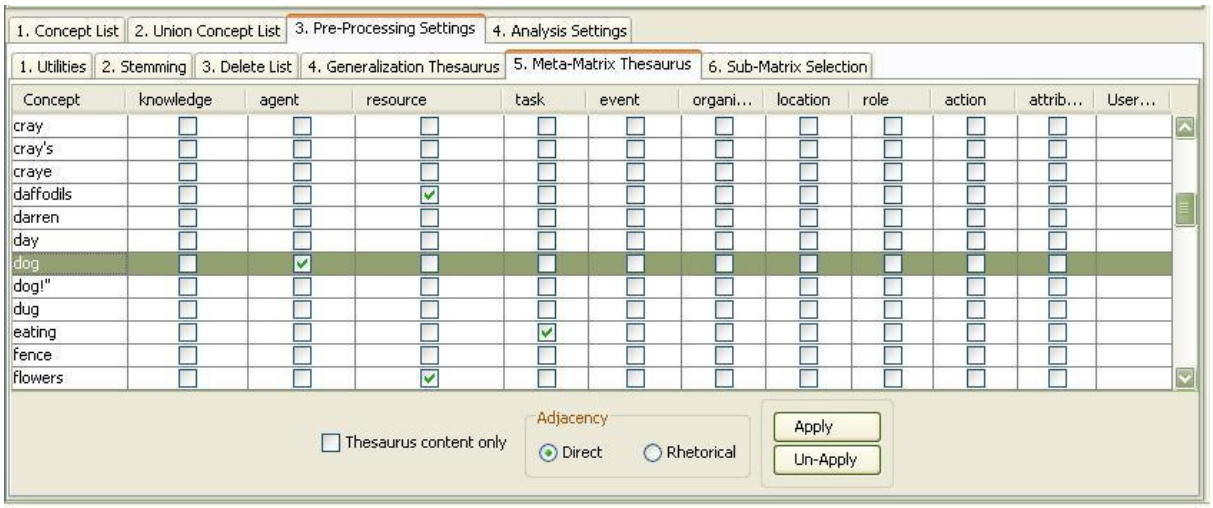
Furthermore, we have prepared a Meta-Matrix Thesaurus that we stored on our machine. This file looks like this:

Mr_Cray/agent
Mrs_Brown/agent
Prof_John_Darren/agent
dog/agent
flowers/resource
lotus/resource
roses/resource
carnations/resource
daffodils/resource
weeds/resource
weeds/task
planting/task
eating/task
yelling/task
met/task

We clicked the **File menu,** selected **Open Meta-Matrix Thesaurus** and chose **Open from file.**
AutoMap searched the opened Meta-Matrix Thesaurus for the words contained in the prepared Thesaurus. When it found a match it assigned the words in the opened Thesaurus the Meta-Matrix categories that were assigned to the same concept in the pre-defined file. Below is the result:



Now we edit the Thesaurus by modifying some of the pre-assignments (e.g., centre) and adding assignments for concepts not assigned to Meta-Matrix categories yet (e.g., contact, copenhagen). Not all concepts were associated with Meta-Matrix categories (e.g., mobile).



In the next step we applied the **Apply Meta-Matrix Thesaurus** with the following settings:

Setting	Resulting text
Not Thesaurus content only (default)	mr . cray's brown < agent> < task> the < resource> blossom at 10 am . mrs . brown was unhappy with the < agent> . she < task> at it saying "you impossible dog!" but the < agent> kept < task> the < resource> and < task> . she asked mr . cray to stop the < agent> . he couldn't . mrs . brown planted < resource> and weeded the garden . the silly <

	agent> % dug up the < resource> looking for a vole on june 12 , 1880 . weeding was no longer needed . prof . darren , mrs . brown & mr . cray < task> the next day to concoct a plan . < agent> and mrs . brown put up a scarecrow . she thought it would scare the < agent> . mr . craye put up a fence . problem solved . then mrs . brown planted < resource> , < resource> , < resource> , and < resource> .
Thesaurus content only, Direct Adjacency (default if Thesaurus content only is chosen)	. < agent> < task> < resource> . . < agent> . < task> < agent> < task> < resource> < task> . . < agent> . . . < resource> . < agent> < resource> . . . . . < task> . < agent> . . < agent> . . . . . < resource> < resource> < resource> < resource> .
Thesaurus content only, Rhetorical Adjacency	xxx . xxx xxx < agent> < task> xxx < resource> xxx xxx xxx xxx . xxx . xxx xxx xxx xxx xxx < agent> . xxx < task> xxx xxx xxx xxx xxx xxx xxx xxx < agent> xxx < task> xxx < resource> xxx < task> . xxx xxx xxx . xxx xxx xxx xxx < agent> . xxx xxx . xxx . xxx xxx < resource> xxx xxx xxx xxx . xxx xxx < agent> xxx xxx xxx xxx < resource> xxx xxx xxx xxx xxx xxx xxx xxx . xxx xxx xxx xxx xxx . xxx xxx xxx . xxx xxx . xxx xxx xxx . xxx < task> xxx xxx xxx xxx xxx xxx xxx . < agent> xxx xxx . xxx xxx xxx xxx xxx . xxx xxx xxx xxx xxx xxx < agent> . xxx . xxx xxx xxx xxx . xxx xxx . xxx xxx . xxx xxx < resource> < resource> < resource> xxx < resource> .

8. Sub-Matrix Selection

The Sub-Matrix Selection enables the user to re-translate concepts represented by a Meta-Matrix category in order to run Sub-Matrix Analysis. If input texts (no matter if they were pre-processed with a Delete list or not) were used in order to generate the Concept List for the Meta-Matrix Thesaurus, concepts represented by a Meta-Matrix category will be translated into text-level concepts. If input texts (no matter if they were pre-processed with a Generalization Thesaurus before applying the Meta-Matrix Thesaurus, concepts represented by a Meta-Matrix category will be translated

into key concepts.
The **Thesaurus content only option** always automatically applies for the **Sub-Matrix Selection**.

8.1 Select Sub-Matrix Categories|

Precondition: Sub-Matrix Selection can only be performed if Meta-Matrix Thesaurus was applied.
There are 4 ways to select sub matrices:

8.1.1 If you do not have a sub matrix selection file available

Create and modify a sub matrix selection.

8.1.2 If you do not have a sub matrix selection file available and want to select the full meta matrix (means all cells in the meta matrix)

Go to the **File** menu, Select **Open Sub-Matrix Selection**, click on **Select Full Meta Matrix.** AutoMap dynamically generates all combinations of meta matrix categories as specified in the meta matrix thesaurus, including user-defined categories, that represent all cells of the meta matrix and display these combinations in the left window on **P2, 8. Sub-Matrix Selection**.
You can modify this sub matrix selection.
This is the full meta-matrix:

agent/agent/knowledge/organization/task-event/resource/location/role/action/attribute
knowledge/agent/knowledge/organization/task-event/resource/location/role/action/attribute
organization/agent/knowledge/organization/task-event/resource/location/role/action/attribute
task-event/agent/knowledge/organization/task-event/resource/location/role/action/attribute
resource/agent/knowledge/organization/task-event/resource/location/role/action/attribute
location/agent/knowledge/organization/task-event/resource/location/role/action/attribute
role/agent/knowledge/organization/task-event/resource/location/role/action/attribute
action/agent/knowledge/organization/task-event/resource/location/role/action/attribute
attribute/agent/knowledge/organization/task-event/resource/location/role/action/attribute

8.1.3 If you have a sub matrix selection file available

Go to the **File** menu, Select **Open Sub-Matrix Selection**, click on **Open from file.**
You can modify this sub matrix selections.

8.1.4 If you want to write your own sub matrix selection file and load it into AutoMap:

Build your own sub matrix selection file outside of AutoMap, save it in .txt format, and load it into AutoMap. Go to the **File** menu, Select **Open Sub-Matrix Selection**, click on **Open from file.** You can modify this sub matrix selections.

8.2 Create or Modify Sub-Matrix Selection

Go to **P2, 8. Sub-Matrix Selection**, click on a Sub-Matrix category you wish to select and hit the **> Add in same line** button. The selected category appears in the right text field on **P2, (tab no. 8) Sub-Matrix Selection**.

You can select as many Sub-Matrix categories per row as you wish by clicking on a category in the left window on **P2, tab no. 8 Sub-Matrix Selection** and move it to the right window by clicking the **> Add in same line** button. Additionally, you can select as many rows of sub matrix selections as you wish by clicking on a category in the left window on **P2, tab no. 8 Sub-Matrix Selection** and move it to the right window by clicking the **> Add in new line** button. To add a further category to a new line, first single click on this category in the right window on **P2, tab no. 8 Sub-Matrix Selection**, so that the category is highlighted in blue. To **unselect** a selected Sub-Matrix category, click on the row in the right window and hit the **< Remove line** button on **P2, tab no. 8 Sub-Matrix Selection**. The row will disappear from the right window.

Tip!

- *If you wish to analyze any relation between e.g. agents and organizations, you need to select agent/ organization in one row, and organization/ agent in another row. If you wish to analyze any relations among all agents, select agent/agent.*
- *If you wish to analyze any relations among all agents and their relation with all organizations, select agent/agent/organization and in another row organization/agent/agent/.*

8.3 Apply Sub-Matrix Selection

1. The **Thesaurus content only option** always automatically applies for the **Sub-Matrix Selection**. Select an adjacency option. Adjacency can be either direct (default) or rhetorical.
   If you do not change the adjacency option, direct adjacency will be applied.
2. To apply your Meta-Matrix Thesaurus with the settings you have specified click the **Apply Sub-Matrix Selection** button on the **Sub-Matrix Selection** Index card.
   AutoMap uses the entries in the Thesaurus to search the text(s) for meta-matrix tags. If a match is found it will be translated into the related concept. The Thesaurus is case insensitive.
3. See the pre-processed texts on the **P1, tab no. 6 Sub-Matrix Text.**

8.3.1 Direct Adjacency

Direct adjacency means that original distances of concepts that represent meta-matrix categories will neither be visualized nor considered for analysis.

To choose the direct adjacency click the **Direct** button in the **Adjacency** field on **P2, tab no. 8 Sub-Matrix Selection** index card. Then apply the Sub-Matrix Selection.
If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis.
As a result, only concepts that represent meta-matrix categories are displayed on **P1, tab no. 6 Sub-Matrix Text** index card and will be considered for analysis. All concepts in the resulting text appear directly adjacent to each other.

8.3.2 Rhetorical Adjacency

To choose the rhetorical adjacency click the **Rhetorical** button in the **Adjacency** field on **P2, tab no. 8 Sub-Matrix Selection** index card. Then apply the Sub-Matrix Selection. If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis.

As a result, concepts that represent meta-matrix categories are displayed on **P1, tab no. 6 Sub-Matrix Text** and will be considered for analysis. Original distances of concepts that represent meta-matrix categories will be visually symbolized by placeholders (xxx) and considered for analysis.

8.4 Un-Apply a Sub-Matrix Selection

To un-apply a **Sub-Matrix Selection** that was applied to the data, go to **P2, tab no. 8 Sub-Matrix Selection** index card and hit the **Un-Apply** button. The **tab no. 6 Sub-Matrix Text** on **P1** will be cleared.

8.5 Save Sub-Matrix Selection

Apply the Sub-Matrix Selection before you save it.
To save a Sub-Matrix Selection (the content of the right window on **P2, tab no. 8 Sub-Matrix Selection**, click the File menu, select **Save applied Sub-Matrix Selection** as applied.

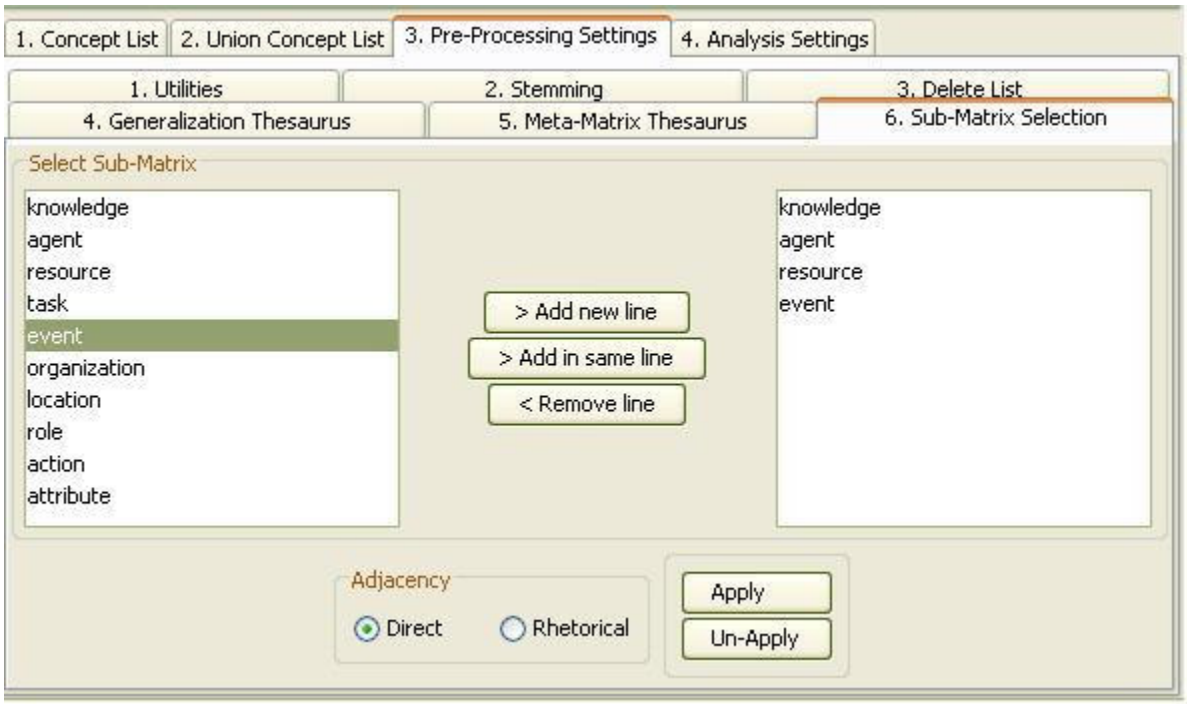
8.6 Save text(s) after Sub-Matrix Selection

To save the text(s) after the application of the **Sub-Matrix Selection**, click the **File menu**, select **Save Text(s) after Sub-Matrix Selection applied.** All texts are automatically saved in a folder called "preprocessed" in the root directory of AutoMap. The filename will be after_SubMatrixSelection_NameOfYourText.txt".

8.7 Example for Sub-Matrix Selection

The Sub-Matrix Selection as shown here is based on the example for the Meta-Matrix Thesaurus.

We opened **P2, tab no. 3 - Pre-Processing Settings, tab no. 6- Sub-Matrix Selection**.

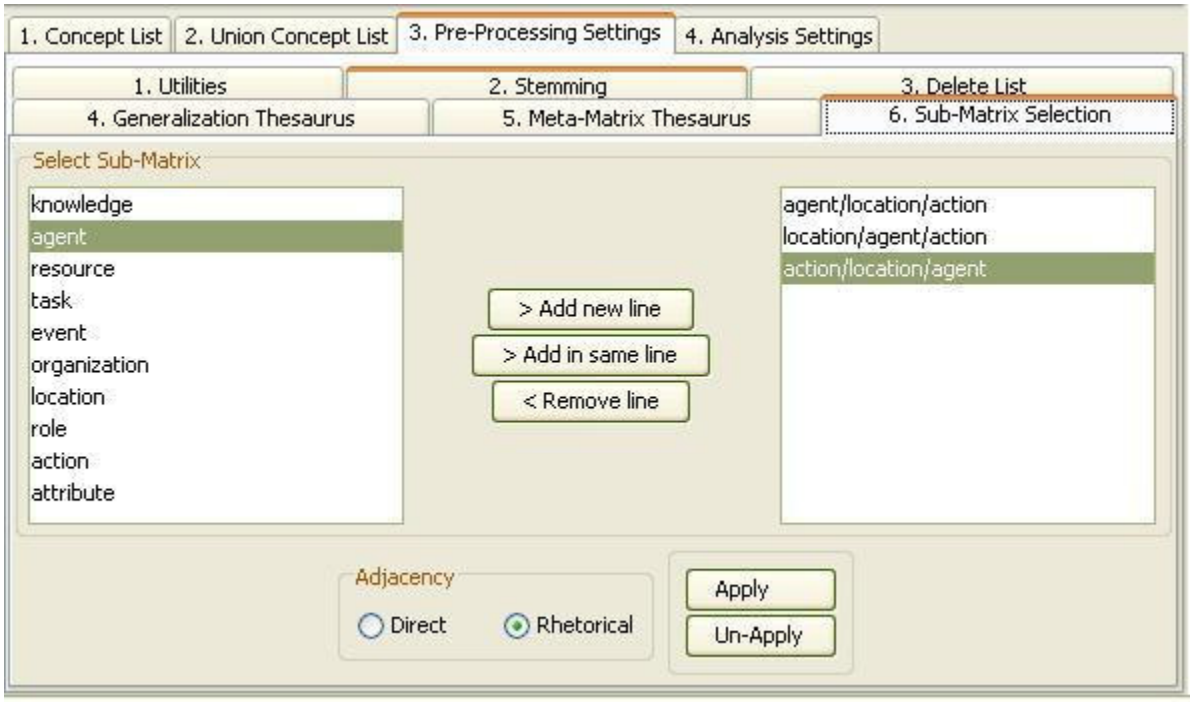


Then we applied the **Sub-Matrix Selection** with the following settings:

Setting	**Resulting text**
Select: Agent/ Location/ Action Location/ Agent/ Action Action/ Location/ Agent Direct Adjacency (default)	. dog.. dog. dog.. dog.... dog,.. .,... prof_john_darren.. dog.....,,,.
Select: Agent/ Location/ Action Location/ Agent/ Action Action/ Location/ Agent Rhetorical Adjacency (see also next picture)	xxx. xxx xxx dog xxx xxx xxx xxx xxx xxx xxx. xxx. xxx xxx xxx xxx xxx dog. xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx dog xxx xxx xxx xxx xxx. xxx xxx xxx. xxx xxx xxx xxx dog. xxx xxx. xxx. xxx xxx xxx xxx xxx xxx xxx. xxx xxx dog xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx xxx

	xxx. xxx, xxx. xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx. prof_john_darren xxx xxx. xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx dog. xxx. xxx xxx xxx xxx. xxx xxx. xxx xxx. xxx xxx xxx, xxx, xxx, xxx xxx.
Select: Full Meta-matrix Direct Adjacency (default)	. dog eating lotus.. dog. yelling dog eating flowers weeds.. dog... roses. dog roses,.. .,.. met. prof_john_darren.. dog..... lotus, carnations, daffodils, roses.
Select: Full Meta-matrix Rhetorical Adjacency (default)	xxx. xxx xxx dog eating xxx lotus xxx xxx xxx xxx. xxx. xxx xxx xxx xxx xxx dog. xxx yelling xxx xxx xxx xxx xxx xxx xxx xxx dog xxx eating xxx flowers xxx weeds. xxx xxx xxx. xxx xxx xxx xxx dog. xxx xxx. xxx. xxx xxx roses xxx xxx xxx xxx. xxx xxx dog xxx xxx xxx xxx roses xxx xxx xxx xxx xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx xxx xxx. xxx, xxx. xxx xxx xxx. xxx met xxx xxx xxx xxx xxx xxx xxx. prof_john_darren xxx xxx. xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx dog. xxx. xxx xxx xxx xxx. xxx xxx. xxx xxx. xxx xxx lotus, carnations, daffodils, xxx roses.

Detailed example for Sub-Matrix Selection:

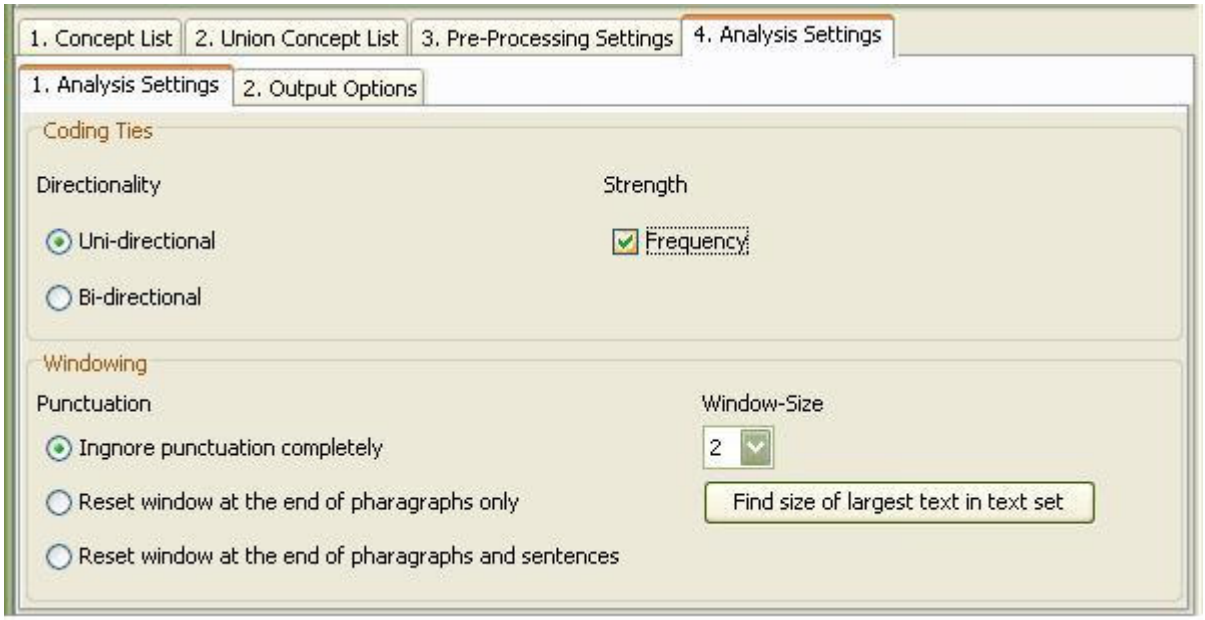


Statement Formation Choices

Statement formation choices can be made after pre-processing data and before running analysis. These choices define *if, how, and where concepts will be linked*. Statement formation settings will be applied to the highest level of data pre-processing that was applied. If no pre-processing was performed, statement formation settings apply to the original input text. If the user does not modify the statement formation settings AutoMap uses a set of default settings.

1. Analysis Settings

To specify the Analysis Settings or make the Analysis Settings, use the **Analysis Settings Index Card** .
Your settings will be automatically applied in the analysis. You do not need to confirm them.
If you do not want to change any of the suggested options a set of standard settings will be applied. The screen shot below shows an example of the Automap Analysis Settings tab in **P2**.

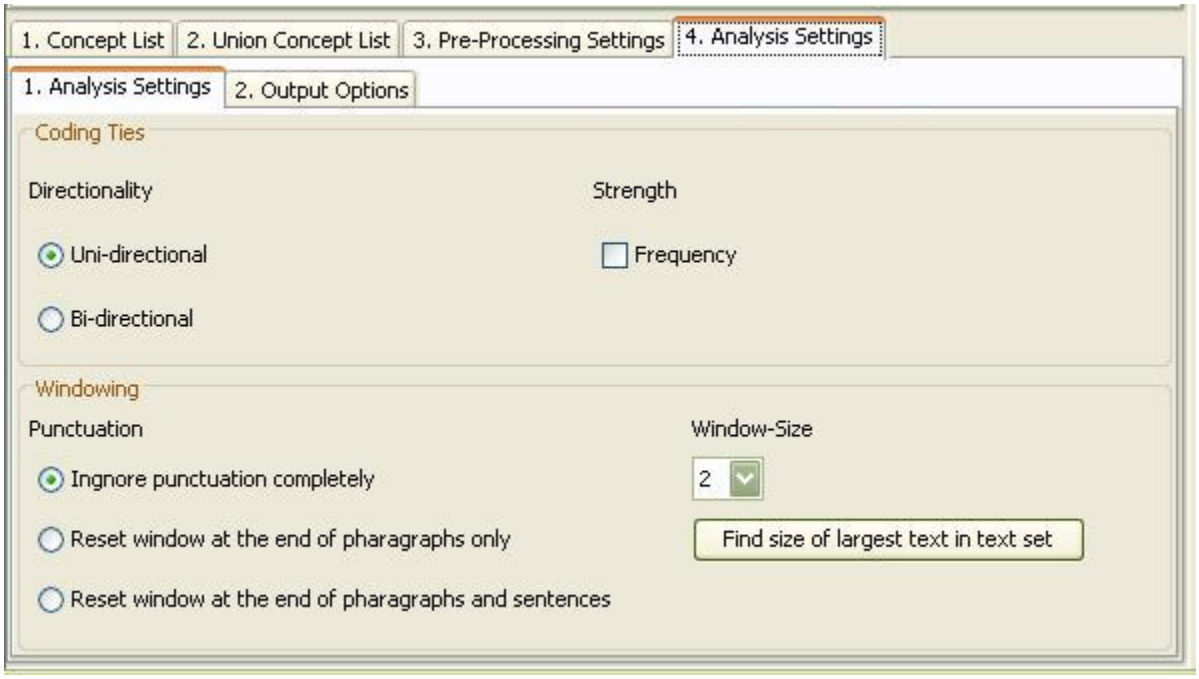


Overview on the possible Settings:

Coding Ties Specify the way statements are counted.	**Directionality** Select one of the following two possibilities by checking the button.	Uni-Directional (When coding a tie, only 1st->2nd concept should be noted)
		Bi-Directional (When coding a tie, both 1st <-> 2nd concept shall be noted)
	Strength Strength will be printed out by default. To not print out Strength uncheck the item.	**Frequency** (The cumulative frequency of every existing statement.) **Item not checked:** Existence of frequency will be printed out (binary result).
Windowing Windowing is a method that codes the text as a map by placing relationships between pairs of Concepts that occur within a window.	**Punctuation** Select one option by checking the radio button.	**Ignore punctuation completely** (Statements will be placed between all concepts.)
		Reset window at end of paragraphs only (Statements will be placed only within every single paragraph.)
		Reset window at end of paragraphs and sentences (Statements will be placed only within every single sentence.)
	Window Size Select one window size by using the number chooser.	Window size between **2 and 100.** The Window Size defines how distant concepts can be and still have a relation ship. Only concepts in same window can form statements.

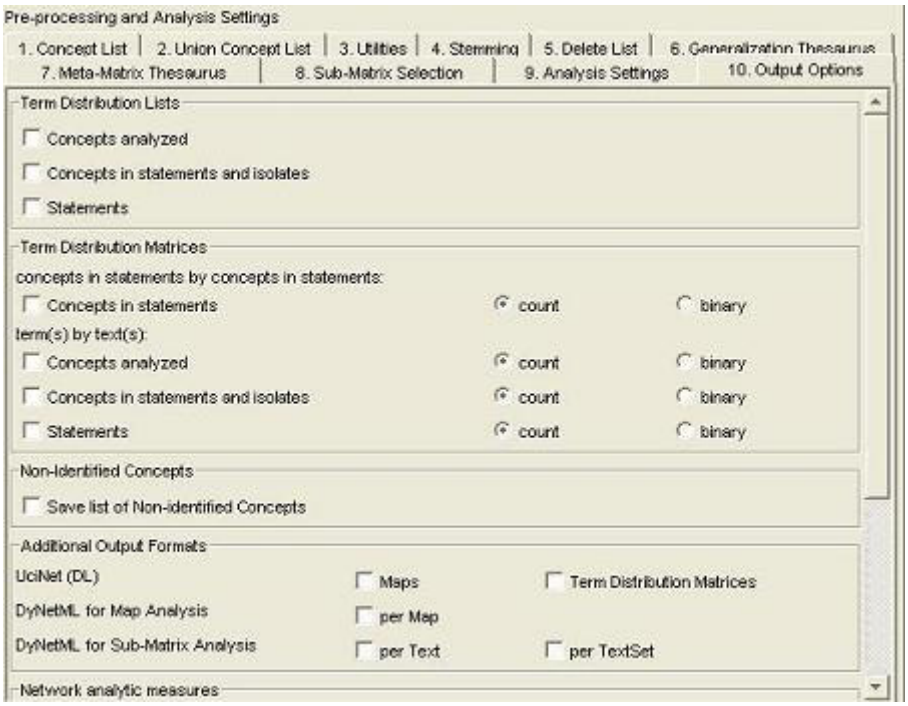
1.1 Default Settings

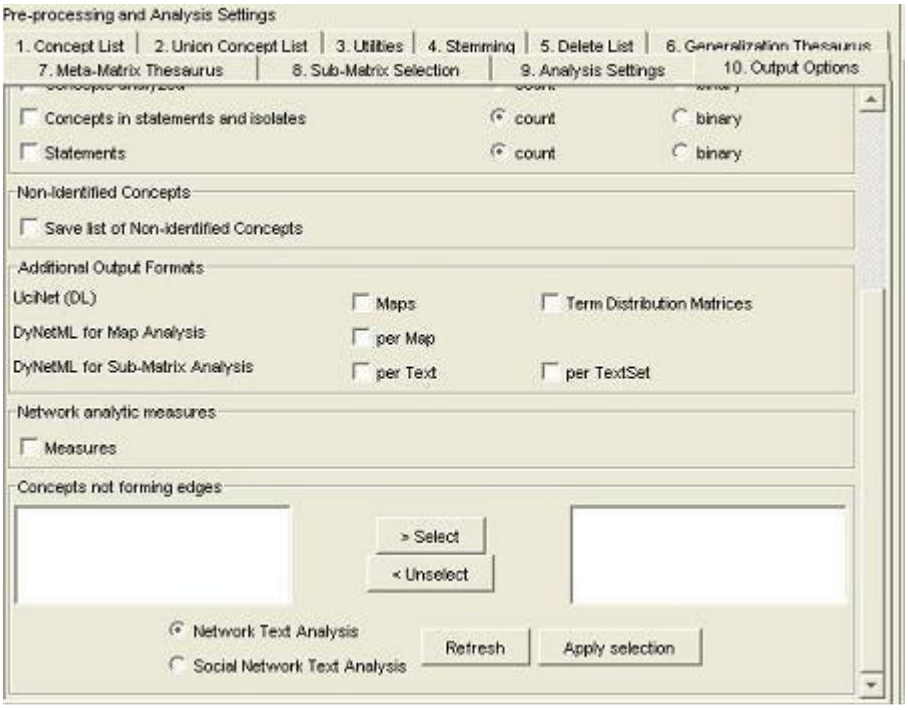
If you do not want to change any of the suggested options the analysis will be done with the following Standard Settings:



Output Options

The map and the statistic output generated by AutoMap are displayed on P3 and P4, respectively. In addition to that AutoMap offers further output options that can be chosen on **P2, 10. Output Options** index card. All additional outputs are only generated after analyses were run.





1. Term Distribution Lists and Matrices

Term distribution list and matrices as output options for all types of **multiple** analysis can be chosen in the upper two fields of **P2, tab no. 10. Output Options**.

Points To Consider

- No list or matrix will be generated by default.
- All list or matrix are only generated if the user checks a list he or she wants to be generated and runs an analysis of any type.
- If pre-processing was performed, the list relates to the stage of pre-processing that was used for the analysis. If several pre-processing techniques were applied, analysis will always be run on the highest stage of pre-processing.
- The requested lists and matrices will be automatically saved in a folder called **Term Distribution Lists and Matrices** under the root directory of AutoMap.
- The **Term Distribution Lists and Matrices** folder is overwritten with every new analysis you run. So if you want to save the results of a current **Term Distribution Lists and Matrices** folder just rename the folder.
  For each Term Distribution List checked two output lists are generated.
- For each Term Distribution Matrix checked one output matrix is generated.
  See the output examples for the content of each list.

1.1 Types and content of Term Distribution Lists and Matrices

Output Type		Name of output	Content of output
Term Distribution List	**Concepts analyzed**	List of concept analyzed.csv	Concept, Text, Frequency
		Statistics of concepts analyzed.csv	Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts
	Concepts in statements and isolates	List of concept in statements.csv	Concept, Text, Frequency
		Statistics of concept in statements.csv	Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts
		List of isolates in statements.csv	Concept, Text, Frequency
		Statistics of isolates.csv	Concept, Cumulated sum across text set, Number of text concept occurs in,

			Percentage of texts concept occurs in, Texts
	Statements	List of statements.csv	Statement, Text, Frequency
		Statistics of statements.csv	Statement, Cumulated sum across text set, Number of text statement occurs in, Percentage of texts statement occurs in, Texts
Term Distribution Matrices	**Concepts in statements by concepts in statements:**		
	Concepts in statements	Matrix of concepts in statements.csv	Matrix of Concept that were linked into statements (first row) by Concept that were linked into statements (first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept
	Term(s) by text(s):		
	Concepts analyzed	Matrix of concept analyzed.csv	Matrix of Concept (union of concepts listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept
	Concepts in statements and isolates	Matrix of concept in statements.csv	Matrix of Concept (union of concepts listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept
	Statements	Matrix of statements.csv	Matrix of Statements (union of statements listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of statement If binary was chosen, cells denote existence (1) or absence (0) of statement

1.2 Example for Term Distribution List and Matrices

Extracts from the Denmark.txt and USA.txt files were used as input:

Our Text 1.txt: Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.

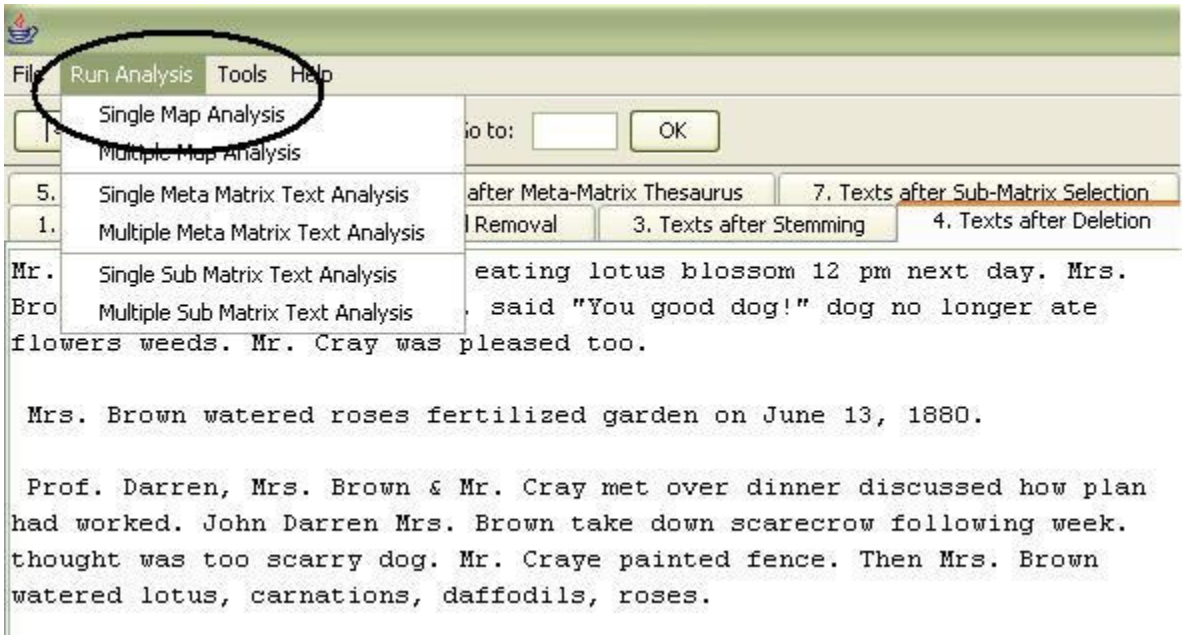
Our Text II.txt: Mr. Cray's brown dog stopped eating the lotus blossom at 12 pm the next day. Mrs. Brown was now happy with the dog. She said "You good dog!" The dog no longer ate the flowers and weeds. Mr. Cray was pleased too. Mrs. Brown watered the roses and fertilized the garden on June 13, 1880. Prof. Darren, Mrs. Brown & Mr. Cray met over dinner and discussed how the plan had worked. John Darren and Mrs. Brown would take down the scarecrow the following week. She thought it was too scary for the dog. Mr. Craye painted his fence. Then Mrs. Brown watered lotus, carnations, daffodils, and roses.

Then AutoMap's **Extensive Delete List**, **Direct Adjacency** was applied to both texts. The Delete List was extended by further *non-content bearing words* that appeared in the sample texts (*a an and as at awhile but for from happening he her her hers him his i in into it its me mine my nor of or our she so that the their theirs them they to us was we were what who whoever whom whomever will would you your yours yourself*). Below is the resulting texts:

Our Text 1.txt: Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy with dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses.

Our Text II.txt: Mr. Cray's brown dog stopped eating lotus blossom 12 pm next day. Mrs. Brown was now happy with dog. said "You good dog!" dog no longer ate flowers weeds. Mr. Cray was pleased too. Mrs. Brown watered roses fertilized garden on June 13, 1880. Prof. Darren, Mrs. Brown & Mr. Cray met over dinner discussed how plan had worked. John Darren Mrs. Brown take down scarecrow following week. thought was too scary dog. Mr. Craye painted fence. Then Mrs. Brown watered lotus, carnations, daffodils, roses.

Then semantic network analysis was run on both texts using AutoMap's default settings. The black ellipse in the screen shot below shows how to run a Single Map Analysis (Semantic Network Analysis):

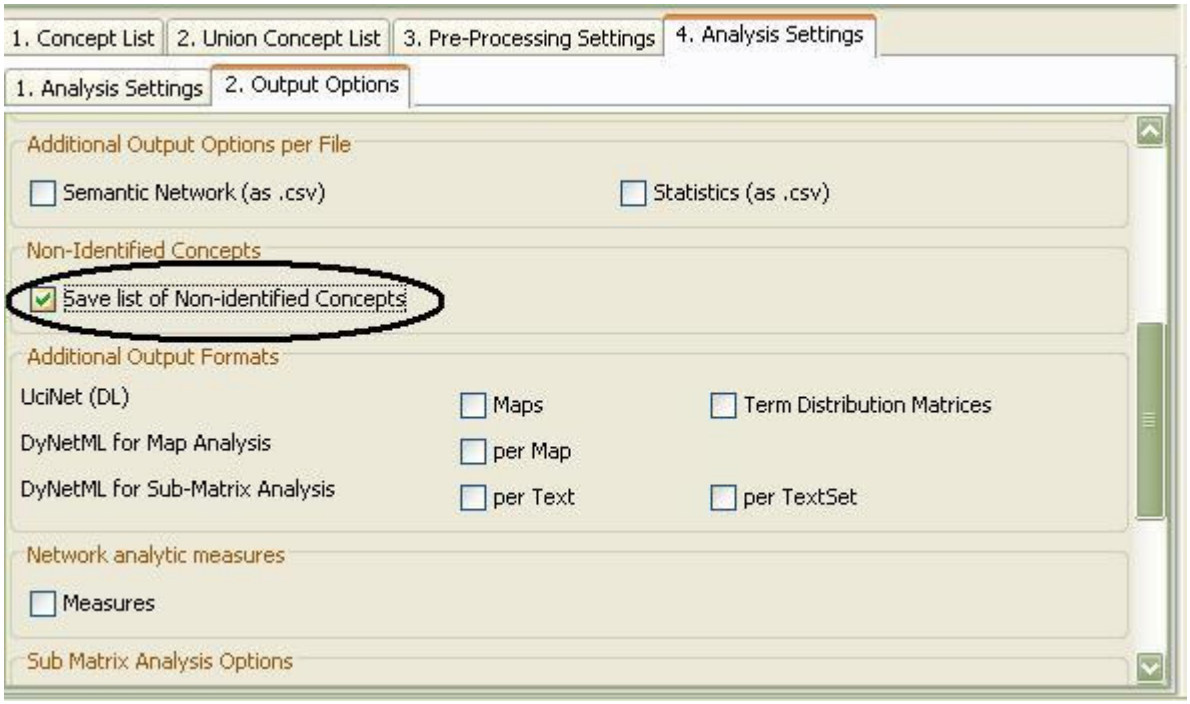


All outputs provided on the Output Options panel were checked. Again, the default settings were used. All output lists are saved in this folder (Term_Distribution_Lists_and_Matrices.zip).

2. Save Non-Identified concepts

Purpose: Save a list of all concepts that are remaining in the pre-processed texts and that are not:
- Denoted in a delete list
- Denoted in any of the thesauri

To create a list of these words, go to on **P2, tab no. 4 - Analysis Settings, tab no. 2 - Output Options** index card, select the **Non-Identified Concepts** field and check **Save list of Non-Identified Concepts**. Run any type of Analysis. The list will be stored under the root directory of AutoMap as Non_identified_concepts.csv. The black ellipse in the screen shot below shows where to find this option.



3. Additional Output Formats

Two additional data formats are offered:

- **DL** for e.g. UCINET (Borgatti, Everett & Freeman, 2002; for more information about UCINET see http://www.analytictech.com/ucinet.htm).
- **DyNetML** (Tsvetovat, Reminga & Carley, 2004; for more information about DyNetML see http://www.casos.cs.cmu.edu/projects/dynetml/).

Note: You can also user the Network Converter to convert data.

3.1 DL:

If maps and/or term distribution matrices are generated, these files can be additionally stored in the **UCINET DL** format.

Some points to consider when storing in UCINET DL Format:

- In order to do this check "Maps" and/ or "Term distribution matrices" items on the "Additional Output Formats" field on **P2, tab no. 10 Output Options** index card.
- The maps are stored with the extension .dl in the "DL_UciNet_Format" folder in the root directory of AutoMap.
- The term distribution matrices are stored with the extension .dl in the "DL_UciNet_Format" folder in the root directory of AutoMap.

- If you wish to perform UCINET analysis on your AutoMap results, import the data into UCINET. If you wish to visualize your AutoMap results, import the data into UCINET and use the visualization tools provided through UCINET.

Note: Since networks extracted with AutoMap are directed, matrices representing these networks can be rectangular. If a DL file needs to be generated from a rectangular matrix AutoMap by default squares this matrix before converting it into DL format.

3.2 DyNetML:
Results of Map Analysis and Sub-Matrix Text Analysis can be output in DyNetML format.

Purpose: Generate DyNetML representation of maps (mental models) generated with Map Analysis.
How to: Check the "per Map" checkbox on **P2, 10. Output Options** index card, Additional Output Formats, DyNetML for Map Analysis.
Output: The resulting DyNetML files will be stored as NameOfText.xml in the **xml** folder under the root directory of AutoMap.

After Map Analysis:
By default, all entities in the applied ontology are considered as nodes, and all statements between entities in the applied ontology (either in anterior or posterior or both positions) are represented as edges in DyNetML. The user is given the option to exclude entities of any applied ontology from being considered as nodes and thus forming statements, but are considered as attributes of other entities in the ontology that are forming nodes and link into statements. An example would be the category "attribute", which represent information that is inherent to a certain node. The concept "teacher" or "male" e.g. might be considered as attributes that relate to the entity agent. In order to use ontologies in a way ontologies that not only represent entities, but also features of entities. do this follow this procedure:

1. Use the drag and drop labeled **Concepts not forming edges** at the bottom of **P2, 10. Output Options** index card. Click **Refresh** to automatically generate a list of all entities in the applied ontology.
2. Use the **>** and **<** buttons to move entities from one window to the other.
3. Apply your decision by hitting the **Apply Selection** button.
4. Before running Sub-Matrix Text Analysis decide whether you want to perform "Network Text Analysis" (NTA) or "Social Network Text Analysis" (SNTA). To select a type use the toggle button at the bottom of **P2, 10. Output Options** index card. Both types are a form of Sub-Matrix Text Analysis, but differ in what they measure:
    1. NTA: If one measures textual network, then entities of the applied ontology that should not be considered as nodes, but as inherent

information of nodes, fill structural position and semantic function in a text. Thus they contribute to a texts' density, and therefore should be taken into consideration as statements for maps and stats. The resulting DyNetML file will contain nodes and inherent information on nodes (if any contained in the window that an entity and an attribute co-occur).

2. SNTA: If one measures a social network that is represented in or extracted from a text, then links from entities of the applied ontology that should not be considered as nodes, but as inherent information of nodes, to actual other nodes are not to be represented in the maps and stats, since this would over fit the social network. The resulting DyNetML file will NOT inherent information on nodes.

Note: Whether to use NTA or SNTA is a "text-philosophical question" and the answer depends upon what the user wants to measure: a textual network or a social network.

Run Sub-Matrix Text Analysis.

Two options for **creating DyNetML files** are offered. Either one or both options can be selected per analysis:
1. Create one DyNetML file per map and text. To do this, check **per Text** in the **Additional Output Formats** field on **P2, 10. Output Options** index card.
The DyNetML files will be stored as NameOfText.xml and in a folder called xml under the root directory of AutoMap.

2. Create one DyNetML file that unifies all maps. To do this, check **per TextSet** in the **Additional Output Formats** field on **P2, 10. Output Options** index card.
The DyNetML file will be stored as consolidated_map.xml in a folder called xml under the root directory of AutoMap.

3.3 Examples for additional Output Formats

Extracts from the Our Text I.txt and Our Text II.txt files were used as input:

Our Text 1.txt: Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.

Our Text II.txt: Mr. Cray's brown dog stopped eating the lotus blossom at 12 pm the next day. Mrs. Brown was now happy with the dog. She said "You good dog!" The dog no longer ate the flowers and weeds. Mr. Cray

was pleased too. Mrs. Brown watered the roses and fertilized the garden on June 13, 1880. Prof. Darren, Mrs. Brown & Mr. Cray met over dinner and discussed how the plan had worked. John Darren and Mrs. Brown would take down the scarecrow the following week. She thought it was too scary for the dog. Mr. Craye painted his fence. Then Mrs. Brown watered lotus, carnations, daffodils, and roses.

Then AutoMap's **Extensive Delete List, Direct Adjacency** was applied to both texts. The Delete List was extended by further non-content bearing words that appeared in the sample texts (a an and as at awhile but for from happening he her her hers him his i in into it its me mine my nor of or our she so that the their theirs them they to us was we were what who whoever whom whomever will would you your yours yourself). Resulting texts:

Our Text I.txt: Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown unhappy with dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12, 1880. Weeding no longer needed.

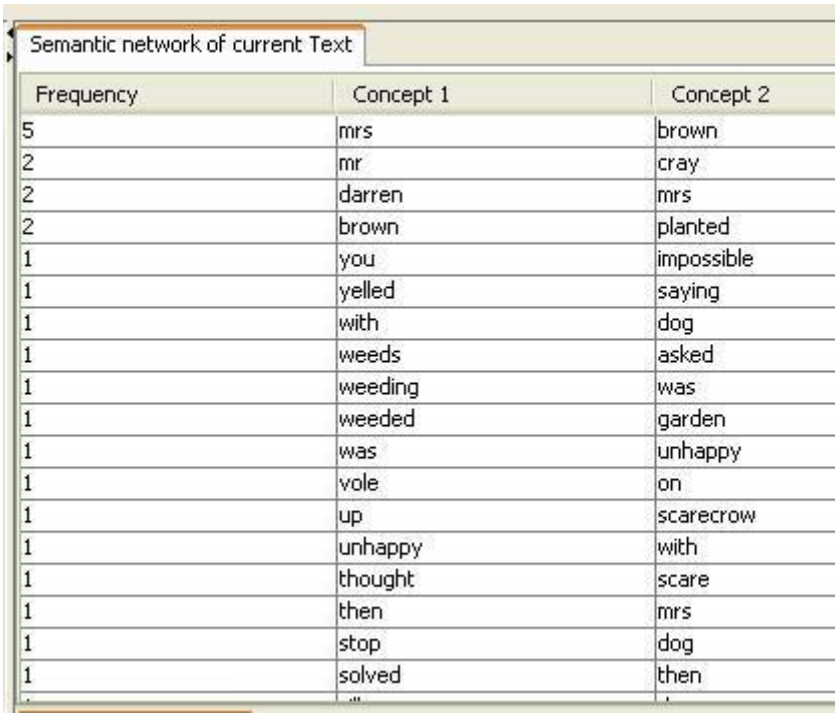
Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses.

Our Text II.txt: Mr. Cray's brown dog stopped eating lotus blossom 12 pm next day. Mrs. Brown now happy with dog. said "You good dog!" dog no longer ate flowers weeds. Mr. Cray pleased too.

Mrs. Brown watered roses fertilized garden on June 13, 1880.

Prof. Darren, Mrs. Brown & Mr. Cray met over dinner discussed how plan had worked. John Darren Mrs. Brown take down scarecrow following week. thought too scary dog. Mr. Craye painted fence. Then Mrs. Brown watered lotus, carnations, daffodils, roses.

Then map analysis was run on both texts using the AutoMap's default settings. All term distribution matrices provided on the Output Options panel were checked. Again, the default settings were used. The screen shot below shows the resulting analysis that should be displayed in **P3**.



Semantic network of current Text

Frequency	Concept 1	Concept 2
5	mrs	brown
2	mr	cray
2	darren	mrs
2	brown	planted
1	you	impossible
1	yelled	saying
1	with	dog
1	weeds	asked
1	weeding	was
1	weeded	garden
1	was	unhappy
1	vole	on
1	up	scarecrow
1	unhappy	with
1	thought	scare
1	then	mrs
1	stop	dog
1	solved	then

4. Network Analytic Measures

AutoMap supports the computation of *network analytic measures* per map or network and per concept or node.

Note: Only Network Analytic Measures for directed networks were implemented into AutoMap. The reason for this is that AutoMap outputs are always directed in order to adequately represent the linear structure of texts.

To create measures follow these steps:

1. On **P2, tab no. 10 Output Options**, select the **Network analytic measures** field and check **Measures**.
2. After analysis is run, measures will be displayed on **P4, 2. Network analytic measures** index card.
   A file that collects all measures per text set will be saved as overall_network_analytix_measures.csv in the results folder under the root directory of AutoMap.

Note: Texts have a linear structure. Therefore, we only implemented Network Analytic Measures for directed networks (digraphs) into AutoMap.

The following Table explains the measures that can be computed:

Name of measure	**Calculation of measure**	**Name of and reference for corresponding**	**Corresponding name of**

		SNA measure	measure in Cube
Concept (node) level measures, direct connectivity			
Concept outdegree	Total number of statements with concept in anterior position	Outdegree, expansiveness, actor degree centrality (Wasserman & Faust 1994); Prestige, Influence (Mrvar)	Local imageability
Concept indegree	Total number of statements with concept in posterior position	Indegree, receptivity, popularity, actor degree prestige (Wasserman & Faust 1994); Prestige, Support (Mrvar)	Local evokability
Concept outdegree centrality	Total number of statements with concept in anterior position/ Number of unique concepts in text	Outdegree Centrality (Wasserman & Faust 1994); Relative Influence (Mrvar)	None
Concept indegree centrality	Total number of statements with concept in posterior position/ Number of unique concepts per text	Indegree Centrality (Wasserman & Faust 1994); Relative Support (Mrvar)	None
Total degree	Concept indegree + concept outdegree	-	Local density
Map (graph) level measures, direct connectivity			
Text outdegree centrality	Sum (largest observed outdegree – outdegree of concepts)/(Number of unique concepts in text) 2	Group outdegree centralization (Wasserman & Faust, 1994)	None
Mean concept outdegree centrality	Sum (outdegree)/ Number of unique concepts in text	Mean outdegree (= Mean indegree) (Wasserman & Faust 1994)	None
Variance of concept outdegree centrality	Sum(sum outdegree – mean outdegree) 2 / Number of unique concepts in text	Variance of outdegree (Wasserman & Faust 1994, p.127-128)	None
Text indegree	Sum (largest observed indegree – indegree of	Group indegree	None

centrality	concepts)/(Number of unique concepts in text) 2	centralization (Wasserman & Faust, 1994)	
Mean concept indegree centrality	Sum (indegree)/ Number of unique concepts in text	Mean indegree (=Mean outdegree)	None
Variance of concept indegree centrality	Sum (sum indegree – mean indegree) 2 / Number of unique concepts in text	Variance of indegree (Wasserman & Faust 1994, p.127-128)	None
Density	Number of statements/ Possible number of statements	Density (Wasserman & Faust 1994, p.129, Scott 1991, p.74) Wasserman and Faust use g(g-1) as denominator:, we use (g*g), because unique concept can form statement with same unique concept (e.g. agent-agent)	None
Concept (node) level measures, indirect connectivity			
Concept closeness centrality	Minimum possible total distance from node i to all other nodes/ Sum of all geodesics between node i and all other nodes	Closeness (Wasserman & Faust, Mrvar) According to Wasserman and Faust (1994, p.200) group level closeness centrality is not computed	None
Concept betweenness centrality	sum ((Number of all geodesics between all nodes that go through node i)/(Number of geodesics between node i and all other nodes))/((Number of unique concepts in text -1)(Number of unique concepts in text -2))	Betweenness (Gould 1987, Mrvar)	None

Concept proximity prestige	Number of concepts directly or indirectly adjacent to node i	Proximity Prestige (Wasserman & Faust 1994, Mrvar)	None
Map (graph) level measures, indirect connectivity			
Text Proximity Prestige	Sum (Proximity Prestige (all unique concepts in text)/ Number of unique concepts in text	Group level proximity prestige (Wasserman & Faust 1994)	None

---

Analyses

1. Semantic Network Analysis

Semantic Network Analysis can be run on original Input text(s) or texts that have been pre-processed with a Delete List and/ or a Generalization thesaurus.
Before you run map analysis make sure that you have completed the following steps:

- Pre-processed the texts if you wish to.
- Specified the analysis settings. If you do not do so, the default setting default settings will be applied.
- Specified additional output options if you wish to. If you do not select additional output options, the default outputs default outputs will be generated.

If you wish to analyze a single text, click the **Run Analysis menu** and select **Single Map Analysis.**
If you wish to analyze a set of texts, click the **Run Analysis menu** and select **Multiple Map Analysis**.

The Results will be displayed on **P3** on the **Map** index card and on **P4** on the **Stat** index card.
If you had requested additional outputs, those will be generated and stored under the directories specified under the Section Additional Outputs.

Other Semantic Network Analysis Points to Consider

- If you have analyzed multiple texts, you can browse through the results and see the related texts.
  Results of multiple analyses are automatically saved in folder named “output” under directory where AutoMap 2.0 is installed.

- This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as a stat_output.xls file that contains the stats of all texts.
- The “Stat Output” folder is overwritten with every new analysis you run. So if you want to save the results of a current “Stat Output” folder just rename the folder.

1.1 Example for Semantic Network Analysis

An extract from the Denmark text was used as input:

```
Reporters said hundreds of people emerged from shops in Copenhagen city
centre to see what was happening, and used their mobile phones to
contact their families.
```

The text was pre-processed with AutoMap's customized extensive Delete List. These are the resulting texts:

Input text	Tool used	Setting	Resulting text
Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying "You impossible dog!" But the dog kept eating the flowers and weeds. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted roses and weeded the garden. The	AutoMap's customized extensive Delete List: a an and as at but for he her her hers him his i it its me mine my nor of or our she	Direct Adjacency (default)	Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy with dog. yelled saying "You impossible dog!" dog kept eating flowers weeds. asked Mr. Cray stop dog. couldn't. Mrs. Brown planted roses weeded garden. silly dog % dug up roses looking vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met next day concoct plan. John Darren Mrs. Brown put up scarecrow. thought scare dog. Mr. Craye put up fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses..

silly dog % dug up the roses looking for a vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan. John Darren and Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, and roses.	so that the their theirs them they to us we who whoever whom whomever will would you your yours yourself	Rhetorical Adjacency	Mr. Cray's brown dog ate xxx lotus blossom xxx 10 am. Mrs. Brown was unhappy with xxx dog. xxx yelled xxx xxx saying "You impossible dog!" xxx xxx dog kept eating xxx flowers xxx weeds. xxx asked Mr. Cray xxx stop xxx dog. xxx couldn't. Mrs. Brown planted roses xxx weeded xxx garden. xxx silly dog % dug up xxx roses looking xxx xxx vole on June 12, 1880. Weeding was no longer needed. Prof. Darren, Mrs. Brown & Mr. Cray met xxx next day xxx concoct xxx plan. John Darren xxx Mrs. Brown put up xxx scarecrow. xxx thought xxx xxx scare xxx dog. Mr. Craye put up xxx fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, xxx roses.

Next we run Map Analysis on both texts using AutoMap's default Analysis Settings.

These are the Map and Statistics outputs for the first text (direct adjacency):	These are the Map and Statistics outputs for the second text (rhetorical adjacency):

Map:

5	mrs
2	mr
2	darren
2	brown
1	you
1	worked
1	with
1	week
1	weeds
1	watered
1	was
1	too
1	thought
1	then
1	take
1	stopped
1	scary
1	scarecrow
1	said
1	roses
1	prof
1	pm
1	pleased
1	plan
1	painted
1	over
1	on
1	now
1	no
1	next
1	mr
1	met
1	lotus
1	longer
1	june
1	john
1	how
1	happy
1	had
1	good
1	garden

Map:

2	mr
2	brown
1	you
1	worked
1	weeds
1	watered
1	was
1	too
1	then
1	take
1	stopped
1	said
1	prof
1	pleased
1	plan
1	over
1	on
1	now
1	no
1	next
1	mr
1	met
1	lotus
1	longer
1	june
1	john
1	happy
1	had
1	good
1	garden
1	following
1	fence
1	dog
1	discussed
1	day
1	darren
1	craye
1	cray's
1	cray
1	carnations
1	brown
1	1880
1	13

	1 12

```
Stat:

# of concepts analyzed:
unique: 62
total: 88

# of concepts in
statements:
unique: 54
total: 60

# of isolated concepts:
unique: 12
total: 13

# of statements:
unique: 54
total: 60

Density (based on
Statements):
unique: 0.87
total: 0.97
```

```
Stat:

# of concepts analyzed:
unique: 60
total: 84

# of concepts in statements:
unique: 55
total: 61

# of isolated concepts:
unique: 6
total: 7

# of statements:
unique: 55
total: 61

Density (based on
Statements):
unique: 0.92
total: 1.02
```

Note: For more information about the impact of coding choices on map analysis results you might have a look at our publications (http://www.casos.cs.cmu.edu/projects/automap/publications.html).

2. Meta Matrix Text Analysis

Meta Matrix Text Analysis can be run on a text or a set of texts that were pre-processed with a Meta-Matrix Thesaurus. It enables the classification and analysis of concepts in texts according to the Meta-Matrix model ontology and categories of the resulting inter and intra-related sub-matrices (Diesner & Carley, 2005), Meta-Matrix Text Analysis and the social systems represented in texts. Meta-matrix based analysis of properties of social systems by investigating the inter and intra-connections between the matrices contained in the meta-matrix (cells in Table 1) can provide insight into the complex structure of social systems.

The Meta-Matrix Model used in AutoMap:

Meta-Matrix	Agent	Knowledge	Resource	Task/ Event	Organization	Location	Action	Role	Attribute
Agent	Social network	Knowledge network	Capability network	Assignment network	Membership network	Agent location network			
Knowledge		Information	Training network	Knowledge	Organizational knowledge	Knowledge location			

		network		requirement network	network	network			
Resource			Resource network	Resource requirement network	Organizational capability network	Resource location network			
Task/ Event				Precedence network	Organizational assignment network	Task/Event location network			
Organization					Interorganizational network	Organizational location network			
Location						Proximity network			
Action									
Role									
Attribute									

Before you run Meta-Matrix analysis make sure that you have completed the following actions:

- Pre-processed the text(s) with a Meta-Matrix Thesaurus. If you also want to use a Delete List and/ or a Generalization Thesaurus make sure to use these tools before you use the Meta-Matrix Thesaurus.
- Specified the analysis settings. If you do not do so, the default setting default settings will be applied.
- Specified additional output options if you wish to. If you do not select additional output options, the default outputs default outputs will be generated.

Analyzing Text(s)

If you wish to analyze a single text, click the **Run Analysis menu** and select **Single Meta Matrix Text Analysis.**

If you wish to analyze a set of texts, click the **Run Analysis menu** and select **Multiple Meta Matrix Text Analysis**.

The Results will be displayed on P3 on the **Map** tab and on P4 on the **Stat** tab respectively. If you had requested additional outputs, those will be generated and stored under the directories specified under the Section Additional Outputs.

If you have analyzed multiple texts, you can browse through the results and see the related texts.
Results of multiple analyses are automatically saved in a folder named “output” under the directory where AutoMap 2.0 is installed. This output folder contains a

map file (*nameOfText.map*) and a stat file (*nameOfText.stat*) for each text analyzed as well as *stat_output.xls file* a that contains the stats of all texts. The “Stat Output” folder is overwritten with every new analysis you run. So if you want to save the results of a current “Stat Output” folder just rename the folder.

2.1 Example for Meta Matrix Text Analysis

An extract from the Our Text I.txt was used as input:

```
Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was
unhappy with the dog. She yelled at it saying "You impossible dog!" But
the dog kept eating the flowers and weeds. She asked Mr. Cray to stop
the dog. He couldn't. Mrs. Brown planted roses and weeded the garden.
The silly dog % dug up the roses looking for a vole on June 12, 1880.
Weeding was no longer needed.

Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan.
John Darren and Mrs. Brown put up a scarecrow. She thought it would
scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs.
Brown planted lotus, carnations, daffodils, and roses.
```

The text was pre-processed with the Meta-Matrix-Thesaurus. This are the resulting texts:

Setting	Resulting text
Not Selected - Thesaurus content only (default)	mr . < agent> < agent> dog ate lotus < event> < attribute> am . mrs . < agent> was unhappy with dog . yelled saying < agent resource> impossible dog!" dog kept eating flowers weeds . asked mr . cray stop dog . couldn't . mrs . < agent> planted roses weeded garden . silly dog < attribute> dug up roses looking vole on june < attribute> , < attribute> . weeding was no longer needed . prof . < agent> , mrs . < agent> < attribute> mr . cray met next day < task> plan . john < agent> mrs . < agent> put up scarecrow . thought scare dog . mr . < agent> put up fence . problem solved . then mrs . < agent> planted lotus , < resource> , < resource> , roses .
Thesaurus content only, Direct Adjacency (default if Thesaurus content only is chosen)	. < agent> < agent> < event> < attribute> . . < agent> . < agent resource> . . . . . < agent> . < attribute> < attribute> < attribute> . . . . < agent> . < agent> < attribute> . < task> . < agent> . < agent> . . . <

	agent> . . . < agent> < resource> < resource> .
Thesaurus content only, Rhetorical Adjacency	xxx . < agent> < agent> xxx xxx xxx < event> < attribute> xxx . xxx . < agent> xxx xxx xxx xxx . xxx xxx < agent resource> xxx xxx xxx xxx xxx xxx xxx . xxx xxx . xxx xxx xxx . xxx . xxx . < agent> xxx xxx xxx xxx . xxx xxx < attribute> xxx xxx xxx xxx xxx xxx xxx < attribute> < attribute> . xxx xxx xxx xxx xxx . xxx xxx xxx . < agent> xxx . < agent> < attribute> xxx . xxx xxx xxx xxx < task> xxx . xxx < agent> xxx . < agent> xxx xxx xxx . xxx xxx xxx . xxx . < agent> xxx xxx . xxx xxx . xxx xxx . < agent> xxx xxx < resource> < resource> xxx .

Then we run Map Analysis on both texts using AutoMap's default Analysis Settings. These are the results:

	Not Selected - Thesaurus content only (default)	Thesaurus content only, Direct Adjacency	Thesaurus content only, Rhetorical Adjacency
Map	1 10 am 1 12 1880 1 1880 weeding 1 am mrs 1 asked mr 1 ate lotus 1 blossom 10 1 brown was 1 carnations daffodils 1 concoct plan 1 couldn't mrs 1 cray stop 1 cray's brown 1 craye put up 1 daffodils roses 1 day concoct 1 dog yelled 1 dug up 1 eating flowers 1 fence problem 1 flowers weeds 1 garden silly 1 impossible dog 1 john darren	1 10 am 1 12 1880 1 1880 weeding 1 am mrs 1 asked mr 1 ate lotus 1 blossom 10 1 brown was 1 carnations daffodils 1 concoct plan 1 couldn't mrs 1 cray stop 1 cray's brown 1 craye put up 1 daffodils roses 1 day concoct 1 dog yelled 1 dug up 1 eating flowers 1 fence problem 1 flowers weeds 1 garden silly 1 impossible	1 blossom 10 1 brown was 1 carnations daffodils 1 concoct plan 1 couldn't mrs 1 cray stop 1 cray's brown 1 craye put up 1 daffodils roses 1 day concoct 1 dog yelled 1 dug up 1 eating flowers 1 fence problem 1 flowers weeds 1 garden silly 1 impossible dog 1 john darren 1 june 12 1 kept eating 1 longer needed 1 looking vole 1 lotus carnations

	1 june 12 1 kept eating 1 longer needed 1 looking vole 1 lotus carnations 1 met next 1 mr craye 1 needed prof 1 next day 1 no longer 1 on june 1 plan john 1 planted roses 1 problem solved 1 prof darren 1 put up 1 put up fence 1 roses weeded 1 saying you 1 scare dog 1 scarecrow thought 1 silly dog 1 solved then 1 stop dog 1 then mrs 1 thought scare 1 unhappy with 1 up scarecrow 1 vole on 1 was unhappy 1 weeded garden 1 weeding was 1 weeds asked 1 with dog 1 yelled saying 1 you impossible 2 brown planted 2 darren mrs 2 mr cray	dog 1 john darren 1 june 12 1 kept eating 1 longer needed 1 looking vole 1 lotus carnations 1 met next 1 mr craye 1 needed prof 1 next day 1 no longer 1 on june 1 plan john 1 planted roses 1 problem solved 1 prof darren 1 put up 1 put up fence 1 roses weeded 1 saying you 1 scare dog 1 scarecrow thought 1 silly dog 1 solved then 1 stop dog 1 then mrs 1 thought scare 1 unhappy with 1 up scarecrow 1 vole on 1 was unhappy 1 weeded garden 1 weeding was 1 weeds asked 1 with dog 1 yelled saying 1 you impossible 2 brown planted 2 darren mrs 2 mr cray 5 mrs brown	1 met next 1 mr craye 1 needed prof 1 next day 1 no longer 1 on june 1 plan john 1 planted roses 1 problem solved 1 prof darren 1 put up 1 put up fence 1 roses weeded 1 saying you 1 scare dog 1 scarecrow thought 1 silly dog 1 solved then 1 stop dog 1 then mrs 1 thought scare 1 unhappy with 1 up scarecrow 1 vole on 1 was unhappy 1 weeded garden 1 weeding was 1 weeds asked 1 with dog 1 yelled saying 1 you impossible 2 brown planted 2 darren mrs 2 mr cray 5 mrs brown
Stat	File: # of concepts analyzed:	File: # of concepts analyzed: unique: 62	File: # of concepts analyzed: unique: 62

	unique: 62 total: 88 # of concepts in statements: unique: 80 total: 87 # of isolated concepts: unique: 0 total: 0 # of statements: unique: 80 total: 87 Density (based on Statements): unique: 1.29 total: 1.4	total: 88 # of concepts in statements: unique: 80 total: 87 # of isolated concepts: unique: 0 total: 0 # of statements: unique: 80 total: 87 Density (based on Statements): unique: 1.29 total: 1.4	total: 88 # of concepts in statements: unique: 80 total: 87 # of isolated concepts: unique: 0 total: 0 # of statements: unique: 80 total: 87 Density (based on Statements): unique: 1.29 total: 1.4

Note: For more information about the impact of coding choices on map analysis results please visit us on the web (http://www.casos.cs.cmu.edu/projects/automap/publications.html).

3. Sub Matrix Text Analysis

Sub Matrix Text Analysis distills one or several sub-networks from the meta-matrix network and retranslates the meta-matrix entities into the text-level concepts that represent these Meta-Matrix categories. This routine enables a more thorough analysis of particular cells of the meta-matrix (Diesner & Carley, 2004c). Sub Matrix Text Analysis can be run on a text or a set of texts that were pre-processed with a Meta-Matrix Thesaurus and from that Sub-Matrices were selected.

Before you run Sub-Matrix analysis make sure that you have completed the following actions:

- Pre-processed the text(s) with a Meta-Matrix Thesaurus. If you also want to use a Delete List and/ or a Generalization Thesaurus make sure to use these tools before you use the Meta-Matrix Thesaurus.
- Selected Sub-Matrix Categories.
- Specified the analysis settings. If you do not do so, the default setting default settings will be applied.
- Specified additional output options if you wish to. If you do not select additional output options, the default outputs default outputs will be generated.

The user is given the option to exclude entities of any applied ontology from being considered as nodes and thus forming statements, but are considered as attributes of other entities in the ontology that are forming nodes and link into statements. An example would be the category "attribute", which would represent information that is inherent to a certain node. The concept "teacher" or "male" might be considered as attributes that relate to the entity agent.

To use ontologies (in a way ontologies not only represent entities, but also features of entities) follow these steps:

1. Use the drag and drop labeled **Concepts not forming edges** at the bottom of **P2, tab no 10. Output Options**. Click **Refresh** to automatically generate a list of all entities in the applied ontology.
2. Use the **>** and **<** buttons to move entities from one window to the other.
3. Apply your decision by hitting the **Apply Selection** button.
4. Before running Sub-Matrix Text Analysis decide whether you want to perform "Network Text Analysis" (NTA) or "Social Network Text Analysis" (SNTA). To select a type use the toggle button at the bottom of **P2, 10. Output Options** index card. Both types are a form of Sub-Matrix Text Analysis, but differ in what they measure:
    1. NTA: If one measures textual network, then entities of the applied ontology that should not be considered as nodes, but as inherent information of nodes, fill structural position and semantic function in a text. Thus they contribute to a texts' density, and therefore should be taken into consideration as statements for maps and stats. The resulting DyNetML file will contain nodes and inherent information on nodes (if any contained in the window that an entity and an attribute co-occur).
    2. SNTA: If one measures a social network that is represented in or extracted from a text, then links from entities of the applied ontology that should not be considered as nodes, but as inherent information of nodes, to actual other nodes are not to be represented in the maps and stats, since this would over fit the social network. The resulting DyNetML file will NOT inherent information on nodes.

       Whether to use NTA or SNTA is a "text-philosophical question" that's answer depends upon what the user wants to measure - a textual network or a social network.

5. Run Sub-Matrix Text Analysis.

If you wish to analyze a single text, click the **Run Analysis menu** and select **Single Meta-Matrix analysis.**
If you wish to analyze a set of texts, click the **Run Analysis menu** and select **Multiple Meta-Matrix Analysis**.

The Results will be displayed on **P3** on the **Map** index card and on **P4** on the **Stat** index card.
If you had requested additional outputs, those will be generated and stored under the directories specified under the Section Additional Outputs.

If you have analyzed multiple texts, you can browse through the results and see the related texts.
Results of multiple analysis are automatically saved in a folder named “output” under the directory where AutoMap 2.0 is installed. This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as stat_output.xls file a that contains the stats of all texts.
The “Stat Output” folder is overwritten with every new analysis you run. So if you want to save the results of a current “Stat Output” folder just rename the folder.

3.1 Example for Sub Matrix Text Analysis

This example is based on the example for Sub Matrix Selection.

Our Text I.txt. was used as input:

```
Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was
unhappy with the dog. She yelled at it saying "You impossible dog!" But
the dog kept eating the flowers and weeds. She asked Mr. Cray to stop
the dog. He couldn't. Mrs. Brown planted roses and weeded the garden.
The silly dog % dug up the roses looking for a vole on June 12, 1880.
Weeding was no longer needed.

Prof. Darren, Mrs. Brown & Mr. Cray met the next day to concoct a plan.
John Darren and Mrs. Brown put up a scarecrow. She thought it would
scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs.
Brown planted lotus, carnations, daffodils, and roses.
```

The text was pre-processed with the Meta-Matrix-Thesaurus. Then we selected the full meta-matrix. These are the resulting texts:

Setting	Resulting text
Select: Knowledge/Agent/Resource	. dog lotus.. dog. dog flowers.. dog... roses. dog roses,.. .,..... dog..... lotus, carnations, daffodils, roses.
Select: Knowledge/Agent/Resource	xxx. xxx xxx dog xxx xxx lotus xxx xxx xxx xxx. xxx. xxx xxx

	xxx xxx xxx dog. xxx xxx xxx xxx xxx xxx xxx xxx xxx dog xxx xxx xxx flowers xxx xxx. xxx xxx xxx. xxx xxx xxx xxx dog. xxx xxx. xxx. xxx xxx roses xxx xxx xxx xxx. xxx xxx dog xxx xxx xxx xxx roses xxx xxx xxx xxx xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx xxx xxx. xxx, xxx. xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx. xxx xxx xxx xxx. xxx xxx xxx xxx xxx. xxx xxx xxx xxx. xxx xxx xxx xxx dog. xxx. xxx xxx xxx xxx. xxx xxx. xxx xxx. xxx xxx lotus, carnations, daffodils, xxx roses. xxx

Then we run Map Analysis on both texts using AutoMap's default Analysis Settings. These are the results:

Map	1 lotus carnations 1 met the 1 mr craye 1 neededprof darren 1 next day 1 no longer 1 on june 1 plan john 1 planted roses 1 problem solved 1 put up 1 put up a 1 roses looking 1 saying you 1 scare the 1 scarecrow she 1 she yelled 1 silly dog 1 solved then 1 stop the 1 the silly 1 then mrs 1 thought it 1 to stop 1 unhappy with 1 up the 1 vole on 1 was unhappy 1 weeded the 1 weeding was 1 weeds she 1 with the	1 12 pm 1 13 1880 1 1880 prof 1 and weeds 1 at 12 1 ate the 1 blossom at 1 brown would 1 carnations daffodils 1 cray was 1 cray's brown 1 craye painted 1 daffodils and 1 darren mrs 1 day mrs 1 dinner and 1 discussed how 1 dog the 1 down the 1 eating the 1 fence then 1 fertilized the 1 flowers and 1 following week 1 for the 1 garden on 1 good dog 1 had worked 1 happy with 1 his fence 1 how the 1 it was

	1 would scare 1 yelled at 1 you impossible 2 brown planted 2 mr cray 4 the dog 5 mrs brown	1 john darren 1 june 13 1 longer ate 1 lotus carnations 1 met over 1 mr craye 1 next day 1 no longer 1 now happy 1 on june 1 over dinner 1 painted his 1 plan had 1 pleased too 1 pm the 1 prof darren 1 roses and 1 said you 1 scarecrow the 1 scary for 1 she thought 1 stopped eating 1 take down 1 the scarecrow 1 then mrs 1 thought it 1 too scary 1 was too 1 watered the 1 weeds mr 1 week she 1 with the 1 worked john 1 would take 1 you good 2 brown watered 2 mr cray 3 the dog 5 mrs brown
Stat	# of concepts analyzed: unique: 72 total: 118 # of concepts in statements: unique: 108 total: 117 # of isolated concepts: unique: 0 total: 0	# of concepts analyzed: unique: 68 total: 107 # of concepts in statements: unique: 98 total: 106 # of isolated concepts: unique: 0

	# of statements: unique: 108 total: 117 Density (based on Statements): unique: 1.5 total: 1.63	total: 0 # of statements: unique: 98 total: 106 Density (based on Statements): unique: 1.44 total: 1.56

Outputs

Outputs for Map Analysis, Meta Matrix Text Analysis and Sub Matrix Text Analysis are displayed on P3 on the **Semantic Network** index card and on P4 on the **Stat** index card.
If you have analyzed multiple texts, you can browse through the results and see the related texts.
Results of multiple analyses are automatically saved in a folder named “output” under the directory where AutoMap 2.0 is installed. This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as a stat_output.xls file that contains the stats of all texts.
Additional Outputs will be generated if requested by the user.

1. Semantic Network

After running analysis, the semantic network will be displayed on **P3** on the Semantic **Network** Index card.
The semantic network contains one coded statement per line.
If the Frequency item was checked on the Analysis Settings index card the first column of the semantic network indicates the frequency of every displayed statement.
Each semantic network generated is automatically saved in a folder named “Stat Output” under the directory where AutoMap 1.2 is installed. This output folder contains a semantic network file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as a stat_output.xls file that contains the stats of all texts.

1.1 Example for Semantic Network Output

For examples for semantic network files, see the examples for analysis, the **Semantic Network of current text** index cards.

Semantic networks are output as .CSV (.map in previous versions).

2. Statistics

After running the analysis, the Stat file will be displayed on **P4** on the **Statistics** Index Card.
Each stat file generated is automatically saved in a folder named “Stat Output” under the directory where AutoMap 1.2 is installed. This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as stat_output.xls file a that contains the stats of all texts.

Entries in the stat output and explanation:

Entry	Entry	Explanation
File:		Name of the analyzed text file.
# of concepts analyzed:	unique:	Unique concepts are those that appear only once in a text; the number of total concepts includes those that appear more than once in a given text. All concepts are considered that occurred in the texts that were analyzed.
	total:	
# of concepts in statements:	unique:	Only concepts are considered that linked into statements.
	total:	
# of isolated concepts:	unique:	Only concepts are considered that did not link into statements.
	total:	
# of statements:	unique:	Unique statements are those that appear only once in a text; the number of total statements includes those that appear more than once in a given text.
	total:	
density (based on statements)	unique:	Unique density is the density of the resulting network based on unique statements, total density respectively is the density of the resulting network based on the total number of statements.
	total:	
Analysis Settings	Punctuation:	The Punctuation option chosen by the user.
	Window Size:	The Window Size chosen by the user.
	Directionality:	The Directionality option chosen by the user.

2.1 Example for Statistic Output

For examples of stat files, see the examples for analysis, the **Stat of current text** index cards.

3. Additional Outputs

The map and the statistic output generated by AutoMap are displayed on **P3** and **P4**, respectively.

For all types of multiple analysis a lot more outputs can be generated on demand.

AutoMap Usage "Snapshots"

1. Snapshot: Split Input Text Files

Purpose

Split large text files into smaller ones of minimum equal size.
When to apply it: In order to speed up AutoMap coding.
Input from user: Number of words (NW) that each text file should contain after splitting.

How it works

Each text will be split at the next sentence mark after the number of words that the user had specified. Thus, each resulting split text will contain at least NW words.

Output

N texts that contain at least NW in directory specified by the user. The resulting texts maintain the original filename plus a counter, starting from 0 and going up to N, where N indicates the largest number of texts that an original text had been split up into.

How To

Click the **Tools** menu and select **Open Text File Splitter.** Follow the directions specified in the user interface.

2. Snapshot: Using Compare Maps

Purpose

- Compare individual mental maps and structures of social and organizational systems extracted from texts with AutoMap by using set theory
- Consolidate individual mental maps and structures of social and organizational systems extracted from texts with AutoMap to a degree specified by the user.

How To

Click the **Tools** menu and select **Open CompareMap.** For further instructions consult the CompareMap User's Guide.

3. Snapshot: Merge DyNetML Files

Purpose

Merge multiple DyNetML files into 1 DyNetML file.

Example

This might be needed for example when DyNetML files that was generated per text during Sub-Matrix Text Analysis need to be consolidated into 1 DyNetML file that represent the entire text set.

Output

1 DyNetML file.

How To

Click the **Tools** menu and select **Open DyNetML File Merger.** Follow the directions specified in the user interface.

4. Snapshot: Convert Network Data Formats

Purpose

Convert a file in a specific network data format (CVS, DL, UCINET, DyNetML, VNA) into another network data format.

How To

Click the **Tools** menu and select **Open Matrix Editor**. Follow the directions specified in the user interface.

5. Snapshot: Edit Network Data

Purpose

Edit relational data.

How To

Click the **Tools** menu and select **Open Network Data Format Converter**. Follow the directions specified in the user interface.

6. Snapshot: Visualize Semantic Networks

Purpose

Visualize mental models and social structure.

How To

- Within AutoMap: click the **Tools** menu and select **Open SocialInsight Visualizer**

  Example: Load in DyNetML files created in AutoMap.

- With External Tools: Convert DyNetML or DL files generated with AutoMap in formats required by the external tool you want to use:

  Example: Convert DL into VNA files can be visualized in NetDraw. In NetDraw open VNA file: File > Open > VNA text file > complete.

Clear

There are two ways to clear all index cards

5. To clear all index cards, click the **File menu** and select **Clear Text-Preprocessing Panel.** As a result, **P1** and, if analyses were run, **P3** and **P4** will be cleared. Use this functionality if you want to run analysis on another (set of) text(s) than the one currently loaded.
6. To clear a single index card, Un-Apply the related pre-processing strategy. **P2** cannot be cleared, but pre-processing tools on P2 can be unapplied. To do this, hit the **Un-Apply** button on the index cards on **P2**. If you want to run analysis with another (set of) text(s) you can reuse the pre-processing tools you probably have already built or edited.

Exit

To exit AutoMap, click the **File menu** and select **Exit**.
AutoMap will be closed.

How to cite AutoMap

Carley, K.M., & Diesner, J. (2005). *AutoMap: Software for Network Text Analysis*.

Further Information

For further information on AutoMap please visit:
http://www.casos.cs.cmu.edu/projects/automap

On this **web page** you will find:

- Publications about AutoMap, Network Text Analysis, and Map Analysis
- Installers for AutoMap
- Contact information
- Information about sponsors of AutoMap
- Information about people working on AutoMap

We also provide a online discussion forum for AutoMap:
- to discuss questions related to the AutoMap software;
- and get help from other AutoMap users and the developers of the software with using the tool.

Questions, Bugs, and Comments

Please contact us:
Dr. Kathleen M. Carley (kathleen.carley@cmu.edu)
Jana Diesner (diesner@cs.cmu.edu)

Technical Writer, Matt De Reno (mjdereno@andrew.cmu.edu)

Carnegie Mellon University
School of Computer Science
Institute for Software Research International (ISRI)
Center for Computational Analysis of Social and Organizational Systems (CASOS)
5000 Forbes Avenue
1325 Wean Hall
Pittsburgh, PA, 15213

We provide a online discussion forum for AutoMap:

- To discuss questions related to the AutoMap software.
- And get help from other AutoMap users and the developers of the software with using the tool.

References

Borgatti, S.P., Everett , M.G., & Freeman, L.C. (ac) (2002). *UCINET. for Windows. Software for Social Network Analysis*. Analytic Technologies, Inc.

Burkart, M. (1997). Thesaurus. In M. Buder, W. Rehfeld, T. Seeger & D. Strauch (Eds.), *Grundlagen der praktischen Information und Dokumentation: Ein Handbuch zur Einführung in die fachliche Informationsarbeit*, (pp. 160 - 179). 4th edition. München: Saur.

Kathleen M. Carley, 1997, "Extracting Team Mental Models Through Textual Analysis." Journal of Organizational Behavior, 18: 533-538.

Kathleen M.Carley, 1994, "Extracting Culture through Textual Analysis." Poetics, 22: 291-312.

Kathleen M.Carley and David Kaufer, 1993, "Semantic Connectivity: An Approach for Analyzing Semantic Networks." Communication Theory, 3(3): 183-213.

Kathleen M.Carley, 1993, "Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis." In Marsden P. (Ed), Sociological Methodology, 23: 75-126. Oxford: Blackwell.

Kathleen M. Carley, 1997, "Network Text Analysis: The Network Position of Concepts." Chapter 4 in C. Roberts (Ed.), Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 79-100.

Carley, K.M. & Reminga, J. (2004). *ORA: Organization Risk Analyzer*. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report.

Carley, K.M. (2003). Dynamic Network Analysis. In R. Breiger, K.M. Carley & P. Pattison (Eds.), *Summary of the NRC workshop on social network modeling and analysis,* (pp. 133-145). Committee on Human Factors, National Research Council.

Kathleen M.Carley and Michael Palmquist, 1992, "Extracting, Representing and Analyzing Mental Models." Social Forces, 70(3): 601-636

Kathleen M.Carley, 1993, "Content Analysis," in Asher R.E. et al.(Eds.), The Encyclopedia of Language and Linguistics . Edinburgh, UK: Pergamon Press. Vol. 2: 725-730.

Diesner, J., & Carley, K.M., 2005. "Exploration of Communication Networks from the Enron Email Corpus," Proceedings of the Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining 2005, pp. 3-14. Newport Beach, CA, April 21-23, 2005.

Jana Diesner, Kathleen M. Carley, 2005, "Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a novel method for Network Text Analysis," In V.K. Narayanan & D.J. Armstrong (Eds.) Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations, Chapter 4, Harrisburg, PA: Idea Group Publishing.

Diesner, J., & Carley, K.M. (2004). *AutoMap1.2 - Extract, analyze, represent, and compare mental models from texts*. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-100. URL: http://reports-archive.adm.cs.cmu.edu/isri2004.html (01-22-2004).

Diesner, J., & Carley, K.M. (2005). Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a novel method for Network Text Analysis. Chapter 4 in V.K. Narayanan, & D.J. Armstrong (Eds.), *Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations* (pp.81-108). Harrisburg, PA: Idea Group Publishing.

Jurafsky, D., & Marton, J.H. (2000). *Speech and Language Processing*. Upper Saddle River, New Jersey: Prentice Hall.

David Kaufer and Kathleen M.Carley, 1993, "Condensation Symbols: Their Variety and Rhetorical Function in Political Discourse." Philosophy and Rhetoric, 26(3): 201-226.

Klein, H. (1997). Classification of Text Analysis Software. In R. Klar, & O. Opitz (Eds.), *Classification and knowledge organization: Proceedings of the 20th annual conference of the Gesellschaft für Klassifikation e.V.*, (pp. 255-261). University of Freiburg , 1996, Berlin , New York : Springer. 1997.

Krovetz, Robert (1995). *Word Sense Disambiguation for Large Text Databases*. Unpublished PhD Theis, University of Massachusetts.

Magnini, B., Negri, M., Prevete, R., & Tanev, H. (2002). A WordNet-based approach to Named Entities Recognition. In *Proceedings of SemaNet'02: Building and Using Semantic Networks*, (pp. 38-44). Taipei, Taiwan, August 2002.

Mrvar, A. *Centrality measures*. URL: http://mrvar.fdv.uni-lj.si/sola/info4/uvod/part4.pdf (06-13-2004)

Michael Palmquist, Kathleen M. Carley, and Thomas Dale, 1997, "Two applications of automated text analysis: Analyzing literary and non-literary texts." Chapter 10 in C. Roberts (Ed.), Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 171-189.

Popping, R., & Roberts, C.W. (1997). Network Approaches in Text Analysis. In R. Klar & O. Opitz (Eds.), *Classification and Knowledge Organization: Proceedings of the 20th annual conference of the Gesellschaft für Klassifikation e.V.*, (pp. 381-898). University of Freiburg , Berlin , New York : Springer.

Porter, M.F. 1980. An algorithm for suffix stripping. *I 14* (3): 130-137.

Tsvetovat, M., Reminga, J., & Carley, K.M. (2004). *DyNetML: Interchange Format for Rich Social Network Data*. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, CASOS Technical Report CMU-ISRI-04-105. URL: http://reports-archive.adm.cs.cmu.edu/anon/isri2004/abstracts/04-105.html .

Zuell, C., & Alexa, M. (2001). Automatisches Codieren von Textdaten. Ein Ueberblick ueber neue Entwicklungen. In W. Wirth & E. Lauf (Eds.), *Inhaltsanalyse - Perspektiven, Probleme, Potenziale* (pp. 303-317). Koeln: Herbert von Halem.

Where to learn to more about Dynamic Network Analysis

Kathleen M. Carley, 2003, “Dynamic Network Analysis” in Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, Ronald Breiger, Kathleen Carley, and Philippa Pattison, (Eds.) Committee on Human Factors, National Research Council, National Research Council. Pp. 133-145, Washington, DC.

Kathleen M. Carley, 2002, “Smart Agents and Organizations of the Future” The Handbook of New Media. Edited by Leah Lievrouw and

Sonia Livingstone, Ch. 12, pp. 206-220, Thousand Oaks, CA, Sage.

Kathleen M. Carley, Jana Diesner, Jeffrey Reminga, Maksim Tsvetovat, 2005-forthcoming, Toward an Interoperable Dynamic Network Analysis Toolkit, DSS Special Issue on Cyberinfrastructure for Homeland Security: Advances in Information Sharing, Data Mining, and Collaboration Systems.