Automap

Help Manual

Summer Institute
June 25, 2007

Carnegie Mellon University
School of Computer Science
Center for Computational Analysis of Social and Organizational Systems (CASOS)
5000 Forbes Avenue
1325 Wean Hall
Pittsburgh, PA, 15213
Telephone: 1-412-268-3163
Fax: 1-412-268-1744
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)

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.
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.
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 entities 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 entities 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 entities are actors. As such, a network can be a cell of terrorists, employees of global company or simply a group of friends. However, entities 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 entities (multi-entities) 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, entities in DNA are like atoms in quantum theory, entities can be, though need not be, treated as probabilistic. Whereas entities in a traditional SNA model are static, entities in a DNA model have the ability to learn. Properties change over time; entities 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 entity 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.

Where to learn to more:

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 the mouse over that particular tool.
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.

The Browser Menu allows you to quickly navigate from one text sample.
Text 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.
Parts of Speech

The Parts of Speech feature does exactly that, when text is loaded into Automap this feature will show you what part of speech each word is.

To use the feature:

- In AutoMap menu bar, select **File > Open single file**. A folder chooser window will pop up.
- Navigate to the folder containing the texts you wish to analyze, and single click on it.
- Load your text file and it will appear in **Panel 1 - Original Data**.
- Next go to **Panel 2** and select **3. Pre-Processing Settings**.
- Once in the Pre-Processing Settings, select **1. Utilities**.
- Scroll down the options and single click on **Tag Texts and save Tagging Output**, which is under the heading **Parts of Speech Tagging**.

**NOTE: Data has to be in txt format with .txt extension**

- The Tag Texts option that you have just selected will take your text document and define each part of text.
- Go back to **Panel 1** and select **8. Texts after Parts-of-Speech Tagging**. This option will show you your document with the different parts of speech.
To undo the Parts of Speech Tag, simply go back to Panel 2, 1. Utilities option Parts of Speech Tagging and select the Undo option.

2. Parts of Speech Tagging

This routine associates every word after the highest level of pre-processing applied so far with its Parts of Speech.

In order to tag your texts, go to the Utilities Panel, Parts of Speech Tagging field, and press the Tag Texts button.

To see the resulting tagging, go to the upper left panel, and Select 8. Texts after Parts of Speech tagging.

To undo the tagging, go to the Utilities Panel, Parts of Speech Tagging field, and press the Undo Tagging button.

This POS tagger was implemented based on a Hidden Markov Model. The learning data stems from the Penn Treebank 3 corpus. We are grateful to Alex Rudnicky from CMU for providing the training data to us.
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 occurring 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:
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.
Text Pre-Processing

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**:
   - **Generalization Thesaurus**: Can be applied before Semantic Network Analysis is run. Can be applied before Meta-Matrix Thesaurus is applied.
   - **Meta-Matrix Thesaurus**: Has to be applied if Meta-Matrix Analysis should be run.
   - **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:

Text Pre-Processing

3.1 Named-Entity Recognition

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:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Utilities</td>
<td>2. Stemming</td>
<td>3. Delete List</td>
<td></td>
</tr>
</tbody>
</table>

- Remove Symbols
- Undo Removal of Symbols

Named Entity Recognition (Language Independent)
- [ ] as base for Generalization Thesaurus
- Create list

N-Gram Detection (Language Independent)
- [ ] as base for Generalization Thesaurus

The black ellipses in the screen shot 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:

<table>
<thead>
<tr>
<th>a</th>
<th>John Darren and Mrs. Brown</th>
</tr>
</thead>
<tbody>
<tr>
<td>an</td>
<td>Mr. Cray</td>
</tr>
<tr>
<td>some</td>
<td>He couldn't. Mrs. Brown</td>
</tr>
<tr>
<td>many</td>
<td>Prof. Darren Mrs. Brown &amp; Mr. Cray</td>
</tr>
<tr>
<td>this</td>
<td>13-Jun</td>
</tr>
<tr>
<td>that</td>
<td>But</td>
</tr>
<tr>
<td>these</td>
<td>Mr. Cray's</td>
</tr>
<tr>
<td>those</td>
<td>Problem solved. Then Mrs. Brown</td>
</tr>
<tr>
<td>the</td>
<td>Then Mrs. Brown</td>
</tr>
<tr>
<td>all</td>
<td>12-Jun</td>
</tr>
<tr>
<td>one</td>
<td>Weeding</td>
</tr>
<tr>
<td>every</td>
<td>She</td>
</tr>
<tr>
<td></td>
<td>She asked Mr. Cray</td>
</tr>
<tr>
<td></td>
<td>Mrs. Brown</td>
</tr>
</tbody>
</table>

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.

Symbol Removal

3.21 N-gram Identification: Bigrams

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.
Text PreProcessing

4. 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.txt, 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:
Mr. Cray's brown dog eat the lotus blossom at 10 am. Mrs. Brown be unhappy with the dog. She yelled at it saying "You impossibl dog!" But the dog keep eating the flow and weed. She asked Mr. Cray to stop the dog. He couldn't. Mrs. Brown planted ros and weeded the gard. The silly dog & dig up the ros looking for a vol on Jun 12, 1880. Weeding be no long needed.

Prof. Dar, Mrs. Brown & Mr. Cray meet the next day to concoct a plan. John Dar and Mrs. Brown put up a scarecrow. She think it be scar the dog. Mr. Craye put up a fenc. Problem solved. Then Mrs. Brown planted lotus, carnation, daffodil, and ros.
Text PreProcessing

5. Deletion

The Delete List is not case sensitive.

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**:
The Delete List can be edited. *(see section 5.3 Edit a Delete List)*

**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. *(see section 5.3 Edit a Delete List)*

**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. *(see section 5.3 Edit a Delete List)*

**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

<table>
<thead>
<tr>
<th>Input text</th>
<th>Tool used</th>
<th>Setting</th>
<th>Resulting text</th>
</tr>
</thead>
<tbody>
<tr>
<td>AutoMap's extensive Delete List: a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my,</td>
<td></td>
<td>Direct Adjacency (default)</td>
<td></td>
</tr>
</tbody>
</table>
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's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy with the dog. She yelled at him saying "You impossible dog!" The dog kept eating flowers and weeds. Mrs. Brown planted roses and weeded the garden. Silly dog dug up roses looking for vole on June 12, 1880. Weeding no longer needed. Prof. Darren, Mrs. Brown put up a scarecrow. She thought it would scare the dog. Mr. Cray's brown dog kept eating flowers and weeds. Mrs. Brown planted roses and weeded the garden. Silly dog dug up roses looking for vole on June 12, 1880. Weeding 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's brown dog kept eating flowers and weeds. Mrs. Brown planted roses and weeded the garden. Silly dog dug up roses looking for vole on June 12, 1880. Weeding 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's brown dog kept eating flowers and weeds. Mrs. Brown planted roses and weeded the garden. Silly dog dug up roses looking for vole on June 12, 1880. Weeding 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's brown dog kept eating flowers and weeds. Mrs. Brown planted roses and weeded the garden. Silly dog dug up roses looking for vole on June 12, 1880. Weeding 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's brown dog kept eating flowers and weeds. Mrs. Brown planted roses and weeded the garden.
Text PreProcessing

6. Generalization Thesaurus

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 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:

![Generalization Thesaurus Interface](image)

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" respectively. 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.

<table>
<thead>
<tr>
<th>Pre-processing technique applied</th>
<th>Entries</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Create Named-Entity List</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>John Darren and Mrs. Brown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mr. Cray</td>
</tr>
<tr>
<td></td>
<td></td>
<td>He couldn't. Mrs. Brown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prof. Darren Mrs. Brown &amp; Mr. Cray</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13-Jun</td>
</tr>
<tr>
<td></td>
<td></td>
<td>But</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mr. Cray's</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mr. Craye</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Problem solved. Then Mrs. Brown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Then Mrs. Brown</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-Jun</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weeding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>She</td>
</tr>
<tr>
<td></td>
<td></td>
<td>She asked Mr. Cray</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mrs. Brown</td>
</tr>
</tbody>
</table>


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.

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.

<table>
<thead>
<tr>
<th><strong>4. Create Union Concept List.</strong></th>
<th><em>(Coincides with Union Concept List loaded into AutoMap, see cell below)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>1880</td>
<td>asked</td>
</tr>
<tr>
<td>asked</td>
<td>blossom</td>
</tr>
<tr>
<td>brown</td>
<td>but</td>
</tr>
<tr>
<td>but</td>
<td>concoct</td>
</tr>
<tr>
<td>craye</td>
<td>daffodils</td>
</tr>
<tr>
<td>daffodils</td>
<td>dinner</td>
</tr>
<tr>
<td>dinner</td>
<td>discussed</td>
</tr>
<tr>
<td>discussed</td>
<td>dog</td>
</tr>
<tr>
<td>dog</td>
<td>dog!</td>
</tr>
<tr>
<td>dog!</td>
<td>down</td>
</tr>
<tr>
<td>down</td>
<td>dug</td>
</tr>
<tr>
<td>dug</td>
<td>fence</td>
</tr>
<tr>
<td>fence</td>
<td>for</td>
</tr>
<tr>
<td>for</td>
<td>his</td>
</tr>
<tr>
<td>his</td>
<td>how</td>
</tr>
<tr>
<td>how</td>
<td>it</td>
</tr>
<tr>
<td>it</td>
<td>looking</td>
</tr>
<tr>
<td>looking</td>
<td>lotus</td>
</tr>
<tr>
<td>lotus</td>
<td>on</td>
</tr>
<tr>
<td>on</td>
<td>over</td>
</tr>
<tr>
<td>over</td>
<td>plan</td>
</tr>
<tr>
<td>plan</td>
<td>pm</td>
</tr>
<tr>
<td>pm</td>
<td>problem</td>
</tr>
<tr>
<td>problem</td>
<td>said</td>
</tr>
<tr>
<td>said</td>
<td>saying</td>
</tr>
<tr>
<td>saying</td>
<td>scary</td>
</tr>
<tr>
<td>scary</td>
<td>solved</td>
</tr>
<tr>
<td>solved</td>
<td>stopped</td>
</tr>
<tr>
<td>stopped</td>
<td>take</td>
</tr>
<tr>
<td>take</td>
<td>up</td>
</tr>
<tr>
<td>up</td>
<td>watered</td>
</tr>
<tr>
<td>watered</td>
<td>weeded</td>
</tr>
<tr>
<td>weeded</td>
<td>weeding</td>
</tr>
<tr>
<td>weeding</td>
<td>weeds</td>
</tr>
<tr>
<td>weeds</td>
<td>with</td>
</tr>
<tr>
<td>with</td>
<td>worked</td>
</tr>
<tr>
<td>worked</td>
<td>yelled</td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Input text</th>
<th>Tool used</th>
<th>Setting</th>
<th>Resulting text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mr. Cray's brown dog ate the lotus blossom at 10 am. Mrs. Brown was unhappy with the dog. She yelled at it saying &quot;you impossible dog!&quot; 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 &amp; 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.</td>
<td></td>
<td>Not Selected - Thesaurus content only (default)</td>
<td>mr. cray's brown dog eating the lotus blossom at 10 am. mrs. brown was unhappy with the dog. she yelling at it saying &quot;you impossible dog!&quot; 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 &amp; 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.</td>
</tr>
</tbody>
</table>
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.

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

<table>
<thead>
<tr>
<th>Input text</th>
<th>Tool used</th>
<th>Setting</th>
<th>Resulting text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy with dog. yelled saying &quot;You impossible dog!&quot; 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 &amp; mr. craye 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.</td>
<td>Thesaurus (same as above)</td>
<td>Not selected - Thesaurus content only (default)</td>
<td>mr. cray's brown dog eating lotus blossom 10 am. mrs. brown was unhappy with dog. yelling saying &quot;you impossible dog!&quot; 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 &amp; mr. craye 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.</td>
</tr>
</tbody>
</table>
| Selected - Thesaurus content only, Direct | | | eating... yelling.....,..

....
prof_john_darren.......,,
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.

<table>
<thead>
<tr>
<th>Input text</th>
<th>Tool used</th>
<th>Setting</th>
<th>Resulting text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text after</td>
<td>Thesaurus</td>
<td>Rhetorical Adjacency</td>
<td>mr. cray's brown dog eating xxx lotus blossom xxx 10 am. Mrs. brown was unhappy with xxx dog. xxx yelling xxx xxx saying &quot;you impossible dog!&quot; 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.</td>
</tr>
<tr>
<td>application of</td>
<td>(same as</td>
<td>Thesaurus content only,</td>
<td>prof. darren, mrs. brown &amp; mr. cray met xxx next day xxx concoct xxx plan.</td>
</tr>
<tr>
<td>Customized</td>
<td>above)</td>
<td>Thesaurus content only</td>
<td>prof john darren xxx mrs. brown put up xxx scarecrow. xxx thought xxx xxx scare</td>
</tr>
<tr>
<td>extensive</td>
<td></td>
<td>(default if Thesaurus content only is chosen)</td>
<td></td>
</tr>
<tr>
<td>delete list,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rhetorical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adjacency:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mr. Cray's brown</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dog ate xxx</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lotus blossom xxx</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 am. Mrs. Brown</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>was unhappy with xxx</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dog. xxx yelled xxx</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>xxx saying &quot;You</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
impossible dog!
xxx dog kept eating flowers weeds.
xxx asked Mr. Cray stop dog.
xxx 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.

Thesaurus content only, Direct Adjacency

(xxx dog. mr. craye put up xxx fence. problem solved. then mrs. brown planted lotus, carnations, daffodils, xxx roses.

(xxx dog. mr. craye put up xxx fence. problem solved. then mrs. brown planted lotus, carnations, daffodils, xxx roses.

Thesaurus content only, Rhetorical Adjacency

(xxx dog. mr. craye put up xxx fence. problem solved. then mrs. brown planted lotus, carnations, daffodils, xxx roses. Then Mrs. Brown planted lotus, carnations, daffodils, xxx roses..
Text PreProcessing

7. 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 **NOT** case 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:

```plaintext
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:

```plaintext
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 to 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,
```
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:

<table>
<thead>
<tr>
<th>Setting</th>
<th>Resulting text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Thesaurus content only (default)</td>
<td>mr. cray's brown &lt;agent&gt; &lt;task&gt; the &lt;resource&gt; blossom at 10 am. mrs. brown was unhappy with the &lt;agent&gt; . she &lt;task&gt; at it saying &quot;you impossible dog!&quot; but the &lt;agent&gt; kept &lt;task&gt; the &lt;resource&gt; and &lt;task&gt; . she asked mr. cray to stop the &lt;agent&gt; . he couldn't. mrs. brown planted &lt;resource&gt; and weeded the garden. the silly &lt;agent&gt; dug up the &lt;resource&gt; looking for a vole on june 12, 1880. weeding was no longer needed.</td>
</tr>
</tbody>
</table>
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>. 

<table>
<thead>
<tr>
<th>Thesaurus</th>
<th>Direct</th>
<th>Adjacency</th>
</tr>
</thead>
<tbody>
<tr>
<td>content only, Direct Adjacency (default if Thesaurus content only is chosen)</td>
<td>. &lt; agent&gt; &lt; task&gt; &lt; resource&gt; . . &lt; agent&gt;. &lt; task&gt; &lt; agent&gt; &lt; task&gt; &lt; resource&gt; &lt; task&gt;. . &lt; agent&gt;. . . &lt; resource&gt;. . &lt; agent&gt;. &lt; resource&gt; . . . . &lt; task&gt;. . &lt; agent&gt;. . . &lt; agent&gt;. . . . &lt; resource&gt; &lt; resource&gt; &lt; resource&gt;.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Thesaurus content only, Rhetorical Adjacency</th>
</tr>
</thead>
</table>

6/14/2007 5:04 PM
Text PreProcessing

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 Delete list or not)* 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:


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:

<table>
<thead>
<tr>
<th>Setting</th>
<th>Resulting text</th>
</tr>
</thead>
</table>
| **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 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 xxx xxx dog xxx xxx xxx xxx xxx  
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. |
Detailed example for Sub-Matrix Selection:

<table>
<thead>
<tr>
<th>Select Sub-Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Agent</td>
</tr>
<tr>
<td>Task</td>
</tr>
<tr>
<td>Event</td>
</tr>
<tr>
<td>Organization</td>
</tr>
<tr>
<td>Location</td>
</tr>
<tr>
<td>Role</td>
</tr>
<tr>
<td>Action</td>
</tr>
<tr>
<td>Attribute</td>
</tr>
</tbody>
</table>

---

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 roses 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 dog.
xxx. xxx xxx xxx xxx. xxx xxx. xxx
xxx xxx xxx lotus, carnations,
daffodils, xxx roses.
Anaphora Resolution

Automap can perform anaphora resolution.

Anaphora is a linguistic instance of a grammatical expression referring to another. In a general linguistic sense, an anaphoric expression is one represented by some kind of deictic, a process whereby words or expressions rely absolutely on context. Sometimes this context needs to be identified.

The anaphora resolution function is a pre-processing setting and can be found in Panel P2, Tab no. 3. Pre-processing settings in Automap's main interface. The screen shot below shows where the tool is located.

This tool extracts network information from text that requires additional textual references from within a certain phrase or grammatical construct.

To Access Anaphora Resolution: Panel P2 > Tab no. 3 Pre-Processing Settings > 1. Utilities > scroll down to bottom of window

Some examples of what would constitute an Anaphora Resolution might include:

- The boy took the hamburger and ate it. It is anaphoric under the strict definition (it refers to the hamburger).
- Larry called off work because he felt sick. He is anaphoric (it refers to Larry).

Moreover, think of a sentence being taken out of context and the missing information required to fully explain the meaning.

Example: The President of our company came to visit us yesterday. The visit was the first time he had come to Pittsburgh since 1998.
If the second sentence stands alone by itself, it is necessary to resolve the anaphor.

Automap can apply this pre-process settings, identify Anaphora Resolution instances and then have them isolated and removed, which is activated by selecting the first and middle button under the Anaphora Resolution tool. The third button allows you to output the Anaphora Resolution contained in your text. Each of these preprocessing settings can be undone.
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.

```
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Analysis Settings</td>
<td>2. Output Options</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

**Coding Ties**

**Directionality**
- Uni-directional
- Bi-directional

**Strength**
- Frequency

**Windowing**

**Punctuation**
- Ignore punctuation completely
- Reset window at the end of paragraphs only
- Reset window at the end of paragraphs and sentences

<table>
<thead>
<tr>
<th>Coding Ties</th>
<th>Directionality</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Uni-Directional</td>
<td>Frequency</td>
</tr>
<tr>
<td></td>
<td>Bi-Directional</td>
<td></td>
</tr>
</tbody>
</table>

**Overview on the possible Settings:**

- **Directionality**
  - 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**
  - Frequency (The cumulative frequency of every existing statement.)
  - **Item not checked**: Existence of frequency will be printed out (binary result).
Specify the way statements are counted.

**Windowing**
Windowing is a method that codes the text as a map by placing relationships between pairs of Concepts that occur within a window.

<table>
<thead>
<tr>
<th>Punctuation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ignore punctuation completely</strong> (Statements will be placed between all concepts.)</td>
<td></td>
</tr>
<tr>
<td><strong>Reset window at end of paragraphs only</strong> (Statements will be placed only within every single paragraph.)</td>
<td></td>
</tr>
<tr>
<td><strong>Reset window at end of paragraphs and sentences</strong> (Statements will be placed only within every single sentence.)</td>
<td></td>
</tr>
</tbody>
</table>

| Window Size | Window size between 2 and 100. The Window Size defines how distant concepts can be and still have a relationship. 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:
**Semantic Text Analysis**

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:

![Concept List Example](image)

### 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 **No. 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 where to access the **Refresh Union Concept List** from the drop-down menu bar:

![Dropdown Menu](image)

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.

<table>
<thead>
<tr>
<th>File</th>
<th>Run Analysis</th>
<th>Tools</th>
<th>Help</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Open single file</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Open multiple files (Select folder)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open Delete List</td>
<td>Open Generalization Thesaurus</td>
<td>Open Meta-Matrix Thesaurus</td>
<td>Open Sub-Matrix Selection</td>
</tr>
<tr>
<td>Create and Refresh Union Concept List</td>
<td>Output Storage Manager</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Save Concept List per Text</td>
<td>Save Union Concept List</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Save applied Delete List</td>
<td>Save applied Generalization Thesaurus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Save applied Meta-Matrix Thesaurus</td>
<td>Save Sub-Matrix Selection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Save Text(s) after Stemmer applied</td>
<td>Save Text(s) after Delete List applied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Save Text(s) after Generalization Thesaurus applied</td>
<td>Save Text(s) after Meta-Matrix Thesaurus applied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Save Text(s) after Sub-Matrix selection applied</td>
<td>Clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exit</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 put up 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 scary 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.

4. 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.

### 4.1 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:

<table>
<thead>
<tr>
<th>Input text</th>
<th>Tool used</th>
<th>Setting</th>
<th>Resulting text</th>
</tr>
</thead>
</table>
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 & Mrs. Brown put up a scarecrow. They thought it would scare the dog. Mr. Craye put up a fence. Problem solved. Then Mrs. Brown planted lotus, carnations, daffodils, roses, and roses.

Mr. Cray's brown dog ate lotus blossom 10 am. Mrs. Brown was unhappy with the dog. She yelled saying "You impossible dog!" The dog kept eating flowers weeds. Mrs. Brown asked Mr. Cray to stop the dog. Mrs. Brown planted roses weeded garden. The silly dog dug up 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 & 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, roses.

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

These are the Map and

Rhetorical Adjacency

These are the Map and
<table>
<thead>
<tr>
<th>Statistics outputs for the first text (direct adjacency):</th>
<th>Statistics outputs for the second text (rhetorical adjacency):</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Map:</strong></td>
<td><strong>Map:</strong></td>
</tr>
<tr>
<td>5 mrs</td>
<td>2 mr</td>
</tr>
<tr>
<td>2 mr</td>
<td>2 brown</td>
</tr>
<tr>
<td>2 darren</td>
<td>1 you</td>
</tr>
<tr>
<td>2 brown</td>
<td>1 worked</td>
</tr>
<tr>
<td>1 you</td>
<td>1 weeds</td>
</tr>
<tr>
<td>1 worked</td>
<td>1 watered</td>
</tr>
<tr>
<td>1 with</td>
<td>1 was</td>
</tr>
<tr>
<td>1 week</td>
<td>1 too</td>
</tr>
<tr>
<td>1 weeds</td>
<td>1 then</td>
</tr>
<tr>
<td>1 watered</td>
<td>1 take</td>
</tr>
<tr>
<td>1 was</td>
<td>1 stopped</td>
</tr>
<tr>
<td>1 too</td>
<td>1 said</td>
</tr>
<tr>
<td>1 thought</td>
<td>1 prof</td>
</tr>
<tr>
<td>1 then</td>
<td>1 pleased</td>
</tr>
<tr>
<td>1 take</td>
<td>1 plan</td>
</tr>
<tr>
<td>1 stopped</td>
<td>1 over</td>
</tr>
<tr>
<td>1 scary</td>
<td>1 on</td>
</tr>
<tr>
<td>1 scarecrow</td>
<td>1 now</td>
</tr>
<tr>
<td>1 said</td>
<td>1 no</td>
</tr>
<tr>
<td>1 roses</td>
<td>1 next</td>
</tr>
<tr>
<td>1 prof</td>
<td>1 mr</td>
</tr>
<tr>
<td>1 pm</td>
<td>1 met</td>
</tr>
<tr>
<td>1 pleased</td>
<td>1 lotus</td>
</tr>
<tr>
<td>1 plan</td>
<td>1 longer</td>
</tr>
<tr>
<td>1 painted</td>
<td>1 june</td>
</tr>
<tr>
<td>1 over</td>
<td>1 john</td>
</tr>
<tr>
<td>1 on</td>
<td>1 happy</td>
</tr>
<tr>
<td>1 now</td>
<td>1 had</td>
</tr>
<tr>
<td>1 no</td>
<td>1 good</td>
</tr>
<tr>
<td>1 next</td>
<td>1 garden</td>
</tr>
<tr>
<td>1 mr</td>
<td>1 following</td>
</tr>
<tr>
<td>1 met</td>
<td>1 fence</td>
</tr>
<tr>
<td>1 lotus</td>
<td>1 dog</td>
</tr>
<tr>
<td>1 longer</td>
<td>1 discussed</td>
</tr>
<tr>
<td>1 june</td>
<td>1 day</td>
</tr>
<tr>
<td>1 john</td>
<td>1 darren</td>
</tr>
<tr>
<td>1 how</td>
<td>1 craye</td>
</tr>
<tr>
<td>1 happy</td>
<td>1 cray's</td>
</tr>
<tr>
<td>1 had</td>
<td>1 cray</td>
</tr>
<tr>
<td>1 good</td>
<td>1 carnations</td>
</tr>
<tr>
<td>1 garden</td>
<td>1 brown</td>
</tr>
<tr>
<td>1 1880</td>
<td>1 13</td>
</tr>
<tr>
<td>1 12</td>
<td>1 12</td>
</tr>
<tr>
<td>Stat:</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td># of concepts analyzed:</td>
<td># of concepts analyzed:</td>
</tr>
<tr>
<td>unique: 62</td>
<td>unique: 60</td>
</tr>
<tr>
<td>total: 88</td>
<td>total: 84</td>
</tr>
<tr>
<td># of concepts in statements:</td>
<td># of concepts in statements:</td>
</tr>
<tr>
<td>unique: 54</td>
<td>unique: 55</td>
</tr>
<tr>
<td>total: 60</td>
<td>total: 61</td>
</tr>
<tr>
<td># of isolated concepts:</td>
<td># of isolated concepts:</td>
</tr>
<tr>
<td>unique: 12</td>
<td>unique: 6</td>
</tr>
<tr>
<td>total: 13</td>
<td>total: 7</td>
</tr>
<tr>
<td># of statements:</td>
<td># of statements:</td>
</tr>
<tr>
<td>unique: 54</td>
<td>unique: 55</td>
</tr>
<tr>
<td>total: 60</td>
<td>total: 61</td>
</tr>
<tr>
<td>Density (based on Statements):</td>
<td>Density (based on Statements):</td>
</tr>
<tr>
<td>unique: 0.87</td>
<td>unique: 0.92</td>
</tr>
<tr>
<td>total: 0.97</td>
<td>total: 1.02</td>
</tr>
</tbody>
</table>

**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](http://www.casos.cs.cmu.edu/projects/automap/publications.html)).
Content Analysis

From content analysis to semantic networks

I. Open AutoMap

On Empire:

- Navigate to C:\Documents and Settings\Carley\Desktop\Tools.
- Double-click the Run Automap icon. The Automap Graphical User Interface (GUI) pops up.

From anywhere else: (requires a web connection)

http://www.casos.cs.cmu.edu/projects/automap/software.html

II. Overview of the AutoMap Graphical User Interface (GUI)

Below is a screen capture of the AutoMap GUI:

The AutoMap GUI is divided into four primary quadrants, or panels: P1 (top left), P2 (bottom left), P3 (top right), and P4 (bottom right).

A drop-down menu bar at the top of the window provides access to analysis tools and utilities. The Browse Menu allows you to quickly navigate between texts you have loaded into Automap.

The P2 panel can be edited, but the other panels (P1, P3, P4) cannot. Information displayed on each panel is always related to the information displayed in the other three panels. The Browse Menu relates to all four panels at the same time.
III. Load multiple text documents into AutoMap

In order to analyze multiple texts simultaneously, you will need to store the texts together in one folder.

**Example:** C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextInputFiles

- In the AutoMap menu bar, select File, then **Open Multiple Files**. A folder chooser window will pop up.
- Navigate to the folder containing the texts you wish to analyze, and single-click on it.

  **Example:** C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextInputFiles

- Click the **Open** button in the bottom right corner.

The first text will be displayed in **panel P1** under the tab titled **1.Original Texts**. You can use the Browse Menu to browse through the texts.

Below is a screen capture of the first text display:

![Screen capture of the first text display](image)

The Browse Menu allows you to quickly navigate from one text sample to the next.

IV. Carry out a simple content analysis

A simple content analysis determines the frequencies of all words in a text document.

**Examine the Concept List:**

The Concept List considers concepts in each text file individually. It is created automatically when a text or set of texts is loaded into Automap, and is displayed in **P2** under the tab titled **1.Concept List**.
The Concept List tells you several important facts about each text in your set:

- The number of concepts found in the text displayed in **P1** under the tab titled **1. Original Text**.
- Related frequencies of those concepts
- Cumulated unique concepts and total concepts contained in the data set

**Note:** The number of unique concepts considers each concept only once. 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**. To sort the list by frequency, click on the second-column header **Frequency**.

Below is a screen capture showing part of the Concept List, sorted by frequency, for the first document in our loaded folder of texts (Text1-in-user's guide.txt):

Create and refresh the Union Concept List:

The Union Concept List considers concepts across all texts in a set (unlike the Concept List, which considers each text in the set individually). It is located in **P2** under the tab titled **2. Union Concept List**.

The Union Concept List tells you several important facts about your text set:

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

Before viewing your Union Concept List, you must refresh it. In the Menu Bar, choose **File**, then **Refresh Union Concept List**.

The Union Concept List can be refreshed after each step of pre-processing to visualize the impact of pre-processing operations on the union of concepts.
Below is a screen capture showing where to find the **Refresh Union Concept List** command:

After refreshing your Union Concept List, you can view it under the **2. Union Concept List** tab in **P2**. Below is a screen capture showing where this tab is located in the GUI:

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

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

**V. Save your Union Concept List as a csv file**

To specify a location for your Union Concept List file, choose **File in the Menu Bar, then Output Storage Manager**.

Below is a screen capture showing where to find this in the GUI:
A window will pop up listing all pre-processing output storage. This is the Output Storage Manager.

The Union Concept List appears as the second item in the Output Storage Manager. Click the Save file as... button to the right of the default file pathway. In the file chooser window that pops up, browse to the pathway you want.

**Example:** `C:\Documents and Settings\carley\Desktop\AutomapLesson1\UnionConceptList`

Below is a screen capture showing where to find the correct Save file as... button in the Output Storage Manager:

Click the **Save Settings** button at the bottom right corner of the Output Storage Manager. Close the Output Storage Manager window by clicking on the **red X** at the top right corner.
In the Automap Menu Bar, choose **File**, then Save **Union Concept List**.

**VI. View your csv file in Microsoft Excel**

Navigate to the location of your new csv file and double-click on its icon.

**Example:** C:\Documents and Settings\carley\Desktop\AutomapLesson1\UnionConceptList

A box titled **Open With** pops up. Click on Microsoft Excel, then click **OK**.

**Note:** You must have MS Excel installed on your computer in order to view your csv file this way.

**VII. Save and view your results as a binary csv file**

Follow Steps 3 and 4.

**VIII. Carry out a simple Semantic Network Analysis**

A simple network analysis involves no thesauri and uses all words.

A Semantic Network Analysis can be run on original input texts or on texts that have been pre-processed with a Delete List and/or a Generalization Thesaurus.

**Before running the analysis:**

Make sure you have completed the following steps:

- Pre-process the texts if you wish to
  - Specify the analysis settings (otherwise the default settings will be applied)
  - Specify additional output options if you wish to (if not, the default outputs will be generated)

**About the analysis settings:**

Use the **Analysis Settings** index card to specify the analysis settings. Your settings will automatically be applied to the analysis (you do not need to confirm them).

Below is a screen capture showing where to find the **Analysis Settings** index card in P2:
The following chart lists the possible analysis settings:

<table>
<thead>
<tr>
<th>Coding Ties</th>
<th>Directionality</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specify the way statements are counted.</td>
<td>Select one of the following two possibilities by checking the button.</td>
<td>Uni-Directional (When coding a tie, only 1st-&gt;2nd concept should be noted)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bi-Directional (When coding a tie, both 1st &lt;-&gt; 2nd concept shall be noted)</td>
</tr>
<tr>
<td></td>
<td>Strength</td>
<td>Frequency (The cumulative frequency of every existing statement.)</td>
</tr>
<tr>
<td></td>
<td>Strength will be printed out by default. To not print out Strength uncheck the item.</td>
<td>Item not checked: Existence of frequency will be printed out (binary result).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Windowing</th>
<th>Punctuation</th>
<th>Window Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windowing is a method that codes the text as a map by placing relationships between pairs of Concepts that occur within a window.</td>
<td>Select one option by checking the radio button.</td>
<td>Select one window size by using the number chooser.</td>
</tr>
<tr>
<td></td>
<td>Ignore punctuation completely (Statements will be placed between all concepts.)</td>
<td>Window size between 2 and 100. The Window Size defines how distant concepts can be and still have a relation</td>
</tr>
</tbody>
</table>
Only concepts in the same window can form statements.

If you do not want to change any of the suggested options, the analysis will be done with a set of standard (default) settings.

Below is a screen capture showing the standard settings:

About the output options:

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

Below is a screen capture showing the Output Options index card in P2:
For all types of multiple analysis, term distribution lists and matrices can be chosen as output options in the upper two fields of the 10. Output Options index card in P2.

**Points to consider:**

- No list or matrix is generated by default.
- Lists or matrices are only generated if the user checks the item he or she wants 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 are automatically saved in a folder called Term Distribution Lists and Matrices under the root directory of AutoMap. This folder is overwritten with every new analysis you run. If you want to save the results of a current Term Distribution Lists and Matrices folder, simply rename that folder.
- Two output lists are generated for each Term Distribution List checked.
- One output matrix is generated for each Term Distribution Matrix checked.

The following chart lists the types and content of Term Distribution Lists and Matrices:

<table>
<thead>
<tr>
<th>Output Type</th>
<th>Name of output</th>
<th>Content of output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts analyzed</td>
<td>List of concept analyzed.csv</td>
<td>Concept, Text, Frequency</td>
</tr>
<tr>
<td></td>
<td>Statistics of concepts analyzed.csv</td>
<td>Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts</td>
</tr>
<tr>
<td></td>
<td>List of concept in statements.csv</td>
<td>Concept, Text, Frequency</td>
</tr>
<tr>
<td><strong>Distribution List</strong></td>
<td><strong>Statistics of concept in statements.csv</strong></td>
<td>Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Distribution List</strong></td>
<td><strong>List of isolates in statements.csv</strong></td>
<td>Concept, Text, Frequency</td>
</tr>
<tr>
<td><strong>Statistics</strong></td>
<td><strong>Statistics of isolates.csv</strong></td>
<td>Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts</td>
</tr>
<tr>
<td><strong>Statements</strong></td>
<td><strong>List of statements.csv</strong></td>
<td>Statement, Text, Frequency</td>
</tr>
<tr>
<td><strong>Statistics</strong></td>
<td><strong>Statistics of statements.csv</strong></td>
<td>Statement, Cumulated sum across text set, Number of text statement occurs in, Percentage of texts statement occurs in, Texts</td>
</tr>
</tbody>
</table>
| **Concepts in statements by 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 Distribution Matrices** | **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 |
If count was chosen, cells contain cumulated frequency of concept. If binary was chosen, cells denote existence (1) or absence (0) of concept.

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.

Save your upcoming analysis in DyNetML format:

Before running your Multiple Map Analysis, you must specify that your results will be in DyNetML format so that they can eventually be read into ORA.

On the 10.Output Options index card in P2, scroll down to Additional Output Formats and check the per Map box next to DyNetML for Map Analysis.

Open the Output Storage Manager. (See section V.).

You will see Text Analysis Output Directory listed under the Analysis Output Storage category. Click the Save in folder... button on the far right in that row. In the folder chooser window that pops up, navigate to the folder you want.

Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextOutputFiles

Click Open.

Save these settings and close the Output Storage Manager. (See section V.)

Run the Multiple Map Analysis:

In the Menu Bar, choose Run Analysis, then Multiple Map Analysis.

The results will be displayed on the Semantic Network of Current Text index card in P3 and on the 2.Statistics index card in P4.

Below is a screen capture showing the results of Multiple Map Analysis in P3 and P4:
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 titled output under the directory where Automap is installed.

The output folders are overwritten with every new analysis you run. If you want to save the results of a current analysis folder, simply rename that folder.

**IX. Open ORA**

**On Empire:**

- Navigate to C:\Documents and Settings\Carley\Desktop\Tools.
- Double-click the Run ORA icon. The ORA interface pops up.

From anywhere else: (requires a web connection)

http://www.casos.cs.cmu.edu/projects/ora/software.html

**X. Load your files into ORA**

Click the Load button on the far right. A file chooser window will pop up.

Choose Select Files. Navigate to your NyNetML files (generated in section VIII).

**Example:** C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextOutputFiles

Select the files you want by holding down the key while clicking on each one. Click Open, then Finish.

**Example:** Select Text1-in-user's guide.txt and Text2-in-user's guide.txt
Note: If some file is already loaded into ORA, choose the Append As Additional Meta-Matrix option.

XI. Visualize your meta-matrices

The left-hand portion of the ORA screen is the ORA navigator. In this navigator, select the matrix you want to visualize by clicking on it. Click the Visualize this meta-matrix button near the center of the screen to generate the visualization.

Below are the visualized meta-matrices for Text1-in-user's guide.txt (this page) and Text2-in-user's guide.txt (next page):
XII. Run a Semantic Network Report

In ORA’s Menu Bar, choose Analysis, then Generate Reports. A window will pop up. You will see Select Report at the top of the window. Click the v-shaped icon to the right, and select Semantic Network from the resulting drop-down menu.

Below is a screen capture showing where to find the Semantic Network Report in the Generate Reports pop-up window:
Check the two boxes for Text1-in-user's guide.txt and Text2-in-user's guide.txt inside the field titled **Select one or more meta matrices**.

To specify a filename for the results, click the **Browse** button near the bottom right corner of the **Generate Reports** pop-up window. Navigate to the desired location and type in a filename or use the default name.

**Example:** C:\Documents and Settings\carley\Desktop\AutomapLesson1\SemanticNetworkReport.html

Click the **Next** button, then the **Finish** button at the bottom of the **Generate Reports** pop-up window.

ORA will run. An HTML file will pop up displaying the results.

Return to the ORA window. A small box titled **Complete** has popped up. Click the **OK** button.

Below is the Semantic Network Report generated for Text1-in-user's guide.txt and Text2-in-user's guide.txt:

---

**SEMANTIC-NETWORK REPORT**

Input data: Text1-in-user's guide.txt, Text2-in-user's guide.txt

Start time: Tue Mar 06 10:59:33 2007

This is a comparison of two semantic networks. Each node in the network is considered a Concept, and each edge a Statement connecting two concepts. Statement weights are interpreted as the number of times the statement occured in the underlying input text.
Symmetric difference

The symmetric distance of network A to network B is a new network that contains the entities in A that are not in B.

Saved output networks

The summary statistics across all maps saved to: C:\Documents and Settings\jm cgille_summary_statistics.csv

Produced by ORA developed at CASOS - Carnegie Mellon University
Ontological Text Coding

1. 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

<table>
<thead>
<tr>
<th>Meta-Matrix</th>
<th>Agent</th>
<th>Knowledge</th>
<th>Resource</th>
<th>Task/Event</th>
<th>Organization</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>Social</td>
<td>Knowledge</td>
<td>Capability</td>
<td>Assignment</td>
<td>Membership</td>
<td>Agent location</td>
</tr>
<tr>
<td>Knowledge</td>
<td>Information</td>
<td>Training</td>
<td>Knowledge requirement</td>
<td>Organizational knowledge</td>
<td>Knowledge location</td>
<td></td>
</tr>
<tr>
<td>Resource</td>
<td>Resource</td>
<td>Resource requirement</td>
<td>Organizational capability</td>
<td>Resource location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task/Event</td>
<td>Precedence</td>
<td>Organizational assignment</td>
<td>Task/Event location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Interorganizational</td>
<td>Organizational location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>Proximity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attribute</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 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 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:

<table>
<thead>
<tr>
<th>Setting</th>
<th>Resulting text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Selected - Thesaurus content only (default)</td>
<td>mr . &lt; agent&gt; &lt; agent&gt; dog ate lotus &lt; event&gt; &lt; attribute&gt; am . mrs . &lt; agent&gt; was unhappy with dog . yelled saying &lt; agent resource&gt; impossible dog!&quot; dog kept eating flowers weeds . asked mr . cray stop dog . couldn't . mrs . &lt; agent&gt; planted roses weeded garden . silly dog &lt; attribute&gt; dug up roses looking vole on june &lt; attribute&gt; , &lt; attribute&gt; . weeding was no longer needed . prof . &lt; agent&gt; , mrs . &lt; agent&gt; &lt; attribute&gt; mr . cray met next day &lt; task&gt; plan . john &lt; agent&gt; mrs . &lt; agent&gt; put up scarecrow . thought scare dog . mr . &lt; agent&gt; put up fence . problem solved . then mrs . &lt; agent&gt; planted lotus , &lt; resource&gt; , &lt; resource&gt; , roses .</td>
</tr>
<tr>
<td>Thesaurus content only, Direct Adjacency</td>
<td>. &lt; agent&gt; &lt; agent&gt; &lt; event&gt; &lt; attribute&gt; . . &lt; agent&gt; . &lt; agent resource&gt; . . . . &lt; agent&gt; . &lt; attribute&gt; &lt; attribute&gt; &lt; attribute&gt; . . . .</td>
</tr>
</tbody>
</table>
Then we run Map Analysis on both texts using AutoMap's default Analysis Settings. These are the results:

<table>
<thead>
<tr>
<th>Not Selected - Thesaurus content only (default)</th>
<th>Thesaurus content only, Direct Adjacency</th>
<th>Thesaurus content only, Rhetorical Adjacency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 10 am</td>
<td>1 10 am</td>
<td>1 blossom 10</td>
</tr>
<tr>
<td>1 12 1880</td>
<td>1 12 1880</td>
<td>1 brown was</td>
</tr>
<tr>
<td>1 1880 weeding</td>
<td>1 1880</td>
<td>1 carnations</td>
</tr>
<tr>
<td>1 am mrs</td>
<td>1 am mrs</td>
<td>daffodils</td>
</tr>
<tr>
<td>1 asked mr</td>
<td>1 asked mr</td>
<td>1 concoct plan</td>
</tr>
<tr>
<td>1 ate lotus</td>
<td>1 ate lotus</td>
<td>1 couldn't mrs</td>
</tr>
<tr>
<td>1 blossom 10</td>
<td>1 blossom 10</td>
<td>1 cray stop</td>
</tr>
<tr>
<td>1 brown was</td>
<td>1 brown was</td>
<td>1 cray's brown</td>
</tr>
<tr>
<td>1 carnations</td>
<td>1 carnations</td>
<td>1 craye put up</td>
</tr>
<tr>
<td>daffodils</td>
<td>daffodils</td>
<td>1 craye put up</td>
</tr>
<tr>
<td>1 concoct plan</td>
<td>1 concoct plan</td>
<td>1 daffodils</td>
</tr>
<tr>
<td>1 couldn't mrs</td>
<td>1 couldn't mrs</td>
<td>roses</td>
</tr>
<tr>
<td>1 cray stop</td>
<td>1 cray stop</td>
<td>1 day concot</td>
</tr>
<tr>
<td>1 cray's brown</td>
<td>1 cray's brown</td>
<td>1 dog yelled</td>
</tr>
<tr>
<td>1 craye put up</td>
<td>1 craye put up</td>
<td>1 dog yelled</td>
</tr>
<tr>
<td>1 daffodils roses</td>
<td>1 daffodils</td>
<td>1 dug up</td>
</tr>
<tr>
<td>1 day concot</td>
<td>1 day concot</td>
<td>1 eating flowers</td>
</tr>
<tr>
<td>1 dog yelled</td>
<td>1 dog yelled</td>
<td>1 fence problem</td>
</tr>
<tr>
<td>1 dug up</td>
<td>1 dug up</td>
<td>1 flowers weeds</td>
</tr>
<tr>
<td>1 eating flowers</td>
<td>1 eating</td>
<td>1 garden silly</td>
</tr>
<tr>
<td>1 fence problem</td>
<td>1 fence</td>
<td>1 impossible</td>
</tr>
<tr>
<td>1 flowers weeds</td>
<td>1 flowers</td>
<td>dog</td>
</tr>
<tr>
<td>1 garden silly</td>
<td>1 flowers</td>
<td>1 impossible</td>
</tr>
<tr>
<td>1 impossible</td>
<td>1 fence</td>
<td>dog</td>
</tr>
<tr>
<td>dog</td>
<td>1 problem</td>
<td>1 john darren</td>
</tr>
<tr>
<td>1 john darren</td>
<td>1 flowers</td>
<td>1 john darren</td>
</tr>
<tr>
<td>1 june 12</td>
<td>1 weeds</td>
<td>1 june 12</td>
</tr>
<tr>
<td>1 keeping vole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 lotus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 looking vole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 carnations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 kept eating</td>
<td>1 garden silly</td>
<td>1 met next</td>
</tr>
<tr>
<td>1 longer needed</td>
<td>1 impossible dog</td>
<td>1 mr craye</td>
</tr>
<tr>
<td>1 looking vole</td>
<td>1 john darren</td>
<td>1 needed prof</td>
</tr>
<tr>
<td>1 lotus</td>
<td>1 kept eating</td>
<td>1 next day</td>
</tr>
<tr>
<td>carnations</td>
<td>1 longer</td>
<td>1 mr craye</td>
</tr>
<tr>
<td>1 met next</td>
<td>needed</td>
<td>1 no longer</td>
</tr>
<tr>
<td>1 mr craye</td>
<td>looking vole</td>
<td>1 on june</td>
</tr>
<tr>
<td>1 needed prof</td>
<td>lotus</td>
<td>1 plan john</td>
</tr>
<tr>
<td>1 next day</td>
<td>1 met next</td>
<td>1 planted roses</td>
</tr>
<tr>
<td>1 no longer</td>
<td>1 mr craye</td>
<td>1 problem</td>
</tr>
<tr>
<td>1 on june</td>
<td>1 needed prof</td>
<td>1 prof darren</td>
</tr>
<tr>
<td>1 plan john</td>
<td>1 next day</td>
<td>1 put up</td>
</tr>
<tr>
<td>1 planted roses</td>
<td>1 on june</td>
<td>1 put up fence</td>
</tr>
<tr>
<td>1 problem</td>
<td>1 plan john</td>
<td>1 roses weeded</td>
</tr>
<tr>
<td>solved</td>
<td>1 planted roses</td>
<td>1 saying you</td>
</tr>
<tr>
<td>1 prof darren</td>
<td>1 problem</td>
<td>1 scare dog</td>
</tr>
<tr>
<td>1 put up</td>
<td>solved</td>
<td>1 scarecrow</td>
</tr>
<tr>
<td>1 put up fence</td>
<td>1 prof darren</td>
<td>thought</td>
</tr>
<tr>
<td>1 roses weeded</td>
<td>1 put up</td>
<td>1 silly dog</td>
</tr>
<tr>
<td>1 saying you</td>
<td>1 put up fence</td>
<td>1 solved then</td>
</tr>
<tr>
<td>1 scare dog</td>
<td>1 roses weeded</td>
<td>1 stop dog</td>
</tr>
<tr>
<td>1 scarecrow</td>
<td>1 saying you</td>
<td>1 then mrs</td>
</tr>
<tr>
<td>thought</td>
<td>1 scare dog</td>
<td>1 thought scare</td>
</tr>
<tr>
<td>1 silly dog</td>
<td>1 scarecrow</td>
<td>1 unhappy with</td>
</tr>
<tr>
<td>1 solved then</td>
<td>thought</td>
<td>1 up scarecrow</td>
</tr>
<tr>
<td>1 stop dog</td>
<td>1 silly dog</td>
<td>1 vole on</td>
</tr>
<tr>
<td>1 then mrs</td>
<td>solved then</td>
<td>1 was unhappy</td>
</tr>
<tr>
<td>1 thought scare</td>
<td>1 stop dog</td>
<td>1 weeded</td>
</tr>
<tr>
<td>1 unhappy with</td>
<td>1 then mrs</td>
<td>garden</td>
</tr>
<tr>
<td>1 up scarecrow</td>
<td>1 thought scare</td>
<td>1 weeding was</td>
</tr>
<tr>
<td>1 vole on</td>
<td>1 unhappy with</td>
<td>1 weeds asked</td>
</tr>
<tr>
<td>1 was unhappy</td>
<td>1 up scarecrow</td>
<td>1 with dog</td>
</tr>
<tr>
<td>1 weeded</td>
<td>1 vole on</td>
<td>1 yelled saying</td>
</tr>
<tr>
<td>garden</td>
<td>1 was unhappy</td>
<td>1 you</td>
</tr>
<tr>
<td>1 weeding was</td>
<td>1 weeded</td>
<td>impossible</td>
</tr>
<tr>
<td>1 weeds asked</td>
<td>garden</td>
<td>2 brown planted</td>
</tr>
<tr>
<td>1 with dog</td>
<td>1 weeding was</td>
<td>2 darren mrs</td>
</tr>
<tr>
<td>1 yelled saying</td>
<td>1 weeds asked</td>
<td>2 mr cray</td>
</tr>
<tr>
<td>1 you</td>
<td>1 with dog</td>
<td>5 mrs brown</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stat</th>
<th>File:</th>
<th>File:</th>
<th>File:</th>
</tr>
</thead>
<tbody>
<tr>
<td># of concepts analyzed:</td>
<td># of concepts analyzed:</td>
<td># of concepts analyzed:</td>
<td></td>
</tr>
<tr>
<td>unique: 62</td>
<td>unique: 62</td>
<td>unique: 62</td>
<td></td>
</tr>
<tr>
<td>total: 88</td>
<td>total: 88</td>
<td>total: 88</td>
<td></td>
</tr>
</tbody>
</table>
### 2. 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](http://www.casos.cs.cmu.edu/projects/automap/publications.html)). 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](http://www.casos.cs.cmu.edu/projects/automap/publications.html). 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 entities and link into statements. An example would be the category "attribute", which would represent information that is inherent to a certain entity. 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&RDquo;** folder just rename the folder.

2.1 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. Cray 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:

<table>
<thead>
<tr>
<th>Setting</th>
<th>Resulting text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select: Knowledge/Agent/Resource</td>
<td>. dog lotus.. dog. dog flowers.. dog... roses. dog roses,... ..... dog..... lotus, carnations, daffodils, roses.</td>
</tr>
<tr>
<td># of concepts analyzed:</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td></td>
</tr>
<tr>
<td>unique: 72</td>
<td></td>
</tr>
<tr>
<td>total: 118</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># of concepts analyzed:</th>
</tr>
</thead>
<tbody>
<tr>
<td>unique: 68</td>
</tr>
<tr>
<td>total: 107</td>
</tr>
</tbody>
</table>

1. roses looking
2. saying you
3. scare crow
4. scarecrow she
5. she yelled
6. silly dog
7. solved then
8. stop the
9. the silly
10. mrs.
11. thought it
12. to stop
13. unhappy with
14. up the
15. vole on
16. was unhappy
17. weeded the
18. weeding was
19. weeds she
20. with the
21. would scare
22. yelled at
23. you impossible
24. brown planted
25. mr craye
26. the dog
27. mrs brown
<table>
<thead>
<tr>
<th># of concepts in statements:</th>
<th># of concepts in statements:</th>
</tr>
</thead>
<tbody>
<tr>
<td>unique: 108</td>
<td>unique: 98</td>
</tr>
<tr>
<td>total: 117</td>
<td>total: 106</td>
</tr>
<tr>
<td># of isolated concepts:</td>
<td># of isolated concepts:</td>
</tr>
<tr>
<td>unique: 0</td>
<td>unique: 0</td>
</tr>
<tr>
<td>total: 0</td>
<td>total: 0</td>
</tr>
<tr>
<td># of statements:</td>
<td># of statements:</td>
</tr>
<tr>
<td>unique: 108</td>
<td>unique: 98</td>
</tr>
<tr>
<td>total: 117</td>
<td>total: 106</td>
</tr>
<tr>
<td>Density (based on Statements):</td>
<td>Density (based on Statements):</td>
</tr>
<tr>
<td>unique: 1.5</td>
<td>unique: 1.44</td>
</tr>
<tr>
<td>total: 1.63</td>
<td>total: 1.56</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Output Type</th>
<th>Name of output</th>
<th>Content of output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts analyzed</td>
<td>List of concept analyzed.csv</td>
<td>Concept, Text, Frequency</td>
</tr>
<tr>
<td></td>
<td>Statistics of concepts analyzed.csv</td>
<td>Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts</td>
</tr>
<tr>
<td>Term Distribution List</td>
<td>List of concept in statements.csv</td>
<td>Concept, Text, Frequency</td>
</tr>
<tr>
<td></td>
<td>Statistics of concept in statements.csv</td>
<td>Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts</td>
</tr>
<tr>
<td>Concepts in statements and isolates</td>
<td>List of isolates in statements.csv</td>
<td>Concept, Text, Frequency</td>
</tr>
<tr>
<td></td>
<td>Statistics of isolates.csv</td>
<td>Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts</td>
</tr>
<tr>
<td>Statements</td>
<td>List of statements.csv</td>
<td>Statement, Text, Frequency</td>
</tr>
<tr>
<td></td>
<td>Statistics of statements.csv</td>
<td>Statement, Cumulated sum across text set, Number of text statement occurs in, Percentage of texts statement occurs in, Texts</td>
</tr>
<tr>
<td>Concepts in statements by concepts in statements:</td>
<td>Matrix of concepts in statements by concepts in statements.csv</td>
<td>Matrix of Concept that were linked into statements (first row) by Concept that were linked into statements (first column)</td>
</tr>
<tr>
<td>Term Distribution Matrices</td>
<td>Matrix of concepts in statements.csv</td>
<td>If count was chosen, cells contain cumulated frequency of concept</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If binary was chosen, cells denote existence (1) or</td>
</tr>
</tbody>
</table>
### 1.2 Example for Term Distribution List and Matrices

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

**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.
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 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 > span>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 use 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 entities, 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 entities and link into statements. An example would be the category "attribute", which represent information that is inherent to a certain entity. 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

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

**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.

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 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
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**.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Concept 1</th>
<th>Concept 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>mrs</td>
<td>brown</td>
</tr>
<tr>
<td>2</td>
<td>mr</td>
<td>craye</td>
</tr>
<tr>
<td>2</td>
<td>darren</td>
<td>mrs</td>
</tr>
<tr>
<td>2</td>
<td>brown</td>
<td>planted</td>
</tr>
<tr>
<td>1</td>
<td>you</td>
<td>impossible</td>
</tr>
<tr>
<td>1</td>
<td>yelled</td>
<td>saying</td>
</tr>
<tr>
<td>1</td>
<td>with</td>
<td>dog</td>
</tr>
<tr>
<td>1</td>
<td>weeds</td>
<td>asked</td>
</tr>
<tr>
<td>1</td>
<td>weeding</td>
<td>was</td>
</tr>
<tr>
<td>1</td>
<td>weeded</td>
<td>garden</td>
</tr>
<tr>
<td>1</td>
<td>was</td>
<td>unhappy</td>
</tr>
<tr>
<td>1</td>
<td>vole</td>
<td>on</td>
</tr>
<tr>
<td>1</td>
<td>up</td>
<td>scarecrow</td>
</tr>
<tr>
<td>1</td>
<td>unhappy</td>
<td>with</td>
</tr>
<tr>
<td>1</td>
<td>thought</td>
<td>scare</td>
</tr>
<tr>
<td>1</td>
<td>then</td>
<td>mrs</td>
</tr>
<tr>
<td>1</td>
<td>stop</td>
<td>dog</td>
</tr>
<tr>
<td>1</td>
<td>solved</td>
<td>then</td>
</tr>
</tbody>
</table>

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_analyticx_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:
<table>
<thead>
<tr>
<th>Name of measure</th>
<th>Calculation of measure</th>
<th>Name of and reference for corresponding SNA measure</th>
<th>Corresponding name of measure in Cube</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept outdegree</td>
<td>Total number of statements with concept in anterior position</td>
<td>Outdegree, expansiveness, actor degree centrality (Wasserman &amp; Faust 1994); Prestige, Influence (Mrvar)</td>
<td>Local imageability</td>
</tr>
<tr>
<td>Concept indegree</td>
<td>Total number of statements with concept in posterior position</td>
<td>Indegree, receptivity, popularity, actor degree prestige (Wasserman &amp; Faust 1994); Prestige, Support (Mrvar)</td>
<td>Local evokability</td>
</tr>
<tr>
<td>Concept outdegree centrality</td>
<td>Total number of statements with concept in anterior position/ Number of unique concepts in text</td>
<td>Outdegree Centrality (Wasserman &amp; Faust 1994); Relative Influence (Mrvar)</td>
<td>None</td>
</tr>
<tr>
<td>Concept indegree centrality</td>
<td>Total number of statements with concept in posterior position/ Number of unique concepts per text</td>
<td>Indegree Centrality (Wasserman &amp; Faust 1994); Relative Support (Mrvar)</td>
<td>None</td>
</tr>
<tr>
<td>Total degree</td>
<td>Concept indegree + concept outdegree</td>
<td>-</td>
<td>Local density</td>
</tr>
</tbody>
</table>

**Map (graph) level measures, direct connectivity**

<table>
<thead>
<tr>
<th>Text outdegree centrality</th>
<th>Sum (largest observed outdegree – outdegree of concepts)/( Number of unique concepts in text)</th>
<th>Group outdegree centralization (Wasserman &amp; Faust, 1994)</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean concept outdegree centrality</td>
<td>Sum (outdegree)/ Number of unique concepts in text</td>
<td>Mean outdegree (= Mean indegree) (Wasserman &amp; Faust 1994)</td>
<td>None</td>
</tr>
<tr>
<td>Variance of concept outdegree centrality</td>
<td>Sum(sum outdegree – mean outdegree)^2 / Number of concepts</td>
<td>Variance of outdegree (Wasserman &amp; Faust 1994)</td>
<td>None</td>
</tr>
</tbody>
</table>
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:

<table>
<thead>
<tr>
<th>Entry</th>
<th>Entry</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>File:</td>
<td></td>
<td>Name of the analyzed text file.</td>
</tr>
<tr>
<td></td>
<td>unique:</td>
<td>total:</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td><strong># of concepts analyzed:</strong></td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong># of concepts in statements:</strong></td>
<td>Only concepts are considered that linked into statements.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong># of isolated concepts:</strong></td>
<td>Only concepts are considered that did not link into statements.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong># of statements:</strong></td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>density (based on statements)</strong></td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Analysis Settings</strong></td>
<td>Punctuation:</td>
<td></td>
</tr>
</tbody>
</table>
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.
Entropy

Entropy is the formalization of redundancy and diversity. Thus we say that Information Entropy (H) of a text document (X) where probability p of a word x = ratio of total frequency of x to length (total number of words) of a text document.

\[ H(X) = - \sum_{x \in X} p(x) \log_2 p(x) \]

To run Entropy you will need to load multiple text documents into AutoMap for analysis.

File > Open Multiple Files > Select Folder.

Navigate to the folder containing the texts you wish to analyze, and single click on it.

In the menu bar, select option Run Analysis

Select the Multiple Map Analysis option as shown in the diagram below:

```
Run Analysis  Tools  Help
Single Map Analysis
Multiple Map Analysis
Single Meta Matrix Text Analysis
Multiple Meta Matrix Text Analysis
Single Sub Matrix Text Analysis
Multiple Sub Matrix Text Analysis
```

Once you have selected the Multiple Map Analysis, go to Panel 2, tab no. 4. Analysis Settings, then select Tab no. 2. Output Options.

You will find the information in an excel file, which will be located in the last column of the stats_overall.csv output file.

Reference

Theory:


Empiric application:

CASOS Email Parser (CEMAP)

1. Instructions for the using the CASOS Email Parser (CEMAP) through AutoMap

Introduction

The CASOS Email Parser (CEMAP) that is launched through AutoMap (http://www.casos.cs.cmu.edu/projects/automap/software.html) enables the extraction of different types of network information from emails (e.g. who exchanges information, who provides what information, etc.). The following image shows what types of information can be extracted with CEMAP.

Network type 1 (Social network) represents social network data that can be extracted from email headers (e.g. agent-agent networks). In SN, entities represent people, and edges represent exchanged emails (frequency count). This network type does not require any text coding in AutoMap.

Networks types 2 and 3 are extracted by performing text coding in AutoMap. More specifically, Knowledge networks (KK) represent semantic network or mental models that are contained in the bodies of individual emails. In KK, entities represent knowledge items, and edges represent the co-occurrence of terms in text. For MM, texts are coded in AutoMap according to a taxonomy or ontology (e.g. meta-matrix, while are ontologies can be specified by the user). In MM, entities represent instances of categories (e.g. agent, knowledge, resources) of the ontology, and edges represent co-occurrences of terms in texts.

Networks types 4 and 5 result from the combination of SN with KK and MM, respectively. In type 4, entities represent people and knowledge, and edges represent emails and mental models. In type 5, nodes represent the categories of the taxonomy as specified by the user, and the edges represent the co-occurrence of the terms that represent instances of the taxonomy in the corpus. For the creation of type 4 and 5, the extraction of type 2 and 3, respectively, is mandatory.

CEMAP stores all network data as DyNetML files (a derivate of XML). This data can be analyzed with any package that reads DyNetML. We recommend using the ORA software for this purpose (http://www.casos.cs.cmu.edu/projects/ora/index.html).

How to do it:

1. Extraction of network type 1 data:
   - 1.1. This network type is needed to perform social network analysis (SNA) based on header information only.
1.2. In CEMAP, **Analysis Type**, select the type of your mail box.

1.3. In CEMAP, **Input Data Set**, specify your email client, inbox type, username, and password as semicolon separated values, *e.g.* `pop.gmail.com;INBOX;JohnDoe;JohnDoesPassword`

1.4. In CEMAP, **Enter Destination Directory for Texts**, put in a directory of your choice, *e.g.* `C:\textInput`. The email bodies will be stored in AutoMap format (.txt) under this directory. Make sure that this directory does exits on your machine.

1.5. In CEMAP, **Enter Destination File for Network**, put in a directory AND filename of your choice. The filename extension needs to be .xml. *e.g.* `C:\output\AA.xml`.

1.6. You are now ready to extract network type 1. In order to do so, press the **Start Extraction** button. As a result, CEMAP will extract network data from the email headers (*e.g. agent by agent*) and store it as one XML file under the directory that is specified in the **Destination File for Network**.

1.7. You can pick up the output file and load it into the SNA package of your choice. We recommend using the **ORA package** ([http://www.casos.cs.cmu.edu/projects/ora/download.php](http://www.casos.cs.cmu.edu/projects/ora/download.php)).

To extract networks of type 2, perform Semantic Network Analysis in AutoMap.

To extract networks of type 3, perform Meta-Matrix Text Analysis in AutoMap.

To extract networks of type 4, merge networks of type one and 2.

To extract networks of type 5, merge networks of type one and 3.

CEMAP is a feature in AutoMap that will import email files into AutoMap to analyze the text in the emails, such as who exchanges information, who provides what information, etc.

**To use the feature:**

- In AutoMap menu bar, select **Tools > Open CASOS Email Parser (CEMAP)**.

Once that is selected the **CASOS — Email Analyzer** window will appear. The Email Analyzer window is where you will type in your information of what type of network you would like to create while importing your emails.

**The different types of imports are**
Select the type of Format your email is in, from the drop down menu.

- **MBOX** — File that is stored in standard file folders
- **POP3 and IMAP** — Emails that are on a server
  - Will require Username, User password, Server Name, and Folder Name
- **TXT** — Exported email files from outlook
  - Will require a Data (file) set identifier
- **Dynetml** — Currently under development

Once you have selected what type of format you would like to use, click on the **Start Extraction** button. This will begin the process of importing your emails.
Support

How to cite AutoMap


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


**Where to learn more about Dynamic Network Analysis**


**Entropy references:**

**Theory:**

Empiric application:

Automap

Lessons
Lesson 1

From content analysis to semantic networks

I. Open AutoMap

On Empire:

- Navigate to C: \ Documents and Settings \ Carley \ Desktop \ Tools.
- Double-click the Run Automap icon. The Automap Graphical User Interface (GUI) pops up.

From anywhere else: (requires a web connection)

http://www.casos.cs.cmu.edu/projects/automap/software.html

II. Overview of the AutoMap Graphical User Interface (GUI)

Below is a screen capture of the AutoMap GUI:

The AutoMap GUI is divided into four primary quadrants, as shown above, or panels: P1 (top left), P2 (bottom left), P3 (top right), and P4 (bottom right).

A drop-down menu bar at the top of the window will provide access to analysis tools and utilities. The Browse Menu allows you to quickly navigate between texts you have loaded into Automap.

Panel P2 can be edited, but the remaining panels (P1, P3, P4) cannot. The information that is displayed on each panel is always related to the information displayed in the other three panels. The Browse Menu, however, relates to all four panels at the same time.
III. Load multiple text documents into AutoMap

In order to analyze multiple pieces of text simultaneously, you will need to save the documents together in one folder.

**Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextInputFiles**

- In the AutoMap menu bar, select **File > Open Multiple Files**. A folder selection window will pop up.
- Navigate to the folder containing the documents you wish to analyze, and single-click on it.

**Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\TextInputFiles**

- Click the Open button in the bottom right corner.

The first text will be displayed in **panel P1** under the tab titled **1.Original Texts**. You can use the Browse Menu to navigate through the texts.

Below is a screen capture of the first text display:

![Screen capture of AutoMap panel P1 displaying original text]

The Browse Menu allows you to quickly navigate from one text sample to the next.

IV. Carry out a simple content analysis

A simple content analysis determines the frequencies of all words in a text document.

**Examine the Concept List:**

The Concept List considers concepts in each text file individually. It is created automatically when a document or set of documents is loaded into Automap. This will be displayed in the
**P2 panel** under the tab titled **1. Concept List**.

The Concept List will tell you several important facts about each document in your set:

- The number of concepts found in the text displayed in **P1** under the tab titled **1. Original Text**
- Related frequencies of those concepts
- Cumulated unique concepts and total concepts contained in the data set

**Note:** The number of unique concepts considers each concept only once. The number of total concepts considers repetitions of concepts.

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

Below is a screen capture showing part of the Concept List, sorted by frequency, for the first document in our loaded folder of texts (Text1-in-user's guide.txt):

**Create and refresh the Union Concept List:**

The Union Concept List considers concepts in all texts in a set (unlike the Concept List, which considers each text in the set individually). It is located in **Panel P2** under the tab titled **2. Union Concept List**.

The Union Concept List tells you several important facts about your text set:

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

Before viewing your Union Concept List, you must refresh it. To refresh your Union Concept List, go to the Menu Bar, choose **File > Refresh Union Concept List.**

The Union Concept List can be refreshed after each step of pre-processing. This will allow you to visualize the impact of pre-processing operations on the union of concepts.

Below is a screen capture showing where to find the **Refresh Union Concept List** command:

![Refresh Union Concept List](image)

After refreshing your Union Concept List, you can view it under the **2. Union Concept List tab in Panel P2.** Below is a screen capture showing where this tab is located in the GUI:

![2. Union Concept List tab in Panel P2](image)

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

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

V. Save your Union Concept List as a csv file

To specify a location for your Union Concept List file, choose **File > Output Storage Manager.**

Below is a screen capture showing where to find this in the GUI:
A window will pop up listing all pre-processing output storage. This is the Output Storage Manager.

The Union Concept List appears as the second item in the Output Storage Manager. Click the **Save file as...** option to the right of the default file pathway. In the file chooser window that pops up, browse to the pathway you want.

**Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\ UnionConceptList**

Below is a screen capture showing where to find the correct **Save file as...** button in the Output Storage Manager:
Click the **Save Settings** button at the bottom right corner of the Output Storage Manager. Close the Output Storage Manager window by clicking on the **red X** at the top right corner.

In the Automap Menu Bar, choose **File > Save Union Concept List**.

**VI. View your csv file in Microsoft Excel**

Navigate to the location of your new csv file and double-click on its icon.

*Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\ UnionConceptList*

A box titled **“Open With”** pops up. Click on **Microsoft Excel**, then click **OK**.

**Note:** You must have MS Excel installed on your computer in order to view your csv file this way.

**VII. Save and view your results as a binary csv file**

Follow Steps 3 and 4.

**VIII. Carry out a simple Semantic Network Analysis**

A **Simple Network Analysis** involves no thesauri and uses all words. A **Semantic Network Analysis** can be run on original input texts or on texts that have been pre-processed with a Delete List and/or a Generalization Thesaurus.

**Before running the analysis:**
Make sure you have completed the following steps:

- Pre-process the texts if you wish
- Specify the analysis settings (otherwise the default settings will be applied)
- Specify additional output options if you wish to (if not, the default outputs will be generated)

**About the analysis settings:**

Use the **Analysis Settings** index card to specify the analysis settings. Your settings will be automatically applied to the analysis (you do not need to confirm them).

Below is a screen capture showing where to find the **Analysis Settings** index card in **P2**:

![Analysis Settings Index Card](image)

The following chart lists the possible analysis settings:

<table>
<thead>
<tr>
<th>Coding Ties</th>
<th>Directionality</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Specify the way</strong></td>
<td>Select one of the following two possibilities by checking the button.</td>
<td>Strength will be printed out by default. To not print out Strength uncheck the item.</td>
</tr>
<tr>
<td><strong>Directionality</strong></td>
<td>Uni-Directional (When coding a tie, only 1st-&gt;2nd concept should be noted)</td>
<td>Frequency (The cumulative frequency of every existing statement.)</td>
</tr>
<tr>
<td></td>
<td>Bi-Directional (When coding a tie, both 1st &lt;--&gt; 2nd concept shall be noted)</td>
<td><strong>Item not checked:</strong> Existence of frequency will be printed out (binary result).</td>
</tr>
</tbody>
</table>
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 relationship. Only concepts in same window can form statements.

If you do not want to change any of the suggested options, the analysis will be done with a set of standard (default) settings.

Below is a screen capture showing the standard settings:

About the output options:

The map and the statistic output generated by AutoMap are displayed in Panels P3 and P4, respectively. In addition, AutoMap offers further output options that
can be selected in the **10.Output Options index card in Panel P2**. Any other additional outputs are generated after analyses are run.

Below is a screen capture showing the **10.Output Options index card in P2**:

![Output Options index card in Panel P2](image)

For all types of multiple analysis, term distribution lists and matrices can be selected as output options in the upper two fields of the **10.Output Options index card in P2**.

**Points to consider:**

- No list or matrix is generated by default.
- Lists or matrices are only generated if the user checks the item wanted 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 are applied, analysis will run on the highest stage of pre-processing.
- The requested lists and matrices are automatically saved in a folder called **Term Distribution Lists and Matrices** under the root directory of AutoMap. This folder is overwritten with new analysis you run. If you want to save the results somewhere other than the default Term Distribution Lists and Matrices folder, simply rename that folder.
- Two output lists are generated for each Term Distribution List checked.
- One output matrix is generated for each Term Distribution Matrix checked.

The following chart lists the types and content of Term Distribution Lists and Matrices:

<table>
<thead>
<tr>
<th>Output Type</th>
<th>Name of Content of output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Term Distribution List</td>
<td>output</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Concepts analyzed</td>
<td>List of concept analyzed.csv</td>
</tr>
<tr>
<td></td>
<td>Statistics of concepts analyzed.csv</td>
</tr>
<tr>
<td>Concepts in statements and isolates</td>
<td>List of concept in statements.csv</td>
</tr>
<tr>
<td></td>
<td>Statistics of concept in statements.csv</td>
</tr>
<tr>
<td></td>
<td>List of isolates in statements.csv</td>
</tr>
<tr>
<td></td>
<td>Statistics of isolates.csv</td>
</tr>
<tr>
<td>Statements</td>
<td>List of statements.csv</td>
</tr>
<tr>
<td></td>
<td>Statistics of statements.csv</td>
</tr>
<tr>
<td>Concepts in statements by concepts in statements:</td>
<td></td>
</tr>
<tr>
<td>Concepts in statements</td>
<td>Matrix of concepts in statements.csv</td>
</tr>
<tr>
<td></td>
<td>Matrix of Concept (union of concepts listed in first row) by texts (all text names listed in first column)</td>
</tr>
<tr>
<td>Term(s) by text(s):</td>
<td></td>
</tr>
<tr>
<td>Concepts analyzed</td>
<td>Matrix of concept analyzed.csv</td>
</tr>
</tbody>
</table>
Save your upcoming analysis in DyNetML format:

Before running your Multiple Map Analysis, you must specify that your results will be in DyNetML format so that the results can eventually be read into ORA.

On the **Output Options** index card in Panel P2, scroll down to the option titled “Additional Output Formats” and check the per **Map box** next to “DyNetML for Map Analysis.”

Open the Output Storage Manager. (See section V.)

You will see “Text Analysis Output Directory” listed under the “Analysis Output Storage” category. Click the **Save in folder...** button on the far right in that row. In the folder chooser window that pops up, navigate to the folder you want.

**Example: C:\Documents and Settings\carley\Desktop\AutomapLesson1\ TextOutputFiles**

Click **Open**.

Save these settings and close the Output Storage Manager. (See section V.)

**Run the Multiple Map Analysis:**
In the Menu Bar, choose Run Analysis > Multiple Map Analysis.

The results will be displayed on the Semantic Network of Current Text index card in Panel P3 and on the Statistics index card in Panel P4.

Below is a screen capture showing the results of Multiple Map Analysis in Panels P3 and P4:

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 titled “output” under the directory where Automap is installed.

The output folders are overwritten with every new analysis you run. If you want to save the results of a current analysis folder, simply rename that folder.

IX. Open ORA

On Empire:

- Navigate to C:\Documents and Settings\Carley\Desktop\Tools.
- Double-click the Run ORA icon. The ORA interface pops up.

From anywhere else: (requires a web connection)

http://www.casos.cs.cmu.edu/projects/ora/software.html

X. Load your files into ORA

Click the Load button on the far right. A file chooser window will pop up.
Choose **Select Files**. Navigate to your **NyNetML** files *(generated in section VIII)*.

**Example:** C:\Documents and Settings\carley\Desktop\AutomapLesson1\ TextOutputFiles

Select the files you want by holding down the `<ctrl>` key while clicking on each one. **Click Open, then Finish.**

**Example:**
1. Select Text1-in-user's guide.txt
2. Press the Control key
3. Select Text2-in-user's guide.txt

**Note:** If some file is already loaded into ORA, choose the **Append As Additional Meta-Matrix** option.

**XI. Visualize your meta-matrices**

The left-hand portion of the ORA screen is the ORA navigator. In this navigator, select the matrix you want to visualize by clicking on it. Click the **Visualize this meta-matrix** button near the center of the screen to generate the visualization. Below are the visualized meta-matrices for Text1-in-user's guide.txt and Text2-in-user's guide.txt:

![Text1-in-user's guide.txt](image1)

![Text2-in-user's guide.txt](image2)
These meta-matrices also can be found in:
C:\Documents and Settings\carley\Desktop\AutomapLesson1\MetaMatrices

XII. Run a Semantic Network Report

In ORA's Menu Bar, choose Analysis > Generate Reports. A window will pop up. You will see “Select Report” at the top of the window. Click the v-shaped icon to the right, and select Semantic Network from the resulting drop-down menu.

Below is a screen capture showing where to find the Semantic Network Report in the Generate Reports pop-up window:
Check the two boxes for **Text1-in-user's guide.txt** and **Text2-in-user's guide.txt** inside the field titled "Select one or more meta matrices".

To specify a filename for the results, click the **Browse** button near the bottom right corner of the **Generate Reports** pop-up window. Navigate to the desired location and type in a filename or you can use the default name.

**Example:** C:\Documents and Settings\carley\Desktop\AutomapLesson1\SemanticNetworkReport.html

Click the **Next** button, then the **Finish** button at the bottom of the Generate Reports pop-up window.

ORA will run. An HTML file will pop up displaying the results.

Return to the ORA window. When a small box titled "Complete" has popped up, click the OK button.

Below is the Semantic Network Report generated for Text1-in-user's guide.txt and Text2-in-user's guide.txt:

**SEMANTIC-NETWORK REPORT**

Input data: Text1-in-user's guide.txt, Text2-in-user's guide.txt

Start time: Tue Mar 06 10:59:33 2007

This is a comparison of two semantic networks. Each node in the network is considered a Concept, and each edge a Statement connecting two concepts. Statement weights are interpreted as the number of times the statement occurred in the underlying input text.
Symmetric difference

The symmetric distance of network A to network B is a new network that contains the entities in A that are not in B.

Statements

<table>
<thead>
<tr>
<th>Network</th>
<th>Concepts</th>
<th>Statements</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text1-in-user's guide.txt</td>
<td>74</td>
<td>200</td>
<td>0.0370233</td>
</tr>
<tr>
<td>Text2-in-user's guide.txt</td>
<td>69</td>
<td>184</td>
<td>0.0392157</td>
</tr>
<tr>
<td>Union</td>
<td>98</td>
<td>310</td>
<td>0.032611</td>
</tr>
<tr>
<td>Intersection</td>
<td>45</td>
<td>8</td>
<td>0.0040404</td>
</tr>
</tbody>
</table>

Saved output networks

The summary statistics across all maps saved to: C:\Documents and Settings\jmcgille_summary_statistics.csv

Produced by ORA developed at CASOS - Carnegie Mellon University

This HTML file also can be found in: ??????????
Lesson 2

Processing data: using the delete list

I. Open AutoMap

On Empire:

- Navigate to C:\Documents and Settings\Carley\Desktop\Tools.
- Double-click the Run Automap icon. The Automap Graphical User Interface (GUI) pops up.

From anywhere else: (requires a web connection)

http://www.casos.cs.cmu.edu/projects/automap/software.html

II. Overview of the AutoMap Graphical User Interface (GUI)

Below is a screen capture of the AutoMap GUI:

The AutoMap GUI is divided into four primary quadrants, or panels: P1 (top left), P2 (bottom left), P3 (top right), and P4 (bottom right).

A drop-down menu bar at the top of the window provides access to analysis tools and utilities. The Browse Menu allows you to quickly navigate between texts you have loaded into Automap.

The P2 panel can be edited, but the other remaining panels (P1, P3, P4) cannot. Information displayed on each panel is always related to the information displayed in the other three panels. The Browse Menu relates to all four panels at the same time.
III. Load multiple text documents into AutoMap

In order to analyze multiple texts simultaneously, you will need to store the documents together in one folder.

**Example: C:\Documents and Settings\carley\Desktop\TextInputFiles02-01-07.**

- In the AutoMap menu bar, select **File > Open Multiple Files**. A folder chooser window will pop up.
- Navigate to the folder containing the documents you wish to analyze, and single-click on it.

**Example: C:\Documents and Settings\carley\Desktop\TextInputFiles02-01-07**

- Click the **Open** button in the bottom right corner.

The first text will be displayed in panel **P1** under the tab titled **1.Original Texts**. You can use the Browse Menu to flip through the documents.

Below is a screen capture of the first text display:

![Screen capture of first text display](image_url)

IV. Carry out a simple content analysis

A simple content analysis determines the frequencies of all words in a text document.
Examine the Concept List:

The Concept List considers concepts in each text file individually. It is created automatically when a text or set of texts is loaded into Automap, and is displayed in panel P2 under the tab titled 1.Concept List.

The Concept List tells you several important facts about each text in your set:

- The number of concepts found in the text displayed in panel P1 under the tab titled 1.Original Text
- Related frequencies of those concepts
- Cumulated unique concepts and total concepts contained in the data set

Note: The number of unique concepts considers each concept only once. 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. To resort the list by frequency, click on the second-column header Frequency.

Below is a screen capture showing part of the Concept List, sorted by frequency, for the first document in our loaded folder of texts (Text1-in-user's guide.txt):

Create and refresh the Union Concept List:

The Union Concept List considers concepts across all texts in a set (unlike the Concept List, which considers each text in the set individually). It is located in panel P2 under the tab titled 2.Union Concept List.
The Union Concept List tells you several important facts about your text set:

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

Before viewing your Union Concept List, you must refresh it. From the Menu Bar, choose **File > Refresh Union Concept List**.

The Union Concept List can be refreshed after each step of pre-processing to visualize the impact of pre-processing operations on the union of concepts.

Below is a screen capture showing where to find the **Refresh Union Concept List** command:

![Refresh Union Concept List](image)

After refreshing your Union Concept List, you can view it under the **2. Union Concept List** tab in **panel P2**. Below is a screen capture showing where this tab is located in the GUI:
By default, the list is sorted by decreasing frequency of concepts. To sort the list alphabetically, click on the first-column header Concept. To re-sort the list by Frequency, click on the second-header column Frequency.

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

**V. The basics of pre-processing in AutoMap**

Now you are ready to pre-process the data, or reduce it so that it includes only terms which are relevant to you.

Here are some points to consider:

- Semantic Network Analysis can be run without any prior data pre-processing
- Meta-Matrix Text Analysis and Sub-Matrix Text Analysis require pre-processing

**VI. The basics of delete lists**

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).

For 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

The Delete List is **NOT** case sensitive.
You can use the predefined Delete Lists that AutoMap offers, or create your own Delete List. All Delete Lists can be edited.

**VII. Open a Delete List**

Select the **File** menu, then **Open Delete List**. Choose one of the following options:

- **Open from file**: A file chooser appears. Select a delete list and click Open
- **Open small predefined Delete List**: This opens AutoMap’s predefined small delete list
- **Open extensive predefined Delete List**: This opens AutoMap’s extensive small delete list

Below is a screen capture showing where to access the Delete List utility:

The Delete List is located in **panel P2** in the AutoMap GUI, under the tab titled **3. Pre-Processing Settings**. It appears as a sub-tab titled **3. Delete List**.

Below is a screen capture of this location:
The Delete List can be edited by clicking inside the panel where the delete terms are listed in the **3.Delete List** index card.

The types of Delete Lists available in AutoMap are:

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

- Extensive predefined Delete List
  This 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.

**VIII. Delete concepts and examine the effects on the text and concept list**

**Note:** If you wish to apply a Delete List and a Thesaurus, we recommend applying the Delete List first, followed by the Thesaurus.

When applying a delete list, AutoMap:

- Searches the text or texts for concepts specified in the Delete List
- Deletes matching concepts from the texts
- Displays the resulting texts in **P1** under the **4.Texts After Deletion** tab

Follow these steps to apply a delete list:

1. Choose an adjacency option on the **3.Delete List** index card. The types of adjacency are the following:
   - Direct Adjacency
If this option is chosen, concepts in the text that match concepts specified in the delete list are removed from the texts. As a result, concepts on either side of the deleted concept move together and are treated as directly adjacent to each other for further analysis and visualization.

- **Rhetorical Adjacency**
  If this option is chosen, concepts in the text that match concepts specified in the delete list are removed from the text, but placeholders ("xxx") are inserted to hold their places. The placeholders retain the original distances of the maintained concepts. This is helpful for visual analysis.

If you do not change the adjacency option, AutoMap uses direct adjacency as the default.

2. Click the **Apply Delete List** button on the **3.Delete List** index card. This deletes the concepts that correspond to concepts specified in the Delete List from all loaded texts.

3. View your newly pre-processed texts in **panel P1** under the **4.Texts After Deletion** tab.

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

**IX. Un-apply a Delete List**

To un-apply a Delete List that was just applied to the data, go to the **3.Delete List** index card and click the **Un-Apply Delete List** button. In **panel P2**, the **4.Texts After Deletion** index card will be cleared.

**X. Save an applied Delete List**

To specify a location for your Union Concept List file, choose **File** in the Menu Bar, then **Output Storage Manager**.

Below is a screen capture showing where to find this in the GUI:

![Output Storage Manager](image)

A window will pop up listing all pre-processing output storage. This is the Output Storage Manager.

The Applied Delete List appears in the list on the left-hand side of the Output Storage Manager. Click the **Save file as...** button to the right of the default file pathway. In the file chooser window that pops up, browse to the pathway you want and type in a filename.

**Example: C:\Documents and Settings\carley\Desktop\Lesson 2\SampleDeleteList**
Click the **Save Settings** button at the bottom right corner of the Output Storage Manager. Close the Output Storage Manager window by clicking on the **red X** at the top right corner.

In the Automap Menu Bar, choose **File > Save Delete List**. This saves the Delete List in the location you specified in the Output Storage Manager.

**XI. Save text(s) after application of Delete List**

To determine where your saved texts will end up, choose **File > Output Storage Manager** (as you did in the previous step).

A window will pop up listing all pre-processing output storage. This is the Output Storage Manager.

Find the Texts After Deletion item in the list on the left-hand side of the Output Storage Manager. Click the Save file as... button to the right of the default file pathway. In the file chooser window that pops up, browse to the pathway you want.

**Example: C:\Documents and Settings\carley\Desktop\Lesson 2\TextsAfterDeletion**

Click the **Save Settings** button at the bottom right corner of the Output Storage Manager. Close the Output Storage Manager window by clicking on the **red X** at the top right corner.

In the Automap Menu Bar, choose **File > Save Text(s) after Delete List Applied**. This saves your modified texts in the location you chose in the Output Storage Manager.

**XII. Modify a Delete List**

By clicking inside the text field in the **3. Delete List** index card, you can:

- Add concepts: Press the <Enter> key after typing in a concept so that there is only one concept per line
- Modify concepts: Go to the desired line and retype the concept
- Drop concepts: Highlight the concept and press the <Delete> key

**XIII. Create a delete list from scratch in AutoMap**

There are two ways to create a new delete list:

Within AutoMap

- Go to the **3. Delete List** index card and click inside the text field.
- Concepts must be arranged with one per line. To do this, simply press <Enter> after entering each concept.
- Avoid empty lines.
- When you are finished adding concepts, click the **Apply Delete List** button.

Outside of AutoMap

- Use any text editor to create a Delete List. Keep the following in mind:
  - The general structure of a Delete List requires one concept per line.
  - Avoid empty lines.
  - The Delete List is not case-sensitive.
Save the Delete List by going to File > Save As, and typing in a filename.

Example: C:\Documents and Settings\carley\Desktop\Lesson 2\SampleDeleteList.txt

Open your new Delete List in AutoMap: Go to the File > Open Form File, browse to your file, and click Open.

To delete concepts using this new delete list and examine the effects on the text and concept list, follow Step VIII.

To save a delete list you have created from scratch, follow Step X.

Example: C:\Documents and Settings\carley\Desktop\Lesson 2\AppliedDeleteList

XIV. Open a chosen delete list with Microsoft Excel

To open your delete list outside of AutoMap, navigate to the location in which the delete list was saved, and double-click on it. A window reading <Open With> pops up. Select Microsoft Excel and click OK. Provided Microsoft Excel is installed on your computer, the file will open.

Alternatively, you can copy and paste your delete list from the AutoMap GUI directly into Microsoft Excel. Press <Control> and <A> simultaneously while in the 3.Delete List panel (this selects all concepts in your delete list), then press <Control> and <C> simultaneously (this copies the delete list). Go to Microsoft Excel and press <Control> and <V> simultaneously (this pastes the delete list into Excel).

XV. Modify a delete list in Microsoft Excel

The same rules that applied for delete lists in AutoMap also apply for lists in Excel:

- Use any text editor to create a Delete List. Keep the following in mind:
  - The general structure of a Delete List requires one concept per line.
  - Avoid empty lines.
  - The Delete List is NOT case-sensitive.
  - Save the Delete List by going to File > Save As, and typing in a filename.

- Open your new Delete List in AutoMap: Go to the File > Open Form File, browse to your file, and click Open.

To un-apply a Delete List you have created from scratch and/or modified, follow Step IX.

To load in a modified delete list, follow Step VII.

To delete concepts using a modified delete list and examine the effects on the text and concept list, follow Step VIII.
Lesson 3

Processing data: using a Generalization Thesaurus

I. Open AutoMap

On Empire:

- Navigate to C:\Documents and Settings\Carley\Desktop\Tools.
- Double-click on the Run Automap icon. The Automap Graphical User Interface (GUI) pops up.

From anywhere else: (requires a web connection)

http://www.casos.cs.cmu.edu/projects/automap/software.html

II. Overview of the AutoMap Graphical User Interface (GUI)

Below is a screen capture of the AutoMap GUI:

The AutoMap GUI is divided into four primary quadrants, or panels: P1 (top left), P2 (bottom left), P3 (top right), and P4 (bottom right) as seen in the screen shot above.

A drop-down menu bar at the top of the window provides access to analysis tools and utilities. The Browse Menu allows you to quickly navigate between documents you have loaded into Automap.

The P2 panel can be edited, but the other panels (P1, P3, P4) cannot. Information displayed on each panel is always related to the information displayed in the other three panels. The Browse Menu relates to all four panels at the same time.
III. Load multiple text documents into AutoMap

In order to analyze multiple documents simultaneously, you will need to store the documents together in one folder.

Example: C:\Documents and Settings\carley\Desktop\TextInputFiles02-01-07.

- In the AutoMap menu bar, select File, then Open Multiple Files. A folder chooser window will pop up.
- Navigate to the folder containing the documents you wish to analyze, and single-click on it.

Example: C:\Documents and Settings\carley\Desktop\TextInputFiles02-01-07

- Click the Open button in the bottom right corner.

The first text will be displayed in panel P1 under the tab titled 1.Original Texts. You can use the Browse Menu to browse through the texts.

Below is a screen capture of the first text display:

IV. Carry out a simple content analysis

A simple content analysis determines the frequencies of all words in a text document.

Examine the Concept List:

The Concept List considers concepts in each text file individually. It is created automatically when a document or set of documents is loaded into Automap, and is displayed in panel P2 under the tab titled 1.Concept List.

The Concept List tells you several important facts about each text in your set:

- The number of concepts found in the text displayed in panel P1 under the tab titled 1.Original Text
- Related frequencies of those concepts
- Cumulated unique concepts and total concepts contained in the data set

Note: The number of unique concepts considers each concept only once. 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. To return the list to frequency, click on the second-column header Frequency.

Below is a screen capture showing part of the Concept List, sorted by frequency, for the first document in our loaded folder of texts (Text1-in-user's guide.txt):

Create and refresh the Union Concept List:

The Union Concept List considers concepts across all texts in a set (unlike the Concept List, which considers each text in the set individually). It is located in panel P2 under the tab titled 2.Union Concept List.

The Union Concept List tells you several important facts about your text set:

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

Before viewing your Union Concept List, you must refresh it. To refresh, go in the Menu Bar, choose File, then Refresh Union Concept List.

The Union Concept List can be refreshed after each step of pre-processing to visualize the impact of pre-processing operations on the union of concepts.

Below is a screen capture showing where to find the Refresh Union Concept List command:

After refreshing your Union Concept List, you can view it under the 2. Union Concept List tab in panel P2. Below is a screen capture showing where this tab is located in the GUI:

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

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

V. The basics of pre-processing

AutoMap follows a 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. If needed, you can un-apply each technique after applying it.

If you wish to apply multiple pre-processing techniques, carry out the process in the following order:

1. Named-Entity Recognition
   This utility does not impact the data. It can be used before any type of analysis is run. It can be used before or after Stemming.

2. Collocation/Bigram Identification
   This utility does not impact the data. It can be used before any type of analysis is run.

3. Stemming
   This can be used before any type of analysis is run. It can be used before or after Named-Entity Recognition.

4. Deletion
   This can be used before any type of analysis is run.

5. Thesauri
   - Generalization Thesaurus: This can be applied before Semantic Network Analysis is run. It can be applied before Meta-Matrix Thesaurus is applied.
   - Meta-Matrix Thesaurus: This has to be applied if a Meta-Matrix Analysis will be run.
   - Sub-Matrix Selection: This can only be performed if the Meta-Matrix Thesaurus was applied. It must be applied if a Sub-Matrix Analysis will be run.

This hierarchy is reflected in the numbering of the index card tabs in panels P1 and P2, to make the sequence of the pre-processing steps more intuitive.

Below is a screen capture showing the hierarchy of pre-processing techniques in panels P1 (top) and P2 (bottom):

VI. Open a Generalization Thesaurus
From the Menu bar select **File > Open Generalization Thesaurus**.

The thesaurus will be displayed in panel P2 under the 3.Preprocessing Settings tab, as a sub-tab titled 4.Generalization Thesaurus.

**VII. Apply a thesaurus and examine its effect on the texts and concept list**

*Note: If you wish to apply both a Delete List and a Generalization Thesaurus, be sure to use first the Delete List and then the Thesaurus.*

When applying a thesaurus, you have the following options:

**Thesaurus content only**

If this option is chosen, AutoMap:

1. Searches the texts for concepts specified in the thesaurus.
2. Translates matches into key concepts.
3. Maintains only key concepts in the pre-processed texts. The remaining input text is dropped and is not considered in further pre-processing or analysis. The original distances of the key concepts are not maintained, but punctuation marks (such as at the ends of sentences and paragraphs) are maintained and considered in analysis.
4. Results in all key concepts in the text appearing directly adjacent to each other.

If this option is not chosen, AutoMap:

1. Searches the texts for concepts specified in the thesaurus.
2. Translates matches into key concepts.
3. Keeps the rest of the text as is. This means that 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 are maintained (unless a concept consisting of more than one word was translated into a key concept).

If you do not choose **"thesaurus content only"**, the setting will not be applied. If you do choose **"thesaurus content only"**, you then have another option:

**Direct or rhetorical adjacency**

Direct adjacency means that original distances between concepts that represent the key concepts are neither visualized nor considered for analysis.

Rhetorical adjacency means that the original distances between key concepts are retained and incorporated into later analyses. The original distances are visually symbolized by placeholders (**"xxx"**).

To choose the direct adjacency option, click the Direct button in the Adjacency field in the 4.Generalization Thesaurus index card under 3.Pre-processing tab in panel P2. To choose the rhetorical adjacency option, click the Rhetorical button in the same field.

If you do not change the adjacency option, AutoMap uses direct adjacency as the default. Also, direct adjacency is automatically applied if the **"thesaurus content only"** option is not checked.

To select the thesaurus content only option, check the box for Thesaurus content only the 4.Generalization Thesaurus index card in panel P2. You can then choose either direct or
rhetorical adjacency by clicking in the circles with those labels.

To apply your chosen settings, click the Apply button on the 4.Generalization Thesaurus index card in panel P2.

To switch back to not using the “thesaurus content only” option, un-check the Thesaurus content only box in the 4.Generalization Thesaurus index card in panel P2, then apply the Generalization Thesaurus again.

VIII. Un-apply a Generalization Thesaurus

To un-apply a Generalization Thesaurus that was applied to the data, go to the 4.Generalization Thesaurus index card and click the Un-Apply button. This clears the 5.Texts After Generalization tab in panel P1.

IX. Modify the generalization thesaurus list

You can modify your generalization thesaurus by clicking inside the text field on the 4.Generalization Thesaurus index card in panel P2. Keep in mind the following points:

1. Every line contains Concept / Key Concept (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 columns.
5. The Thesaurus is not case sensitive.

To apply a modified thesaurus and examine its effects on the texts and concept list, follow Step VII.

To un-apply a modified thesaurus, follow Step VIII.

X. Create a generalization thesaurus list from scratch

You can create a Generalization Thesaurus from scratch in AutoMap.

Find the 4.Generalization Thesaurus index card under the 3.Preprocessing Settings tab in panel P2. You will see a text field, composed of two columns titled “concept” and “generalization”, that resembles a Microsoft Excel spreadsheet. To edit items inside this text field, simply double-click on them.

Automap can help with the building of a generalization thesaurus by loading the Union Concept List from the highest level of pre-processing applied into the Generalization Thesaurus field.

To load in the highest-level union concept list:

1. Follow Step 4 to create and refresh the Union Concept List
2. Click the Load Union Concept List button on the 4.Generalization Thesaurus index card.

Once loaded into AutoMap, the Concept List can be refined by applying Named-Entity Recognition and Deletion prior to application of the Generalization Thesaurus. (See Lessons 1 and 2).

To apply the thesaurus you have made from scratch and examine its effect on the texts and concept list, follow Step VII.
XI. Save the Generalization Thesaurus

To specify a location for your Generalization Thesaurus file, select File in the Menu Bar, then Output Storage Manager.

Below is a screen capture showing where to find this in the GUI:

A window will pop up listing all pre-processing output storage. This is the Output Storage Manager.

Find the Applied Generalization Thesaurus item in the list on the left-hand side of the Output Storage Manager. Click the Save file as… button to the right of the default file pathway. In the file chooser window that pops up, browse to the pathway you want and type in a filename.

Example: C:\Documents and Settings\carley\Desktop\Lesson 3\AppliedGeneralizationThesaurus

Click the Save Settings button at the bottom right corner of the Output Storage Manager. Close the Output Storage Manager window by clicking on the red X at the top right corner.

In the Automap Menu Bar, choose File, then Save applied Generalization Thesaurus.

XII. Open the Generalization Thesaurus in Microsoft Excel

Outside of Automap, navigate to the Generalization Thesaurus you have saved and double-click on it. A window titled “Open With” will pop up. Double-click on Microsoft Excel. The file will open in Excel (provided you have Excel installed on your computer).

Alternatively, you can copy and paste your Generalization Thesaurus into Excel directly from Automap. Single-click inside the text field on the 4. Generalization Thesaurus index card in panel P2. Press <Control> and <A&t; simultaneously to select all the text, followed by and simultaneously to copy the selected text. Then go into a blank Excel spreadsheet and press and simultaneously to paste the generalization thesaurus into the spreadsheet.

XII. Modify the Generalization Thesaurus in Microsoft Excel

You can modify your generalization thesaurus in Microsoft Excel just as for a standard spreadsheet. Keep in mind the following points:

1. Every line contains Concept / Key Concept (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 columns.
5. The Thesaurus is not case sensitive.

To load in a thesaurus you have modified in Microsoft Excel, follow Step VI.

To apply this modified thesaurus and examine its effects on the texts and concept list, follow Step VII.

To un-apply this modified thesaurus, follow Step VIII.
Lesson 4

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 were 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.*