Automap User's Guide 2008

Kathleen M. Carley, Dave Columbus,
Matthew J. DeReno and Jana Diesner

July 2008
CMU-ISR-08-123

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Center for the Computational Analysis of Social and Organizational Systems
CASOS Technical Report

This report/document supersedes CMU-ISRI-07-114
"Automap User's Guide 2007", July 2007

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

Key Words: Semantic Network Analysis, Dynamic Network Analysis, Mental Models, Social Networks, AutoMap

Abstract

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

Table of Contents

1 Introduction to AutoMap 1
1.1 An Overview 1
1.2 Network Text Analysis (NTA) 2
1.3 Semantic Network Analysis 2
1.4 Social Network Analysis (SNA) 2
1.5 Where to find out more on SNA? 2
2 Dynamic Network Analysis 3
2.1 Where to learn to more: 4
3 Formats 5
4 Automap web page 5
4.1 Included for download are: 5
5 Main Interface 6
6 GUI Overview 7
7 AutoMap Menus - The MenuBar 8
7.1 Browsing through texts 8
8 AutoMap Menus 8
9 File Menu 9
9.1 Example 9
10 Delete List submenu 13
11 General Thesaurus submenu 14
12 Meta-Network Thesaurus submenu 15
13 Sub-Matrix Thesaurus submenu 15
14 Output Storage Manager 16
14.1 About Map Analysis 18
14.2 Analyzing Text(s) 19
15 Single Map Analysis 19
16 Single Map Analysis Example 20
16.1 Load a text file into AutoMap 20
16.2 Load the extensive Delete List: 20
16.3 Apply Delete List with Direct Adjacency (default) 20
16.4 The Map 21
16.5 The Analysis 22
17 Multiple Map Analysis 22
18 Example 23
18.1 Cleaning up the text 23
18.2 Running the Map Analysis 23
19 Additional Tools 23

19.1 Text data Preparation Tool : 24
19.2 Network Post Processing Tools : 24
19.3 Other Tools : 24
20 Text Data Preparation Tools 25
21 Split Input Texts 25
22 Text Data Preparation Tools 25
23 Merge DyNetML 26
24 Convert Network Data 26
25 Other Tools 27
26 CASOS Email Parser (CEMAP) 28
26.1 The Top textfields 29
26.2 The Directory Buttons 29
26.3 Checkboxes at the bottom 30
26.4 The Start Button 30
26.5 Extraction of network type 1 data: 30
26.6 To use the feature: 31
26.7 The different types of imports are 31
26.8 Type Selection 31
26.9 Examples of POP3 and IMAP setups 32
27 Compare Maps 33
28 Visualize Semantic Network 35
29 Text Display Panel 35
30 Text Processing Panel 36
31 Text Processing Panel - Concept List 36
31.1 The Concept List 36
32 Text Processing Panel - Union Concept List 37
32.1 Create and refresh Union Concept List 37
32.2 Save Union Concept List 38
32.3 Example for Union Concept List 38
32.4 Create a Delete List 39
33 Text Processing Panel 40
34 Text Processing Panel - Utilities 40
34.1 Text before Removal of Symbols and Number 41
34.2 Text after Removal of Symbols and Number 41
34.3 Text after Removal of Symbols 41
35 Text Processing Panel - Bi-Gram Detection 42
35.1 N-gram Identification: Bigrams 42
35.2 Text Example 42
35.3 Results after Bi-Gram list created 43
36 Text Processing Panel - Feature Selection 46
36.2 Opening the file in Excel 46

37 Named Entity Recognition 47
37.1 Text Example 47
37.2 NamedEntities.csv 48
37.3 as base for Generalization Thesaurus 48
38 Text Processing Panel - Parts of Speech Tagging 50
38.1 Text Example 50
39 51
40 Text Processing Panel - Anaphora Resolution 52
41 Text Processing Panel - Extraction of Numerals 53
41.1 Text Example 53
41.2 Extraction of Numerals for Delete List 53
41.3 Extraction of Numerals for Meta Network Thesaurus 54
42 Text Processing Panel - Extraction of Time Data 54
42.1 Text Example for Extraction of Date 54
43 Text Processing Panel - Union Concept List in Context 55
44 Text Processing Panel - Create Lists 55
45 Text Processing Panel - Stemming 56
46 Stemming Example 57
46.1 Original Text 57
46.2 Text After Porter Stemming 57
46.3 Text After KSTEM 57
47 Text Processing Panel - Deletion List 59
48 Text Processing Panel - Deletion List 59
48.1 Four different Delete List can be opened form the menu. 59
48.2 Format of the Delete List 60
48.3 Editing the Delete List 60
48.4 Save an applied Delete List 60
49 Deletion List Example 60
50 Adjacency Options 61
50.1 Direct Adjacency: 61
50.2 Rhetorical Adjacency: 61
51 Text Processing Panel - Generalization Thesaurus 62
52 Open a Generalization Thesaurus 62
53 Editing a Thesaurus 62
54 Creating a General Thesaurus 63
54.1 Create Named-Entity List 63
54.2 Build a Generalization Thesaurus 64
54.3 Add further words 64
55 Applying a Generalization Thesaurus 64
55.1 Apply a Generalization Thesaurus 65
55.2 Un-Apply a Generalization Thesaurus 65

56 Save a Generalization Thesaurus 65
57 Generalization Thesaurus Example 65
57.1 Original Text 65
57.2 Thesaurus Content Only - Off 65
57.3 Thesaurus Content Only - On – Direct Adjacency 66
57.4 Thesaurus Content Only - On – Rhetorical Adjacency 66
58 Create Positive Thesaurus 66
59 Load Union Concept 66
59.1 Creating and Using "Union of Concepts" 66
60 Adjacency Options 67
61 Thesaurus Content Only 68
61.1 Thesaurus Content only - chosen 68
61.2 Thesaurus Content only - NOT chosen 68
62 Meta Network Thesaurus 68
62.1 Related type of Analysis: Meta Network Analysis. 69
63 Open a Meta Network Thesaurus 70
64 Edit Meta-Network Thesaurus 70
64.1 User-defined Category 70
64.2 Multiple User-Defined Categories 70
65 Build a Meta-Network outside of AutoMap 71
65.1 Rules for Building a Thesaurus 71
66 Save an applied Meta Network Thesaurus 71
66.1 Save the Meta Network 71
66.2 Save the Applied text 71
67 Example of a Meta Network Thesaurus 71
68 Apply Meta-Network Thesaurus 73
68.1 Procedure: 73
69 Adjacency Options 73
70 Thesaurus Content Only 74
70.1 Thesaurus Content only - chosen 74
70.2 Thesaurus Content only - NOT chosen 74
71 Sub-Matrix Selection 75
71.2 Select Sub-Matrix 75
71.3 Adjacency 76
71.4 Apply / Un-Apply buttons 76
72 Applying a Sub-Matrix Selection 76
72.1 Apply a Sub-Matrix 76
72.2 Un-Apply a Sub-Matrix 77
72.3 Direct Adjacency 77
72.4 Rhetorical Adjacency 77
73 Example of a Sub-Matrix 77

73.1 Select: 77
73.2 Select: 78
73.3 Select: 78
73.4 Select: 79
74 Saving a Sub-Matrix 80
75 Selecting Sub-Matrix Categories 80
75.1 If you do not have a sub matrix selection file available 80
75.2 If you do not have a sub matrix selection file available and want to select the full Meta Network (means all cells in the Meta Network) 81
75.3 If you have a sub matrix selection file available 82
75.4 If you want to write your own sub matrix selection file and load it into AutoMap: 82
76 Adjacency Options 82
77 Text Processing Panel - Analysis Setting 83
77.1 Analysis Settings 83
78 Text Processing Panel - Analysis 85
79 Text Processing Panel - Analysis Setting 85
79.1 Coding Ties 85
79.2 Windowing 86
80 Output Options 87
81 Term Distribution List 87
81.1 Points To Consider 87
81.2 Types and content of Term Distribution Lists 88
81.3 Concepts analyzed 88
81.4 Concepts in Statements and Isolates 88
81.5 Statements 88
82 Term Distribution Matrices 88
82.1 Concepts in statements by concepts in statements: 89
82.2 Term(s) by text(s): 89
83 Term Distribution List Example 89
84 Additional Output Options 91
85 Additional Output Formats 91
85.2 UciNet (DL) 91
85.3 DyNetML 92
85.4 After Map Analysis: 92
85.5 Run Sub-Matrix Text Analysis. 92
85.6 Examples for additional Output Formats 92
86 Parts of Speech Tagging Options 94
87 Sub Matrix Analysis Options 95
88 Concepts not Forming Edges 95
89 Semantic Networks of Current Text Panel 96
90 Results Panel 97

91 Results Panel — Action Tracer 97
92 Results Panel ⇒ Statistics 97
93 98
94 Network Analytic Measures 99
95 Open Input 101
95.1 Open a single text 101
95.2 Open a Multiple Files 101
96 GUI Overview 101
97 AutoMap Menus - The MenuBar 103
97.1 Browsing through texts 103
98 Concept List 104
99 Union Concept List 105
99.1 Create and refresh Union Concept List 105
99.2 Example for Union Concept List 105
99.3 Creating a Union Concept List 107
99.4 Save Union Concept List 107
100 Named Entity Recognition 107
100.1 Text Example 108
100.2 NamedEntities.csv 109
100.3 Used as a base for Generalization Thesaurus 109
101 Text Processing Panel - Bi-Gram Detection 111
101.1 N-gram Identification: Bigrams 111
101.2 Text Example 111
101.3 Results after Bi-Gram list created 112
102 Clean-up Text 115
102.1 Text before Removal of Symbols and Number 115
102.2 Text after Removal of Symbols and Number 115
102.3 Text after Removal of Symbols 116
102.4 N-gram Identification: Bigrams 116
103 Text Processing Panel - Feature Selection 116
103.2 Opening the file in Excel 117
104 Text Processing Panel - Parts of Speech Tagging 117
104.1 Text Example 118
105 Text Processing Panel - Anaphora Resolution 119
106 Text Processing Panel - Extraction of Numerals 121
106.1 Text Example 121
106.2 Extraction of Numerals for Delete List 122
106.3 Extraction of Numerals for Meta Network Thesaurus 122
107 Text Processing Panel - Extraction of Time Data 122
107.1 Text Example for Extraction of Date 122
108 Text Processing Panel - Union Concept List in Context 124

109 Stemming 124
110 Stemming Example 125
110.1 Original Text 125
110.2 Text After Porter Stemming 125
110.3 Text After KSTEM 126
111 Deletion List 126
111.1 The Two main Delete Lists can be opened form the menu. 126
111.2 Format of the Delete List 126
111.3 Editing the Delete List 127
111.4 Apply a Delete List 127
111.5 Rhetorical adjacency 128
111.6 Save an applied Delete List 128
112 Deletion List Example 128
112.1 Direct Adjacency: 129
112.2 Rhetorical Adjacency: 129
112.3 Direct Adjacency: 129
112.4 Rhetorical Adjacency: 129
113 Generalization Thesaurus 130
113.1 Creating a General Thesaurus 130
113.2 Create Named-Entity List 131
113.3 Build a Generalization Thesaurus 131
113.4 Add further words 132
113.5 Applying a Generalization Thesaurus 132
113.6 Un-Apply a Generalization Thesaurus 132
113.7 Editing a Thesaurus 132
113.8 Save a Generalization Thesaurus 133
114 Generalization Thesaurus Example 133
114.1 Original Text 133
114.2 Thesaurus Content Only - Off 133
114.3 Thesaurus Content Only - On – Direct Adjacency 133
114.4 Thesaurus Content Only - On – Rhetorical Adjacency 134
115 One Mode Network 134
116 Semantic Network Analysis 134
117 Semantic Network 135
117.1 Example for Semantic Network Output 135
117.2 Example for Semantic Network Analysis 135
118 Multi Mode Network 137
119 Analysis Setting 138
119.1 Coding Ties 138
119.2 Windowing 139
120 Meta Network Thesaurus 139

120.1 Open a Meta Network Thesaurus 140
120.2 Edit a Meta Network Thesaurus 140
120.3 Build a Meta Network Thesaurus 140
120.4 Apply a Meta Network Thesaurus 141
120.5 Thesaurus content only 141
120.6 Direct Adjacency 143
120.7 Rhetorical Adjacency 143
120.8 Un-Apply a Meta Network Thesaurus 143
120.9 Save an applied Meta Network Thesaurus 144
120.10 Save text(s) after application of Meta Network Thesaurus 144
121 Example of Meta Network Thesaurus 144
122 Sub-Matrix Selection 148
122.2 Select Sub-Matrix 149
122.3 Adjacency 149
123 Selecting Sub-Matrix Categories 149
123.2 If you do not have a sub matrix selection file available and want to select the full Meta Network (means all cells in the Meta Network) 150
123.3 If you have a sub matrix selection file available 151
123.4 If you want to write your own sub matrix selection file and load it into AutoMap: 151
123.5 Create Sub-Matrix Selection 152
123.6 Applying a Sub-Matrix Selection 152
123.7 Un-Apply a Sub-Matrix 152
123.8 Direct Adjacency 153
123.9 Rhetorical Adjacency 153
123.10 Save Sub-Matrix Selection 153
123.11 Save text(s) after Sub-Matrix Selection 153
124 Example of a Sub-Matrix 153
124.1 Select: 153
124.2 Select: 154
124.3 Select: 154
124.4 Select: 154
125 Saving a Sub-Matrix 155
126 Meta Network Text Analysis 155
126.1 Steps to take before running the Meta Network analysis: 156
126.2 Analyzing Text(s) 156
126.3 Results 156
126.4 Example for Meta Network Text Analysis 156
127 Sub-Matrix Text Analysis 160
127.1 Steps to take before running the Sub-Matrix analysis: 160
127.2 Exclude Entities 161
127.3 The use of ontologies 161

135.13 Run the Multiple Map Analysis: 185
136 ORA 186
136.1 Loading files into ORA 186
136.2 Visualize your meta-matrices 186
136.3 Run a Semantic Network Report 188
136.4 Semantic Network Report 188
136.5 Symmetric difference 189
137 189
138 Lesson 2 190
138.1 Using the Delete List 190
138.2 Basics of delete lists 190
138.3 Open a Delete List 190
138.4 Multiple Delete Lists 193
138.5 Un-apply a Delete List 193
138.6 Save an applied Delete List 193
138.7 Save text(s) after Delete List 193
138.8 Modify a Delete List 193
138.9 Create a delete list from scratch in AutoMap 194
138.10 Modify a delete list in Microsoft Excel 194
139 Lesson 3 194
139.1 Using a Generalization Thesaurus 194
139.2 Load text document(s) into AutoMap 194
139.3 Pre-Processing 194
139.4 Open a Generalization Thesaurus 194
139.5 Apply a Generalization Thesaurus 195
139.6 Apply a Generalization Thesaurus 195
139.7 Un-apply a Generalization Thesaurus 195
139.8 Modify the generalization thesaurus list 195
139.9 Create a generalization thesaurus list from scratch 196
139.10 Save the Generalization Thesaurus 197
139.11 Open the Generalization Thesaurus in Microsoft Excel 197
139.12 Modify the Generalization Thesaurus in Microsoft Excel 197
140 Outputs 198
140.1 Split Input Text Files 198
140.2 Using Compare Maps 198
140.3 Merge DyNetML Files 199
140.4 Convert Network Data Formats 199
140.5 Edit Network Data 199
140.6 Visualize Semantic Networks 199
141 AutoMap Batch 199
141.1 Things to Remember 200

141.2 Running AutoMap Batch 200
142 The AutoMap_script.XML.xml file. 200
142.1 Input 200
142.2 Utilities 200
142.3 PreProcessing 202
143 Basic Terms 207

1 Introduction to AutoMap

1.1 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 Meta-Network 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

1.2 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.

1.3 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

1.4 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.

1.5 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.

2 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.

2.1 Where to learn to more:

- Kathleen M. Carley, 2003, " Dynamic Network Analysis" in Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, Ronald Breiger, Kathleen Carley, and Philippa Pattison, (Eds.) Committee on Human Factors, National Research Council, National Research Council. Pp. 133-145, Washington, DC.
- Kathleen M. Carley, 2002, " Smart Agents and Organizations of the Future" The Handbook of New Media. Edited by Leah Lievrouw and Sonia Livingstone, Ch. 12, pp. 206-220, Thousand Oaks, CA, Sage.
- Kathleen M. Carley, Jana Diesner, Jeffrey Reminga, Maksim Tsvetovat, 2005-forthcoming, Toward an Interoperable Dynamic Network Analysis Toolkit, DSS Special Issue on Cyberinfrastructure for Homeland Security: Advances in Information Sharing, Data Mining, and Collaboration Systems.

3 Formats

There are many types of text formats available. Not all of them work correctly in AutoMap.

- ASCII (American Standard Code for Information Interchange) is the lowest common denominator. There are actually two ASCII codes. The original 128 character, 7-bit code and the expanded 256 character, 8-bit code.
- CSV (Comma Separated Value): a file type that stores tabular data. The format dates back to the early days of business computing. For this reason, CSV files are common on all computer platforms.
- EBCDIC (Extended Binary Coded Decimal Interchange Code)
- HTML (Hypertext Markup Language) : The predominant markup language used for web pages.
- ISO/IEC 8859 : standard for 8-bit character encodings for use by computers.
- RTF (Rich Text Format) : a proprietary document file format developed by DEC in 1987 for cross-platform document interchange. Most word processors are able to read and write RTF documents.
- UTF-8 (Uniform Transformation Format) : It is able to represent any character in the Unicode standard, yet the initial encoding of byte codes and character assignments for UTF-8 is backwards compatible with ASCII. For these reasons, it is steadily becoming the preferred encoding for e-mail, web pages, and other places where characters are stored or streamed.
- XML (Extensible Markup Language) : A general purpose makeup language that allows users to define their own tags.

4 Automap web page

Casos maintains a web site for Automap at http://www.casos.cs.smu.edu/projects/automap.php.

4.1 Included for download are:

Delete Lists : Extensive list, Punctuation delete list, and small list

Generalization Thesauri : Country generalization thesaurus, English abbreviations generalization thesaurus, and English contraction generalization thesaurus.

Meta Network Thesauri : Countries Meta Network thesaurus.

For Map Post Processor : Attribute interpretation thesaurus.

5 Main Interface

The Main Interface consists of two parts: The Menu and the Panels.

- An Overview of the AutoMap GUI
- The File Menu
- The Run Analysis Menu
- The Other Tools Menu
- The Menu Bar
- 1 : Text Display Panel
- 2 : Text Processing Panel
- 3 : Semantic Network Panel
- 4 : Results Panel

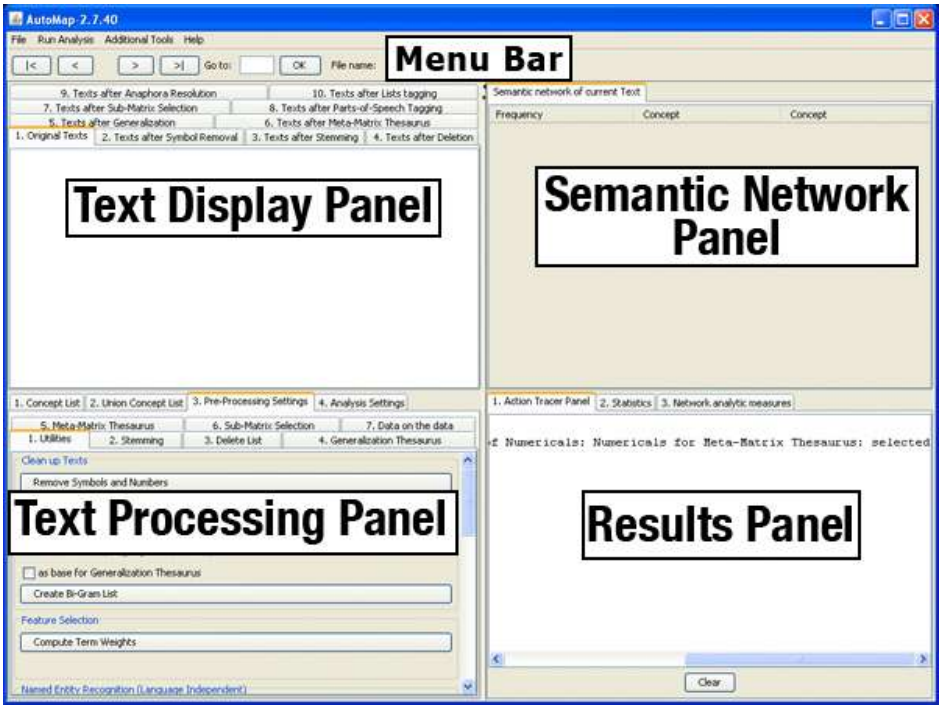
6 GUI Overview

AutoMap's graphical user interface (GUI) is divided into four primary panels. In the images below you can see the references: *1: Text Display Panel* (top left), *2: Text Processing Panel* (bottom left), *3: Semantic Network Panel* (top right) and *4: Results Panel* (bottom right).

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

The panels contain tabs which provide an 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:



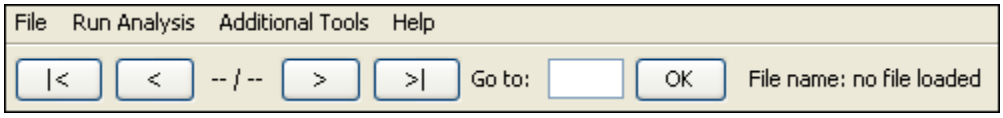
6.1.1 NOTE : The *Text Processing Panel* is the only panel that *can* be edited. The other panels (Text Display Panel, Semantic Network Panel, & Result Panel) display information regarding the currently active file. These panels always relates to each other. The Text Browse Menu relates to all panels 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 Tab in *Results Panel* will log preprocessing utilities applied to your text. This is a handy way to keep track of changes and actions relating to your text.

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.

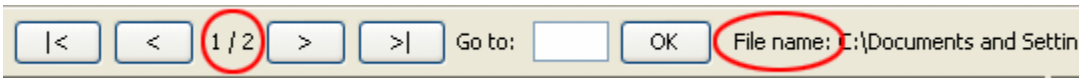
7 AutoMap Menus - The MenuBar



[| <] Go to first text file.
[<] Go to previous text file.
--/-- denotes # of file currently active / total # of files.
[>] Go to next text file
[> |] Go to last text file
Go to: Input the number of the text file you want to bring up.
[OK]
File name: xxx : Display the currently active text file. When AutoMap is first started this displays no file loaded

7.1 Browsing through texts

This function enables you to quickly jump from text example to another. All panels are synchronized in the Browse Menu.
Using the [>] and [<] buttons allow you move back and forth between the currently opened text files. The [|<] and [>|] will send you to the first or last file opened. The numbers in the middle 1 / 2 represent which number file you are viewing / total number of files open.



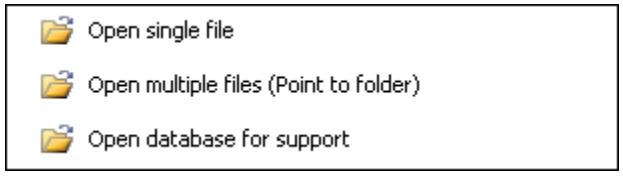
To go to a specific text enter the text number in text field right next to the Go to command and select [OK].
The name of the currently selected text is displayed to the right of File name:.

8 AutoMap Menus

AutoMap contains four menus:

1. The File menu is used for directing AutoMap where to load/save files.
2. The Run Analysis menu contains the main routines for analyzing texts.
3. The Additional Tools menu contains tools with functions that are of a more specialty orientated function.
4. The Help menu contains this file and the About Automap section

9 File Menu

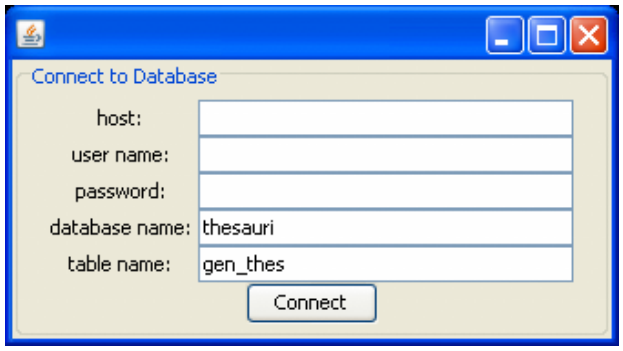


Open Single File : Opens only one file. After selecting this option navigate to the directory where your files are, highlight it, then select [OK].

Open Multiple Files (Point to folder) : This options allows you to open two or more files. The folder you navigate to will appear empty, but it isn't. Before you execute this option make sure all the files (and only these files) are in that directory. If there are any other txt files in the directory they too will be opened.

Open database for support :

This function uses a MySQL database running CASOS schema to add attributes from the database to the output dynetml file. Pressing the button will bring up a dialog box.

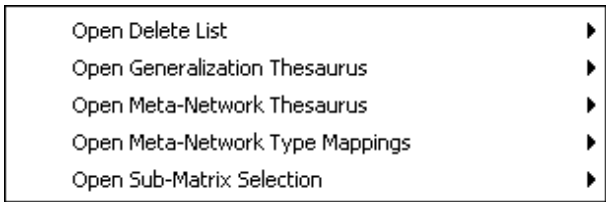


9.1 Example

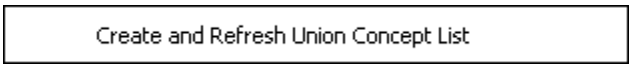
- database host: roboron.casos.cs.cmu.edu
- user name:ID used to access the database
- password: associated with the above user name
- database name: the particular database to connect to
- table name: an optional function

The resulting output dynetml file will contain node attributes from the database in case output dynetml file contains a node with same nodeID (node name) as the database.

Selecting [Connect] will connect the user to the specified database. The attributes from the database will then be automatically added when the dynetml output is created.

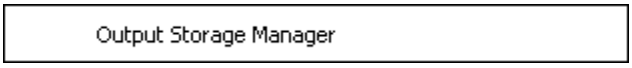


- Delete List : A sub-menu for controlling the opening of the Delete List options.
- Open Generalization Thesaurus : a sub-menu for controlling opening a General Thesaurus options.
- Open Meta Network Thesaurus : A sub-menu for controlling opening Meta Network Thesaurus options.
- Open Meta Network Type Mappings:
- Open Sub-Matrix Thesaurus : A sub-menu for controlling opening Sub-Matrix Thesaurus options.

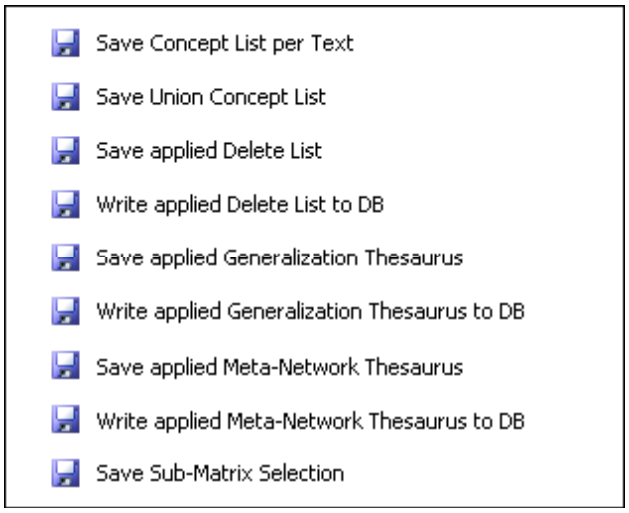


Create and Refresh Union Concept List : Creates a Union Concept List which can be viewed in *Panel 2 => Tab 2. Union Concept List.*

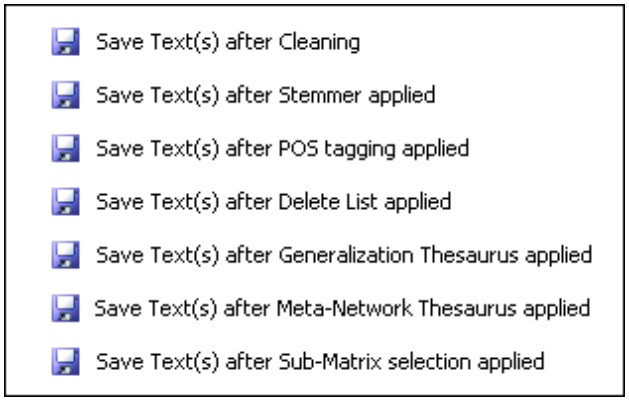
9.1.1 NOTE : Used only when multiple files are loaded.



The Output Storage Manager displays the pathways and names that AutoMap has mapped for each individual output file. Some have editable pathways and others have editable filenames.

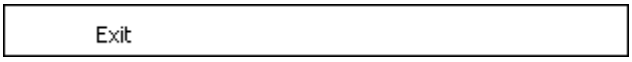


Save text or Write text to a database : After running various functions on a text file you can save the altered lists and thesauri for later use.



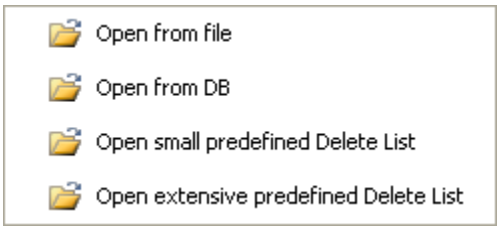
The Save Text(s) after … saves your text file after you have applied one of the processing functions. Each of the Save Text(s) after … corresponds to the tabs in Panel 1.

Clear : Clears ALL the current text files from memory.



Exit : Close the Program.

10 Delete List submenu



Open from file : Navigate to the directory with delete lists. These delete lists are lists in a text file.

Open from DB :

Open small predefined Delete List : The small predefined Delete List is compiled of the most common words that occur in English:

a, an, and, some, many, this, that, these, those, the, all, one, every.

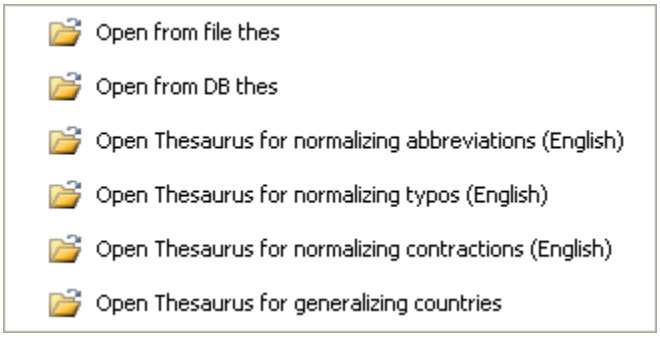
Open extensive predefined Delete List : An Extensive predefined Delete List contains a broader list of common words which includes the small defined delete list.

a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself.

10.1.1 NOTE : 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.

11 General Thesaurus submenu



Open from file thes :

Open from DB thes :

Open Thesaurus for normalizing abbreviations (English) : Loads the Thesaurus for common abbreviations, for example:

BBC, British_Broadcasting_Corporation, CBS, Columbia_Broadcasting_System, and CEO, Chief_Executive_Officer

Open Thesaurus for normalizing typos (English) : Loads the Thesaurus containing commonly misspelled word, for example:

[absense, absence] | [centruy, century] | [manuever, maneuver]

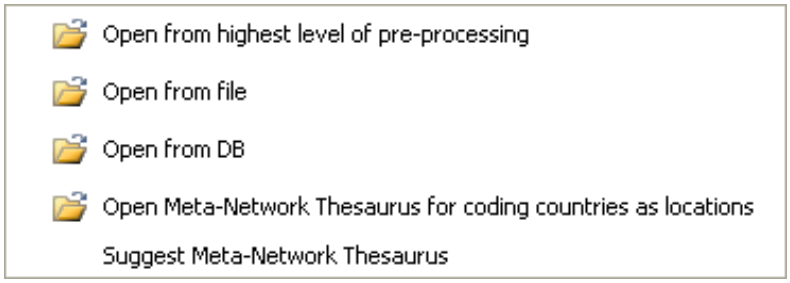
Open Thesaurus for normalizing contractions (English) : Loads the Thesaurus with common contractions, for example:

[we're, we are] | [I'm, I am] | [she'd, she had]

Open Thesaurus for generalizing countries :

[USA, United_States_of_America] | [El Salvador, El_Salvador] | [Sierra Leone, Sierra_Leone]

12 Meta-Network Thesaurus submenu



Open from highest level of pre-processing :

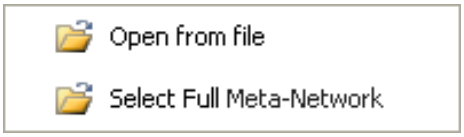
Open from file :

Open from DB :

Open Meta Network Thesaurus for coding countries as locations :

A Meta Network Thesaurus associates concepts with the following Meta Network categories: Agent, Knowledge, Resource, Task/Event, Organization, Location, Action, Role, Attribute, Any user-defined category (as many as the user defines)

13 Sub-Matrix Thesaurus submenu

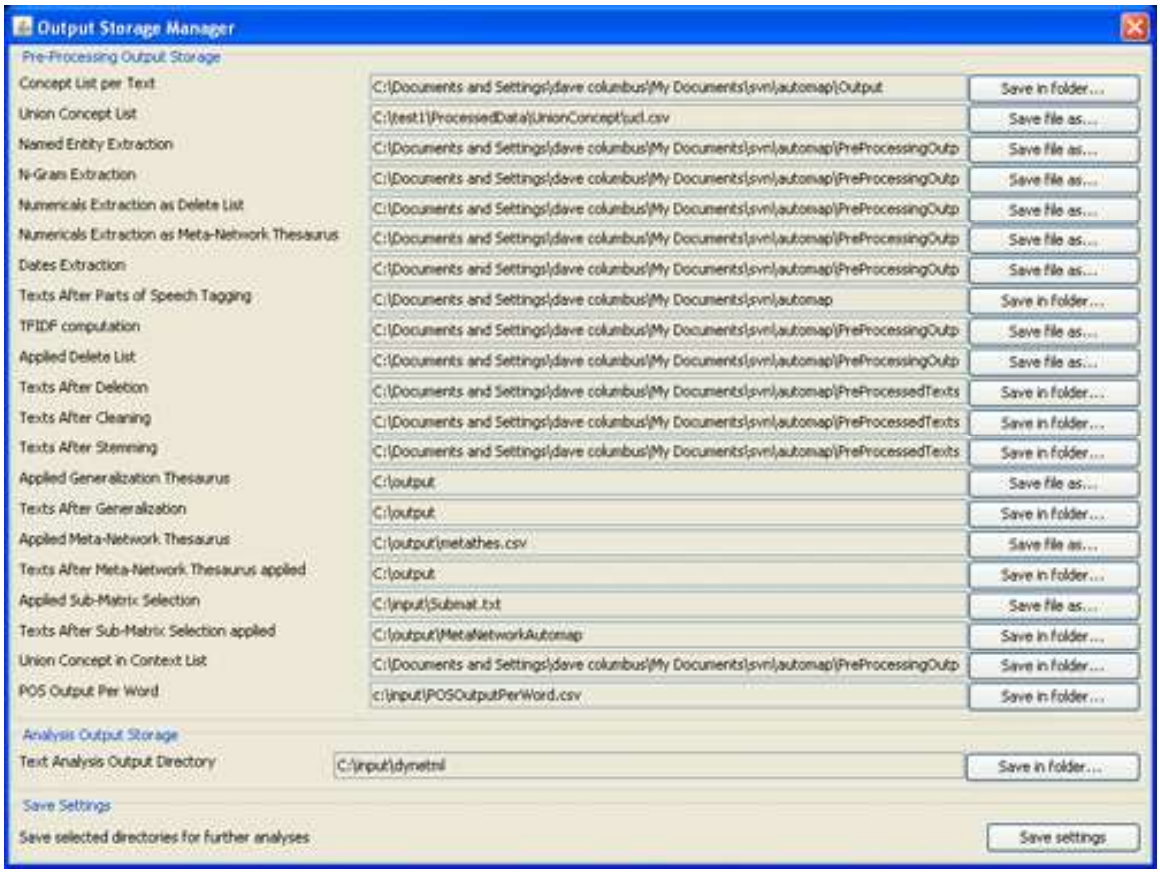


Open from file :

Select full Meta Network :

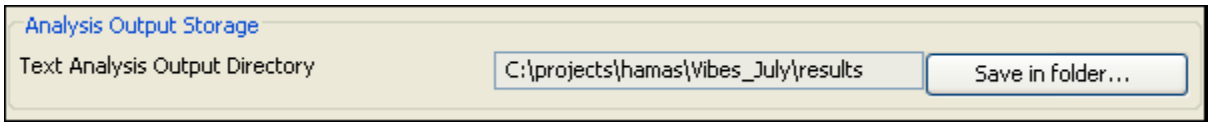
14 Output Storage Manager

This options allows you to specify either what filename to save an output as, *Save file as…* or the destination of a file *Save in folder*. Each item has a default but you may change whatever parameters you chose.

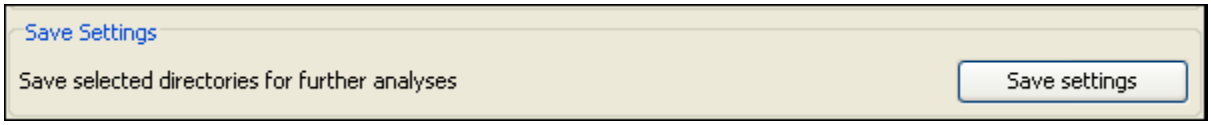


Be aware of the different buttons *Save in folder…* and *Save file as…*. For the *Save file as…* a specific filename needs to be used.

Here is a closer look at the various options.



The directory for Analysis Output Storage has it's own specific folder.



Finally, to save the choices, select [Save settings].

14.2 Analyzing Text(s)

To analyze a single text, use the Single Meta Network Text Analysis. To analyze a set of texts use Multiple Meta Network Text Analysis.

The Results will be displayed in *Panel 3* and in *Panel 4*. If you had requested additional outputs, those will be generated and stored under the directories specified under the Section Additional Outputs.

If multiple texts are being analyzed they can be browsed through with the results seen in the panels of the main interface.

Results of multiple analyses are automatically saved in a folder named output under the directory where AutoMap 2.0 is installed. This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as stat_output.xls file a that contains the stats of all texts.

The Stat Output folder is overwritten with every new analysis you run. If this file is needed for future reference it needs to either be moved to a new directory or renamed.

15 Single Map Analysis

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

- Pre-process the texts if you wish to use.
- Specify the analysis settings or else the default setting default settings will be applied.
- Specify additional output options cause if you do not the default outputs will be generated

The Results will be displayed on the *Semantic Network Panel => Semantic network of current Text* and on the *Results Panel => Tab 3. Statistics*. 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 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.

16 Single Map Analysis Example

16.1 Load a text file into AutoMap

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

16.2 Load the extensive Delete List:

Before running an Map Analysis it's a good idea to run the Delete List to remove the unwanted concepts.

a | and | as | at | but | for | he | her | 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

16.2.1 NOTE : The Delete List can be added to/subtracted from is needed.

16.3 Apply Delete List with Direct Adjacency (default)

Synopsis: Tok'ra are planning attack which kill all System Lord Gou'alds in one fell swoop by infiltrating summit poisoning all simultaneously. present plan military planet Earth. With help SGC United States Military Daniel is able infiltrate summit with Jacob Carter/Selmak backup. O'Neill, rest SG-1, stay on Revenna. plan is going fine until new emmisarry presents some new information Revenna is attacked. All plans are now considered worst things possible.

16.4 The Map

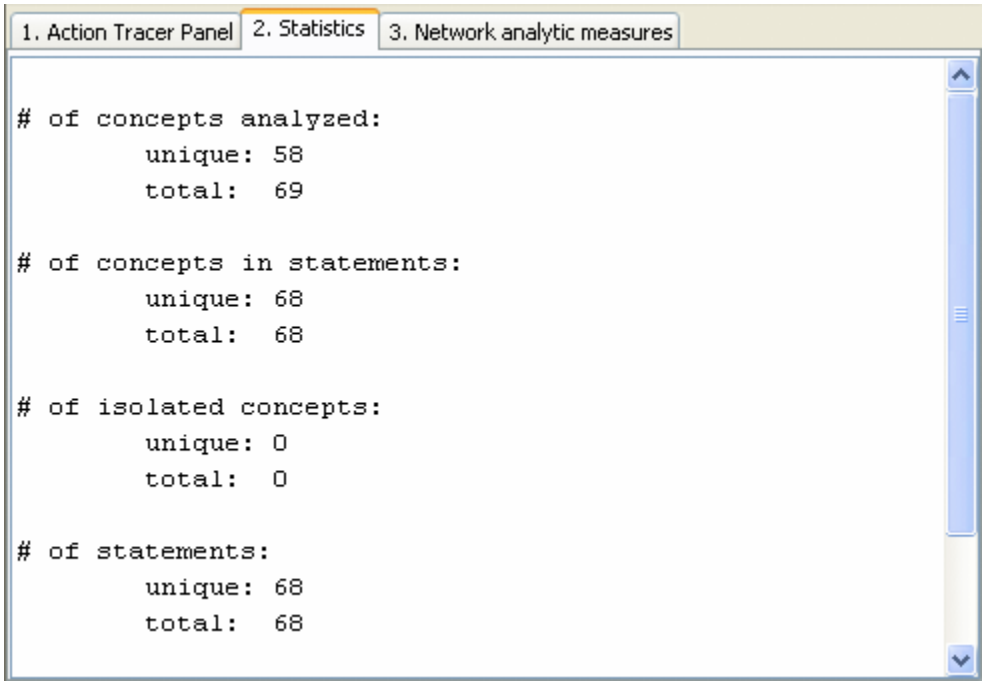
In the *Semantic Network Panel* AutoMap will display the mapped concepts with their frequency of each pair.

Semantic network of current Text

Frequency	Concept	Concept
1	able	infiltrate
1	all	plans
1	all	simultanously
1	all	system
1	are	now
1	are	planning
1	attack	which
1	attacked	all
1	backup	o'neill
1	by	inflitrating
1	carterselmak	backup
1	considered	worst
1	daniel	is
1	earth	with
1	emmisarry	presents
1	fell	swoop
1	fine	until
1	going	fine

16.5 The Analysis

In the *Results Panel => Tab 3. Statistics* AutoMap will display various statistics for the Map Analysis.



17 Multiple Map Analysis

Multiple Map Analysis works like the Single Map Analysis on multiple loaded files.

From the main menu select *File => Open Multiple Files* then navigate to the folder where the text files reside. Then select [OK].

17.1.1 REMEMBER : The folder will only APPEAR empty in this function but it isn't.

Once the files are loaded select from the main menu *Run Analysis => Multiple Map Analysis*. As in the Single Map Analysis it is suggested to run the Delete List before running the Map Analysis.

Various Output Options can be set for these files from *Text Processing Panel => Tab 4. Analysis Settings => Sub-Tab 2. Output Options.*

18 Example

synopsis.txt

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

timeframe.txt

On March 22, 2002 the Ren'al of the Tok'ra arrived @ the SGC to discuss their plan. 5 days later, on March 27, 2002, SG-1 and SG-17 arrive @ Revenna. The summit meeting with the 6 main System Lords* will be happening in 10 days. Daniel is faced with the question, "Will this prevent the System Lords from overthrowing the galaxy?"

*Gou'alds

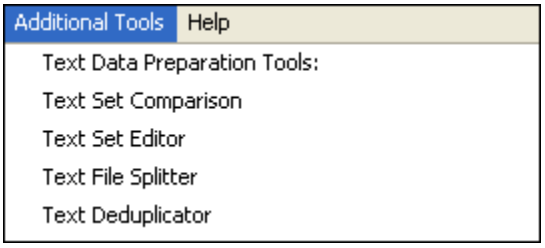
18.1 Cleaning up the text

The Extended Delete List was run on the files to remove unwanted concepts. Next a custom Generalization Thesaurus coordinated concepts.

18.2 Running the Map Analysis

From the main menu select *Run Analysis => Multiple Map Analysis*. This will display in the *Semantic Network Panel* the Semantic network of the current Text and in the *Results Panel => Tab 2. Statistics* the numbers for concepts analyzed, in statements and other computed data. By using the [<] and the [>] buttons in the menu bar the loaded texts can be toggled through and *Semantic Network Panel and Results Panel* will change accordingly.

19 Additional Tools



20 Text Data Preparation Tools

Text Set Comparison :
Text Set Editor:
Text File Splitter :
Text Deduplicator :

21 Split Input Texts

The File Splitters splits large text files into smaller ones of minimum equal size. It's used in order to speed up AutoMap coding. The number of words (NW) is the size that each of the number of text (NT) text file should contain after splitting.

From the main menu select *Tools => Open Text File Splitter*.

Input : 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 : NT 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 NT, where NT indicates the largest number of texts that an original text had been split up into.

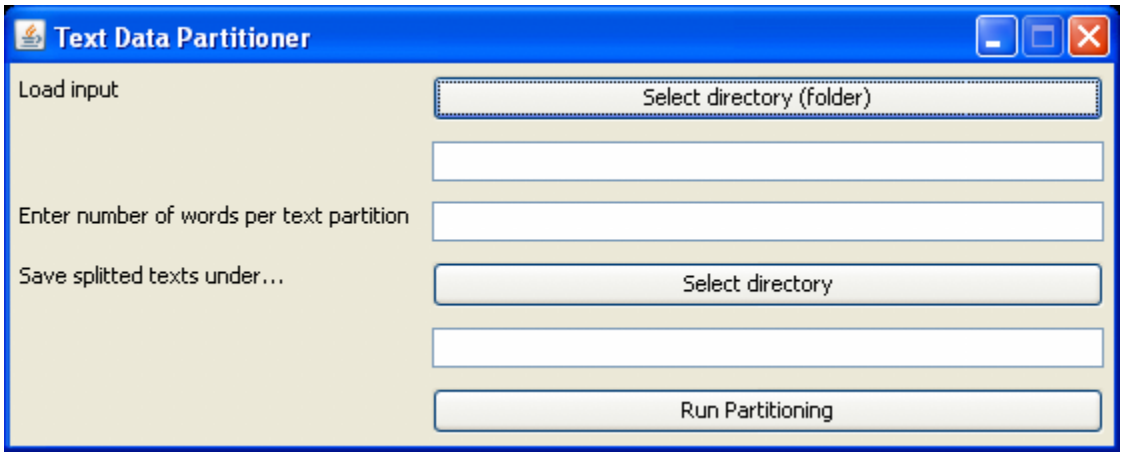
22 Text Data Preparation Tools

Text Set Comparison :
Text Set Editor:
Text File Splitter :
Text Deduplicator :

20 Text Data Preparation Tools

- Text Set Comparison :
- Text Set Editor:
- Text File Splitter :
- Text Deduplicator :

21 Split Input Texts



The File Splitters splits large text files into smaller ones of minimum equal size. It's used in order to speed up AutoMap coding. The number of words (NW) is the size that each of the number of text (NT) text file should contain after splitting.

From the main menu select *Tools => Open Text File Splitter*.

Input : 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 : NT 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 NT, where NT indicates the largest number of texts that an original text had been split up into.

22 Text Data Preparation Tools

- Text Set Comparison :
- Text Set Editor:
- Text File Splitter :
- Text Deduplicator :

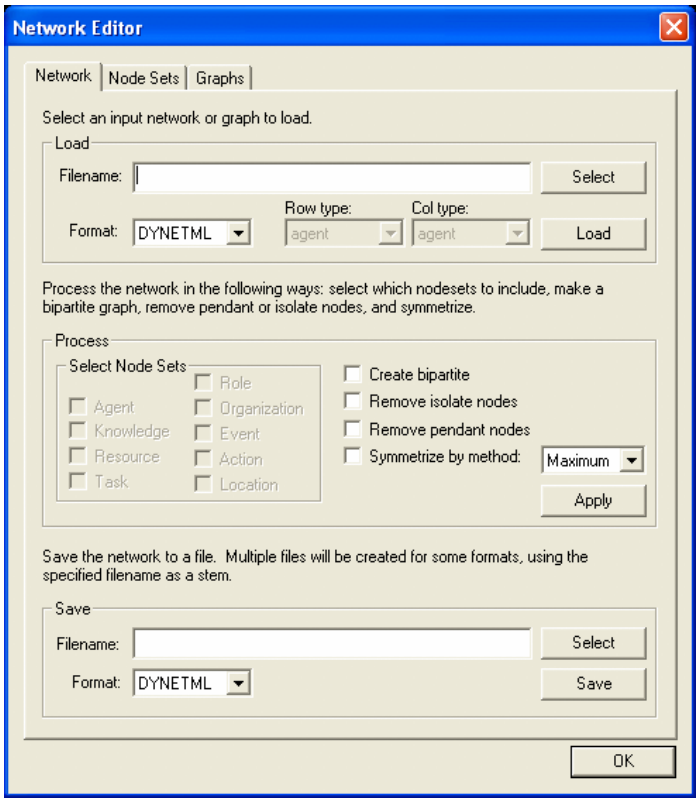
23 Merge DyNetML

The Merge DyNetML function takes multiple DyNetML files and outputs one DyNetML file.

From the main menu select *Tools => Open DyNetML File Merger*.

This could be used when DyNetML files that were generated per text during Sub-Matrix Text Analysis and need to be consolidated into one DyNetML file that represents the entire text set. The Output is a single DyNetML file

24 Convert Network Data



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

From the main menu select *Tools => Open Matrix Editor*.

25 Other Tools

- CEMAP (CASOS Email Parser) :
- CompareMap :
- AutoMap to Database Connector :
- Network Visualizer :
- Matrix Editor :

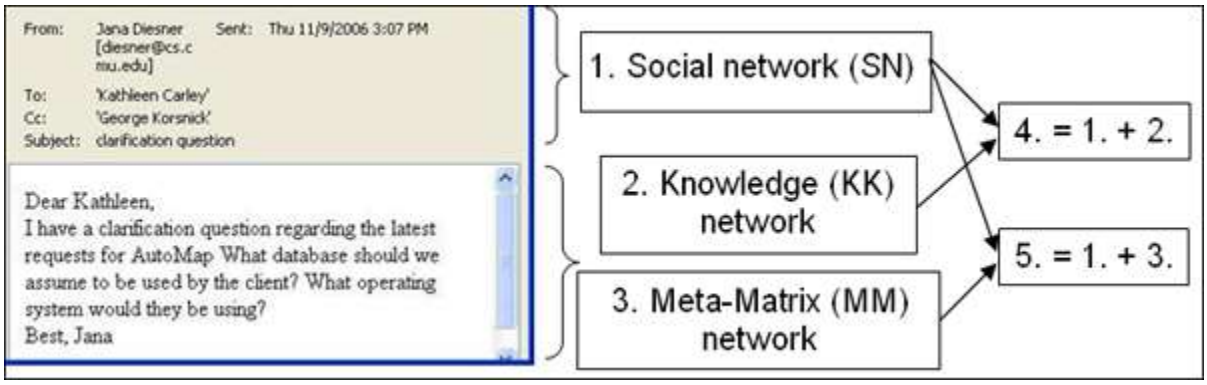
26 CASOS Email Parser (CEMAP)

CEMAP is an ORA tool that primarily allows you to pull network data in the form of emails into ORA. However, the CEMAP tool could be used to convert any sources of data in the world into DyNetML, ORA's native file format, for analysis.

Setting up CEMAP is similar to setting up a new ISP mail client. What follows are procedures to configure this tool and how to use it.

Each email has an identifying number in a long number showing the date, for example: [20060308042751s1500bii66e]. The 20060308 is in the year/month/day format and says this email was sent on March 8, 2006.

The CASOS Email Parser (CEMAP) that is launched through AutoMap/ORA (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 type1 Social network (SN) represents social network data that can be extracted from email headers (e.g. agent-agent networks). In Social Networks, 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 Meta Network (MM), texts are coded in AutoMap according to a taxonomy or ontology (e.g. Meta Network, 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).

26.1 The Top textfields

- Input File -or- HTTP URL

  Input File -or- HTTP URL – Input Data Set. You specify the email client, inbox type, username, and password as semicolon separated values. The location can be either file or server (e.g. pop.gmail.com;INBOX;JohnDoe;JohnDoesPassword).

26.1.1 NOTE : The following information is what is needed to access an email account.

- Server – The IP address or URL of the server.
- Port – The port on the server to connect to.
- UserID – ID of the User.
- Password – The User's Password for accessing the Mailbox.
- Mailbox/Folder – The ID of the mailbox to use.

  lip>File Directories

26.2 The Directory Buttons

- This section allows the user to dictate where the end files will be written.
- Select Output Directory for full document bodies – Directory to save the body of the email text.
- Select Output Directory for document bodies (new content Only) – Directory to save body of the email. Discards all the repeat material that is quoted from previous emails.
- Select Output Directory for document bodies (forward content Only) – Directory to save Forwarding content.
- Select Output Directory for SubjectLine Bodies – Directory to save the Subject Lines of the email.
- Select Output Filename for Header Network file – Directory to write a single containing the Header.

26.3 Checkboxes at the bottom

- Anonymize Agents – Anonymize Agents changes the agents name (only agents) to A# (# is a number) in DyNetML.
- SSL (Secure Sockets Layer) (For IMAP and POP3 only) – Creates a secure connection between a client and a server, over which any amount of data can be sent securely.
- De-Dupe Messages – Ignores identical emails.

  The file prefix makes it possible to put all the files in the same directory

26.4 The Start Button

The Start Extraction button begins the process.

26.5 Extraction of network type 1 data:

1. This network type is needed to perform social network analysis (SNA) based on header information only.
2. In CEMAP, Analysis Type, select the type of your mailbox.
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
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.
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.
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.
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).

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

To extract networks of type 3, perform Meta Network 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.

26.6 To use the feature:

- In the AutoMap menu bar, select Tools => Open CASOS Email Parser (CEMAP).
- In the ORA menu bar, select File => CEMAP Email-Importer….

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.

26.7 The different types of imports are

- Full email
- Header information only
- Forwarded/New emails
- Subject lines only

26.8 Type Selection

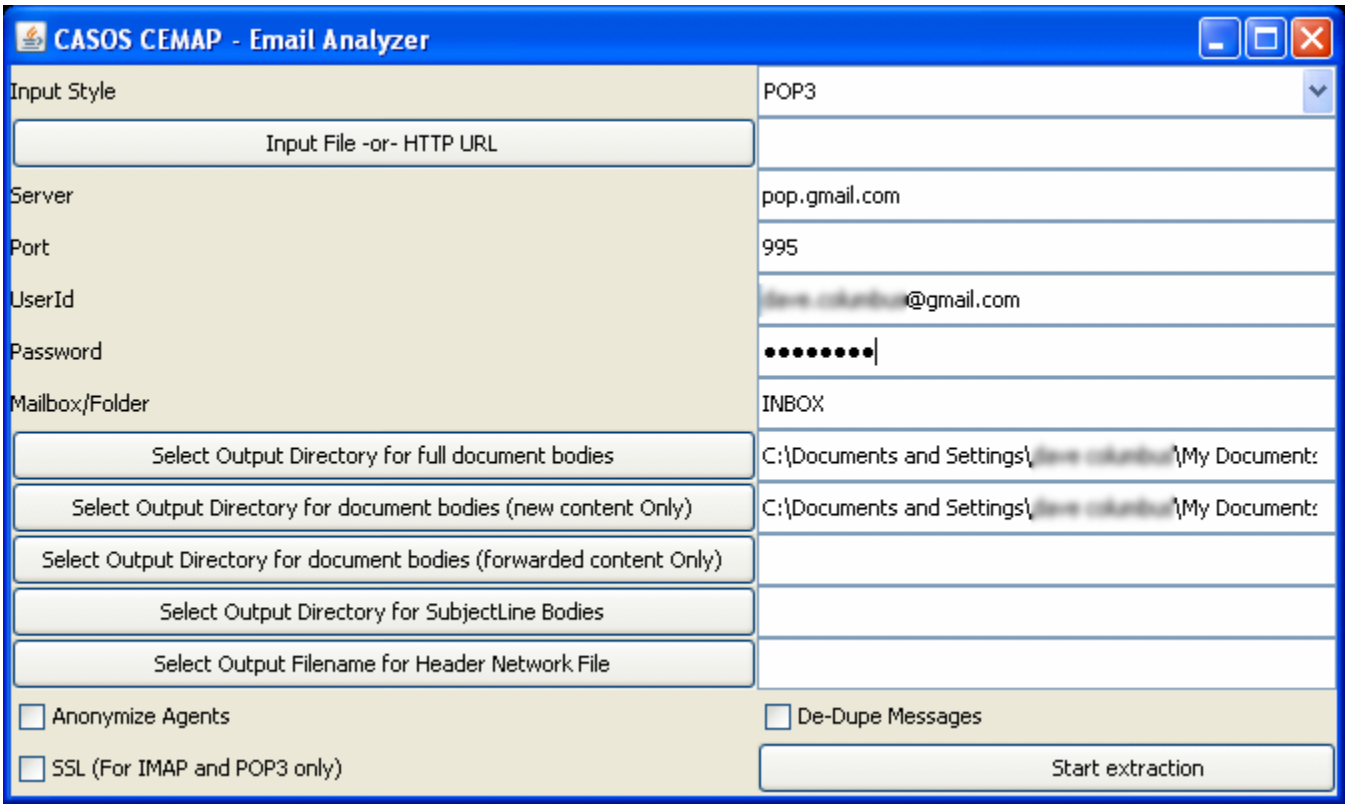
Select the type of Format your email is in, from the drop down menu.

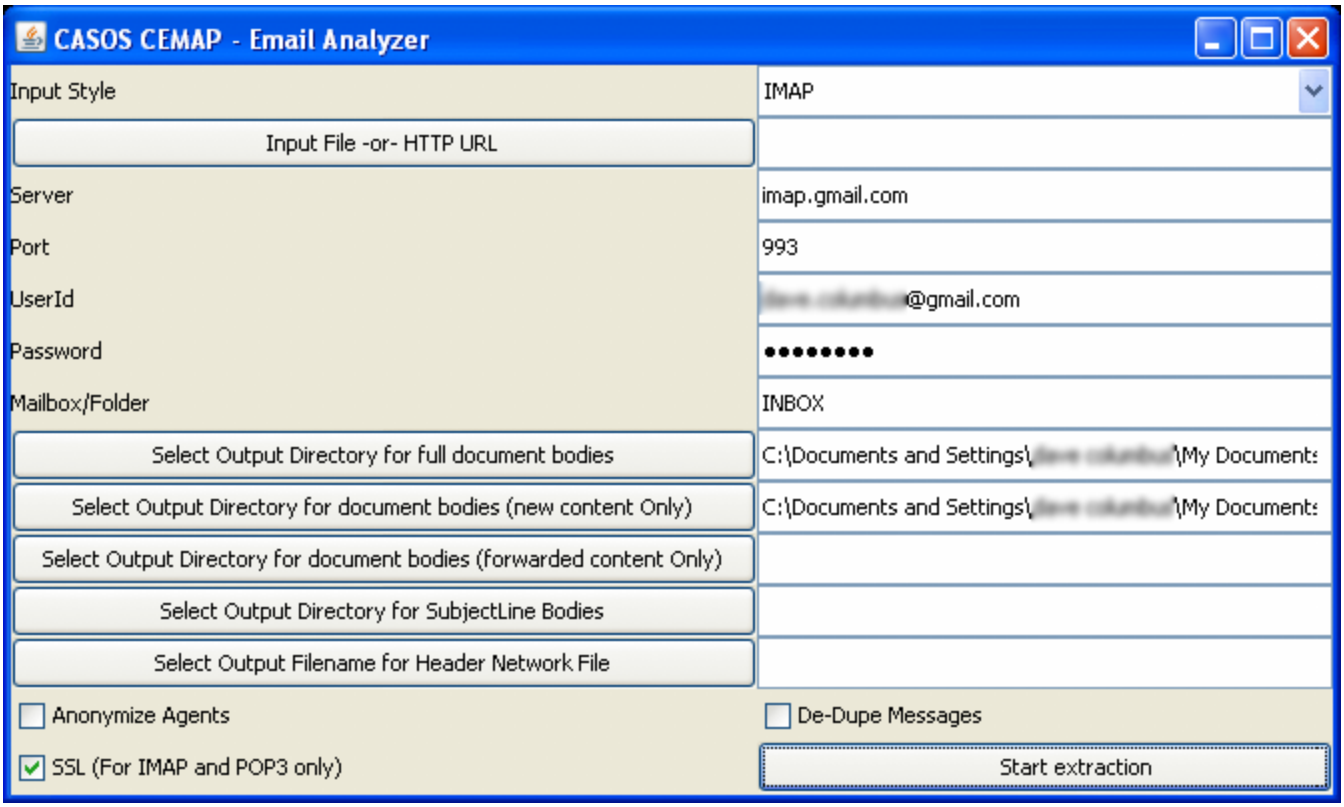
- MBOX – A generic term for a family of related file formats used for holding collections of electronic mail messages. All messages in an mbox mailbox are concatenated and stored as plain text in a single file. The beginning of each message is indicated by a line whose first five characters are "From" (the so-called "From_ line" or "'From' line") and a blank line is appended to the end of each message. For a while, the mbox format was popular because text processing tools can be readily used on the plain text files used to store the e-mail messages. This is want you see when you View Source in a browser.
- POP3 – **(Post Office Protocol)** version 3 (POP3) is an application-layer Internet standard protocol to retrieve e-mail from a remote server over a TCP/IP connection.
- IMAP – **(Internet Message Access Protocol)** Is an application layer Internet protocol operating on port 143 that allows a local client to access e-mail on a remote server.
- TXT – Exported email files from outlook – will require a Data (file) set identifier
- CasosDB – for internal use only
- DyNetml – Proprietary format for ORA.

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.

26.9 Examples of POP3 and IMAP setups

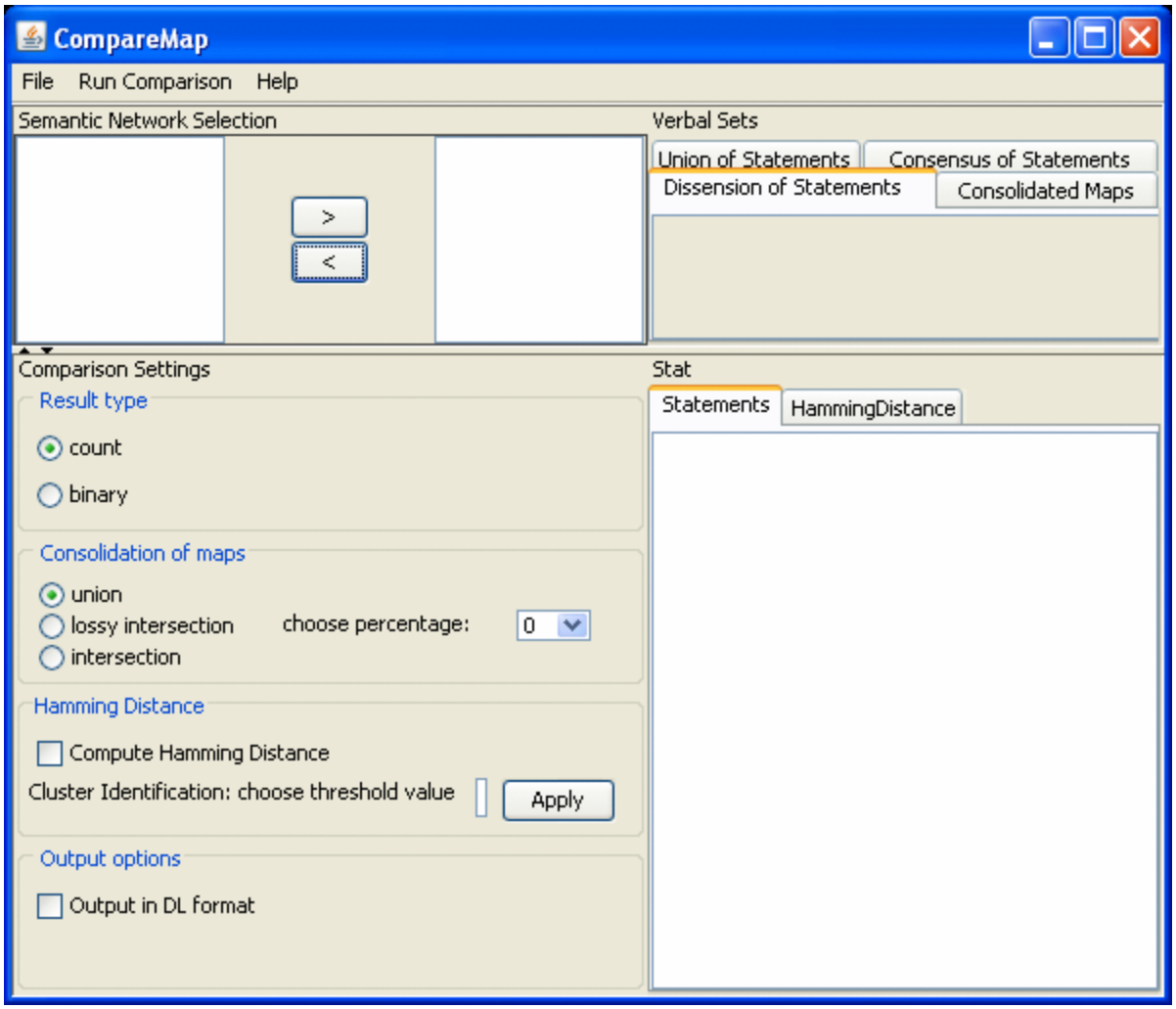
Below are examples of setting up CEMAP with POP or MAP servers. Refer to the documentation with a particular server regarding the exact Server, Port, or other pertinent information necessary.





27 Compare Maps

From the main menu select *Tools ⇒Open CompareMap*.



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.

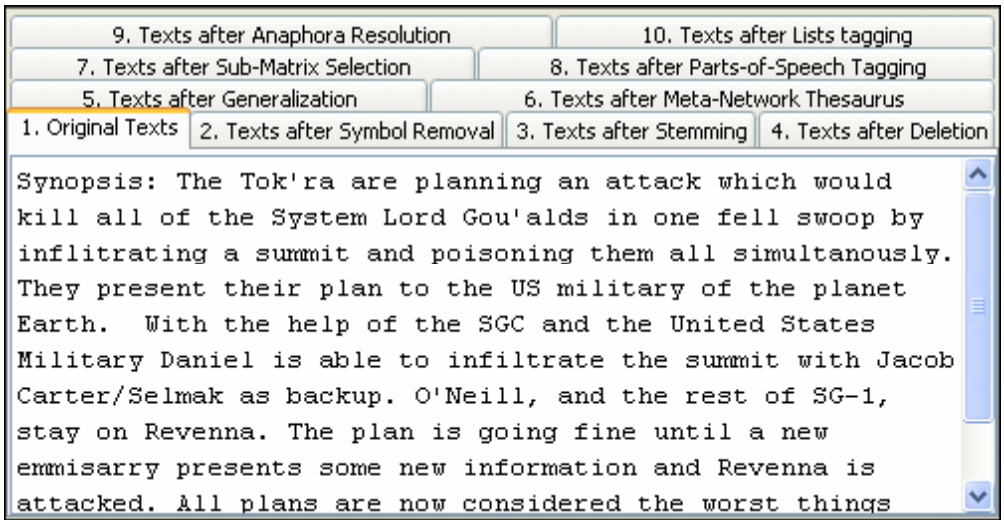
28 Visualize Semantic Network

Purpose : Visualize mental models and social structure.
How To : From the main menu select *Tools => Open SocialInsight Visualizer*.

28.1.1 NOTE : When using an External Tool convert the DyNetML or DL files generated with AutoMap into formats required by the selected external tool. (i.e. Convert DL into VNA files can be visualized in NetDraw. In NetDraw open VNA file: File => Open => VNA text file => complete.)

29 Text Display Panel

The *Text Display Panel* is a display panel only. The user cannot change any text that appears here. It will display texts as they are altered from functions in the *Text Processing Panel*.



1. Original Text : Shows the loaded text(s) as they were originally, without any modifications made by any AutoMap functions.
2. Texts after Symbol Removal : Display text of the file after the Symbols Removal option has been applied.
3. Texts after Stemming : Display text of the file after Stemming has been performed on it.
4. Texts after Deletion : Display text of the file after Deletion has been performed on it.
5. Texts after Generalization : Display text of the file after a Generalized Thesaurus has been applied.
6. Texts after Meta Network Thesaurus : Display text of the file after a Meta Network Thesaurus has been applied.
7. Texts after Sub-matrix Selection : Display text of the file after a Sub-Matrix Thesaurus has been applied.
8. Texts after Parts-of-Speech Tagging : Display text of the file after the Parts of Speech option has been applied.

9. Texts after Anaphora Resolution : Display text of the file after the Anaphora Resolution option has been applied.
10. Texts after Lists Tagging: Display text of the file after the Lists option has been applied.

30 Text Processing Panel

The *Text Processing Panel* is where the user inputs commands and performs functions on the text. The other three panels reflect whatever is done in panel 2.

31 Text Processing Panel - Concept List

31.1 The Concept List

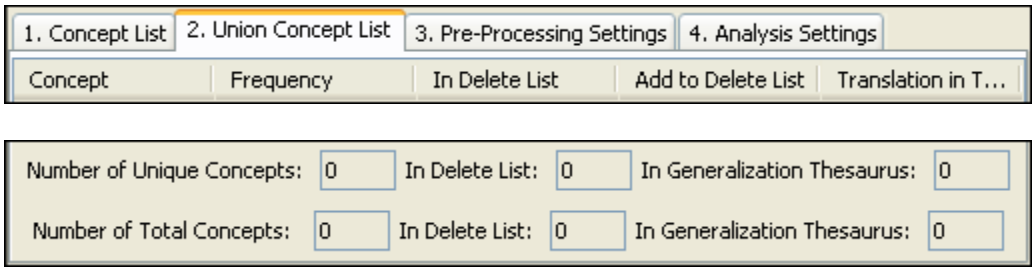
After loading text a Concept List is automatically generated and can be viewed in *Text Processing Panel => Tab 1. Concept List*. The list defaults to the order of Frequency and Concept. You can change the order by clicking the either of those headers. The concept list tells you several important facts about your text:

- The number of concepts found in the text displayed in *Text Display Panel => Tab 1. Original Text*.
- Related frequencies of those concepts.
- Cumulated unique concepts and total concepts contained in the data set.

31.1.1 NOTE : 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.

32 Text Processing Panel - Union Concept List



32.1 Create and refresh Union Concept List

The Union Concept List differs from the Concept List in that it considers concepts across all the texts currently loaded in Automap, rather than only the currently selected 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 in *Text Processing Panel => Tab 2 Union Concept List*.

The union concept list must be refreshed by selected *File => Refresh Union Concept List* before viewing your results. 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. Below is an example of the text and union concept list.

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.

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

32.2 Save Union Concept List

From the main menu select *File => Save Union Concept List*. A file chooser will pop up. Navigate to a directory and save your file.

32.3 Example for Union Concept List

File 1

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

File 2

On March 22, 2002 the Ren'al of the Tok'ra arrived @ the SGC to discuss their plan. 5 days later, on March 27, 2002, SG-1 and SG-17 arrive @ Revenna. The summit meeting with the 6 main System Lords* will be happening in 10 days. Daniel is faced with the question, "Will this prevent the System Lords from overthrowing the galaxy?"

*Gou'alds

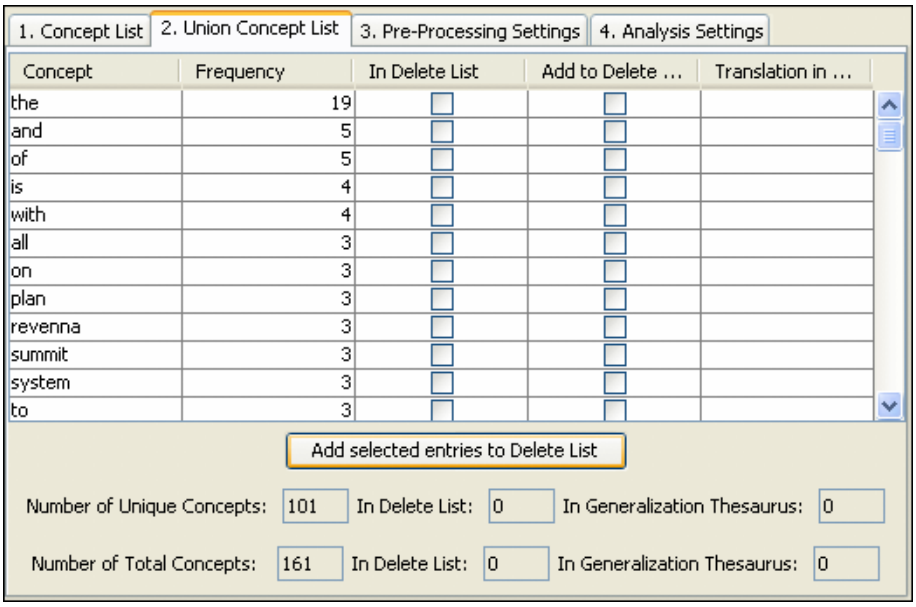
The first step is to load in your text files. This can be done individually or using the *Open Multiple Files* function. In order to use the *Open Multiple Files* move all your files into a directory which contains no other text files.

From the main menu select *File => Open Multiple Files* and navigate to the folder in which you placed your text files. Then select [Open].

To run the Create and Refresh Union Concept Utility based on multiple texts select *File => Create and Refresh Union Concept List*

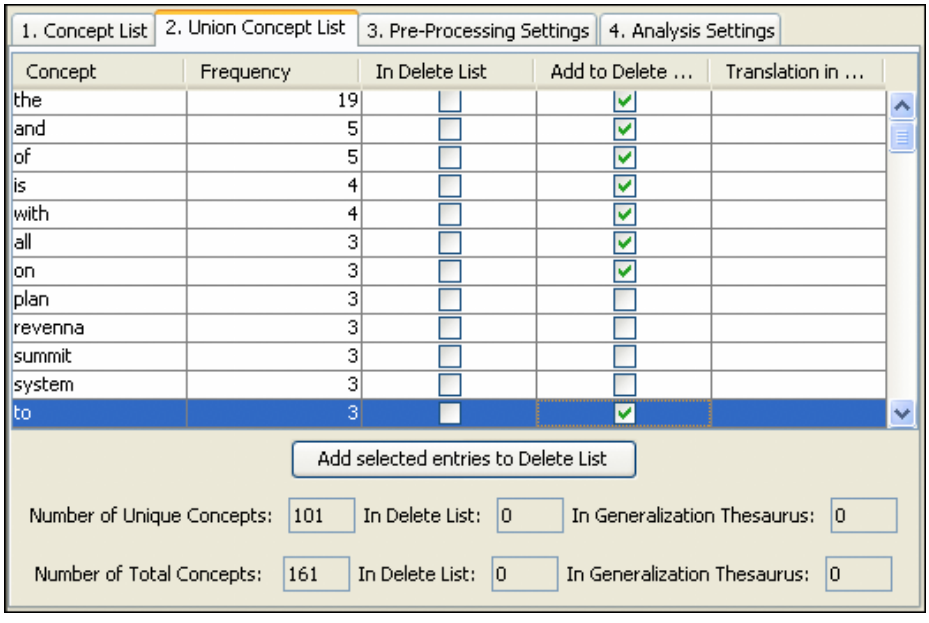
To save your list select from the main menu *File => Save Union Concept List*.

The result is shown in the screen shot below:

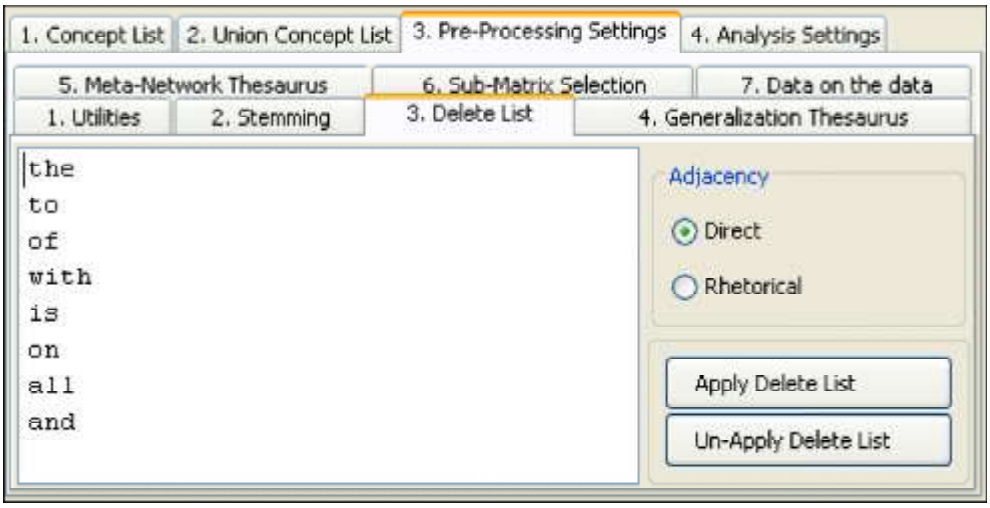


The table is ordered by Frequency but can be changed by clicking in any of the headers. The concept list contains more entries than the interface can display. The displayed Union Concept List indicates there are 101 unique concepts and 161 total concepts in the data set.

32.4 Create a Delete List



Place a checkmark in the boxes for each concept to delete in the Add to Delete… column. When finished, select the [Add selected entries to Delete List]. The concepts pegged for the Delete List now show up in *Text Processing Panel => Tab 3. Pre-Processing Settings => SubTab 3. Delete List*.

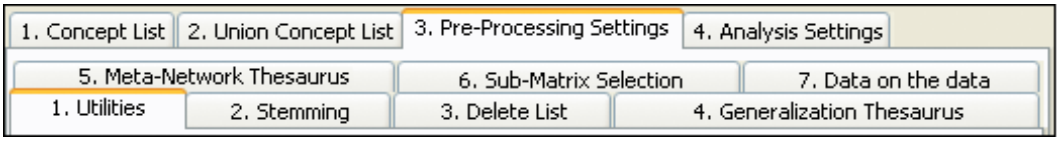


33 Text Processing Panel

Text Processing Panel - PreProcessing is where most of the work is performed.

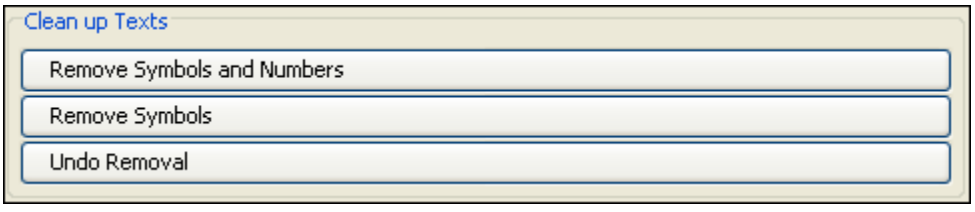
34 Text Processing Panel - Utilities

Holds the tools to work with the text files.



This function removes all characters that are neither a letter nor a number depending on the option chosen. It maintains sentence marks converting 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.

Clean up Texts contains three buttons.



1. Remove Symbols and Number : Removes both numbers and sentence marks.
2. Remove Symbols : Removes sentence marks but retains the numbers
3. Undo Remove : Restores Symbols and/or numbers to the text.

The results of Clean up Text are found in *Text Display Panel => Tab 2. Text after Symbol Removal*.

34.1 Text before Removal of Symbols and Number

On March 22, 2002 the Ren'al of the Tok'ra arrived @ the SGC to discuss their plan. 5 days later SG-1 and SG-17 arrive @ Revenna. The summit meeting with the 6 main System Lords* will be happening in 10 days. Daniel is faced with the question, "Will this prevent the System Lords from overthrowing the galaxy?"

*Gou'alds

34.2 Text after Removal of Symbols and Number

On March the Renal of the Tokra arrived the SGC to discuss their plan. days later SG and SG arrive Revenna. The summit meeting with the main System Lords will be happening in days. Daniel is faced with the question Will this prevent the System Lords from overthrowing the galaxy.

Goualds

This function removed all the numbers 22, 2002, 5, 1, 17, & 6. If also removed the punctuation commas, hyphens, & quote marks and the symbols @, and *

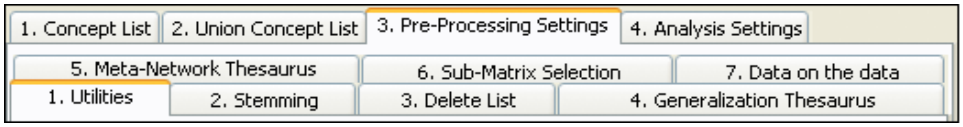
34.2.1 NOTE : Question marks are changed to periods

34.3 Text after Removal of Symbols

On March 22 2002 the Renal of the Tokra arrived the SGC to discuss their plan. 5 days later SG1 and SG17 arrive Revenna. The summit meeting with the 6 main System Lords will be happening in 10 days. Daniel is faced with the question Will this prevent the System Lords from overthrowing the galaxy.

Goualds

35 Text Processing Panel - Bi-Gram Detection



35.1 N-gram Identification: Bigrams

Creates a list of all bigrams* that are contained in the data set opened. The resulting list will be automatically saved under CorrelationList.csv in the root directory of AutoMap.

35.2 Text Example

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

35.3 Results after Bi-Gram list created

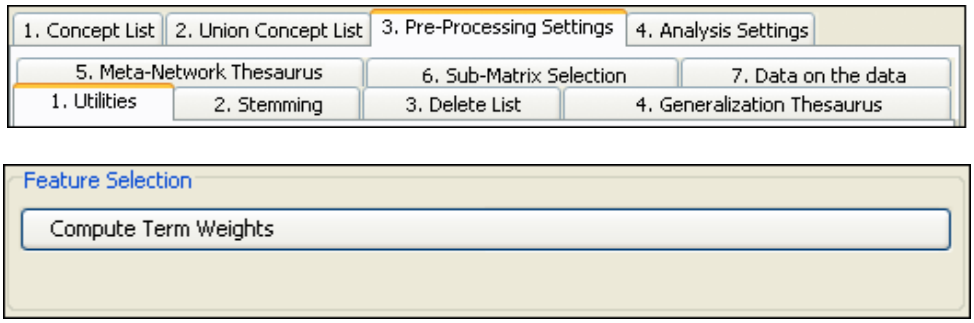
After pressing the Create Bi-Gram List button a file in the AutoMap directory will be created like below.

Frequency,Correlation,Concept,Concept,
3,1.0,of,the
2,0.6666667,and,the
1,0.33333334,a,new
1,0.33333334,a,summit
1,0.33333334,able,to
1,0.33333334,all,of
1,0.33333334,all,plans
1,0.33333334,all,simultanously
1,0.33333334,an,attack
1,0.33333334,and,poisoning
1,0.33333334,and,revenna
1,0.33333334,are,now
1,0.33333334,are,planning
1,0.33333334,as,backup
1,0.33333334,attack,which
1,0.33333334,attacked,all
1,0.33333334,backup,o'neill
1,0.33333334,by,inflitrating
1,0.33333334,carterselmak,as
1,0.33333334,considered,the
1,0.33333334,daniel,is
1,0.33333334,earth,with
1,0.33333334,emmisarry,presents
1,0.33333334,fell,swoop
1,0.33333334,fine,until
1,0.33333334,going,fine
1,0.33333334,gou'alds,in
1,0.33333334,help,of
1,0.33333334,in,one
1,0.33333334,infiltrate,the
1,0.33333334,inflitrating,a
1,0.33333334,information,and
1,0.33333334,is,able
1,0.33333334,is,attacked
1,0.33333334,is,going
1,0.33333334,jacob,carterselmak

1,0.33333334,kill,all
1,0.33333334,lord,gou'alds
1,0.33333334,military,daniel
1,0.33333334,military,of
1,0.33333334,new,emmisarry
1,0.33333334,new,information
1,0.33333334,now,considered
1,0.33333334,o'neill,and
1,0.33333334,of,sg-1
1,0.33333334,on,revenna
1,0.33333334,one,fell
1,0.33333334,plan,is
1,0.33333334,plan,to
1,0.33333334,planet,earth
1,0.33333334,planning,an
1,0.33333334,plans,are
1,0.33333334,poisoning,them
1,0.33333334,present,their
1,0.33333334,presents,some
1,0.33333334,rest,of
1,0.33333334,revenna,is
1,0.33333334,revenna,the
1,0.33333334,sg-1,stay
1,0.33333334,sgc,and
1,0.33333334,simultanously,they
1,0.33333334,some,new
1,0.33333334,states,military
1,0.33333334,stay,on
1,0.33333334,summit,and
1,0.33333334,summit,with
1,0.33333334,swoop,by
1,0.33333334,synopsis,the
1,0.33333334,system,lord
1,0.33333334,the,help
1,0.33333334,the,plan
1,0.33333334,the,planet
1,0.33333334,the,rest
1,0.33333334,the,sgc
1,0.33333334,the,summit
1,0.33333334,the,system
1,0.33333334,the,tok'ra

1,0.33333334,the,united
1,0.33333334,the,us
1,0.33333334,the,worst
1,0.33333334,their,plan
1,0.33333334,them,all
1,0.33333334,they,present
1,0.33333334,things,possible
1,0.33333334,to,infiltrate
1,0.33333334,to,the
1,0.33333334,tok'ra,are
1,0.33333334,united,states
1,0.33333334,until,a
1,0.33333334,us,military
1,0.33333334,which,would
1,0.33333334,with,jacob
1,0.33333334,with,the
1,0.33333334,worst,things
1,0.33333334,would,kill

36 Text Processing Panel - Feature Selection



From the main menu load a file using *File => Open single text file*.
From the main menu select *Create Union Concept List*.
From the main menu select *Output Storage Manager*. Locate the line for TFIDF computation.

36.1.1 NOTE : In Windows the simplest method for viewing the file is to highlight the pathway in the Output Storage Manager and copy it with [Ctrl-C]. Go to an explorer window and put this pathway in the address bar and press [Enter]. The file will automatically open

36.2 Opening the file in Excel

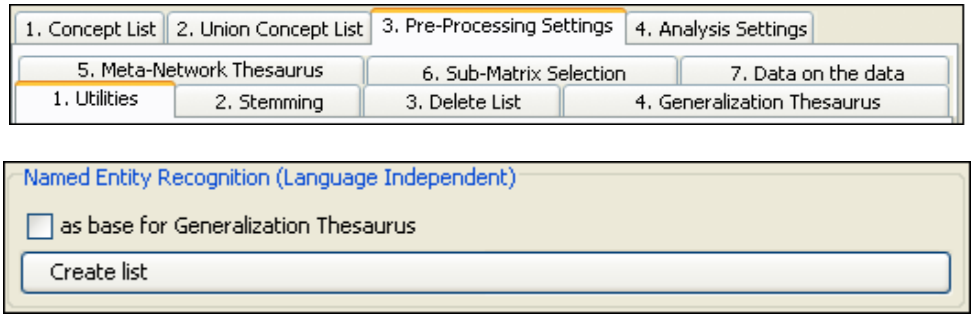
The text file will be in a comma-delimited format.
Open up Excel with a new spreadsheet.
In AutoMap highlight and copy the entire file. Paste this into Excel.
Highlight the column which contains the data. From the main menu select *Tools => Text to Columns…*.

1. Select the Delimited radio button and hit [Next].
2. Make sure that only the Comma checkbox is checked. Make sure in the view area that the information is separated correctly. Press [Next]
3. Press [Finish]

The data can now be sorted by TF*IDF (Term frequency by Inverse Document Frequency) descending order.

36.2.1 This is an Alternative Method to find most important terms or features in text or text set.

37 Named Entity Recognition



Named-Entity Recognition 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.

37.1 Text Example

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

In *Text Processing Panel => Tab 3. Pre-Processing settings => SubTab 1. Utilities* under Named Entity Recognition select the Create List button.
This will create a Named Entity list which can be found in the directory specified in the Output Storage Manager.

37.2 NamedEntities.csv

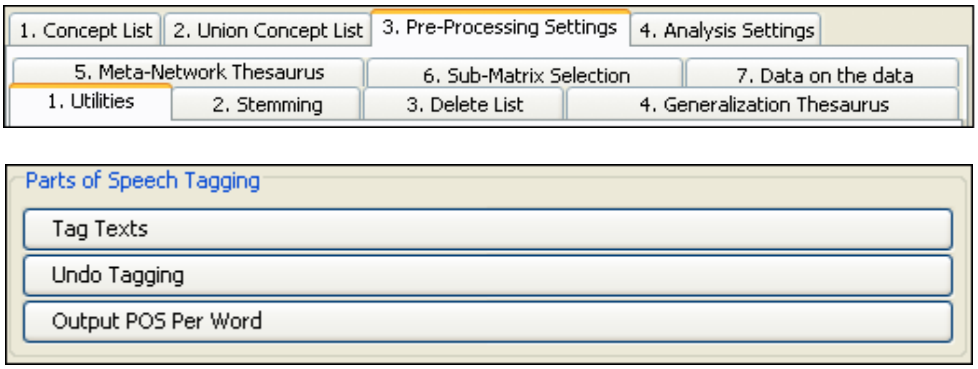
A .csv file is created. The csv file can be edited in a text editor.
Jacob CarterSelmak
US
SG
System Lord Goualds
Earth. With
Synopsis The Tokra
ONeill
All
United States
Revenna. The
Jacob CarterSelmak
System Lord
They
United States Military Daniel
Revenna
SGC

37.3 as base for Generalization Thesaurus

If a checkmark is placed in the box before the button is selected the NamedEntities.csv file takes on a different format as used in the Generalization Thesaurus.
Jacob CarterSelmak,,Jacob CarterSelmak/ Jacob CarterSelmak
US,,US/ US
SG,,SG/ SG
System Lord Goualds,,System Lord Goualds/ System Lord Goualds
Earth. With,,Earth. With/ Earth. With
Synopsis The Tokra,,Synopsis The Tokra/ Synopsis The Tokra
ONeill,,ONeill/ ONeill
All,,All/ All
United States ,,United States / United States
Revenna. The,,Revenna. The/ Revenna. The
Jacob CarterSelmak ,,Jacob CarterSelmak / Jacob CarterSelmak
System Lord ,,System Lord / System Lord
They,,They/ They

United States Military Daniel,,United States Military Daniel/ United States Military Daniel
Revenna,,Revenna/ Revenna
SGC,,SGC/ SGC

38 Text Processing Panel - Parts of Speech Tagging



This feature will show you what part of speech each word is. This routine associates every word after the highest level of pre-processing applied so far with its Parts of Speech. 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.

- The Tag Texts option will take the text document and define each part of text.
- The Undo Tagging button removes any tags in the text
- The Output POS Per Word.

38.1 Text Example

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

Click the Tag Text button and view the results in *Panel 1 => Tab 8. Texts after Parts-of-Speech Tagging*.

Synopsis/VERB The/NOISE Tokra/NN are/VERB planning/VERB an/NOISE attack/NN which/NOISE would/MODAL kill/VERB all/NOISE of/NOISE the/NOISE System/NNP Lord/NNP Goualds/VERB in/NOISE one/NUM fell/VERB swoop/NOISE by/NOISE/NN a/NOISE summit/NN and/NOISE poisoning/VERB them/ANA all/NOISE./NOISE ./. They/ANA present/VERB their/NOISE plan/NN to/NOISE the/NOISE US/NOISE military/ADJ of/NOISE the/NOISE planet/NN Earth./NOISE ./. With/NOISE the/NOISE help/VERB of/NOISE the/NOISE SGC/NN and/NOISE the/NOISE United/NNP States/NNPS Military/NNP Daniel/NNP is/VERB able/ADJ to/NOISE infiltrate/NN the/NOISE summit/NN with/NOISE Jacob/NNP CarterSelmak/VERB as/NOISE backup./NOISE ./. ONeill/VERB and/NOISE the/NOISE rest/NN of/NOISE SG/NN stay/VERB on/NOISE Revenna./NOISE ./. The/NOISE plan/NN is/VERB going/VERB fine/ADJ until/NOISE a/NOISE new/ADJ emmisarry/NN presents/VERB some/NOISE new/ADJ information/NN and/NOISE

Revenna/NN is/VERB attacked./NOISE ./. All/NOISE plans/NN are/VERB now/NOISE considered/VERB the/NOISE worst/ADJ things/NN possible./NOISE ./.

To remove the tags select the Undo Tagging button.

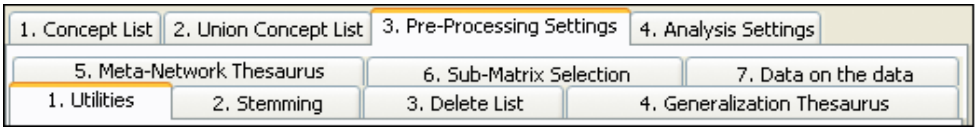
To save the Parts-of-Speech as a .csv file use the Output POS Per Word button. This file can be opened in Excel.

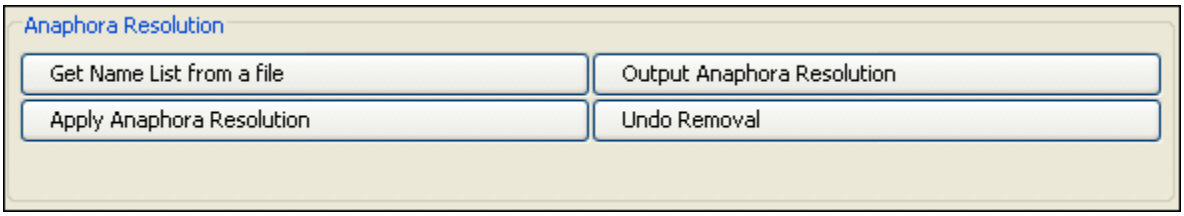
Below is a table of the Parts of Speech used

Code	Part of Speech	Code	Part of Speech
CC	Coordinating conjunction	PRP$	Possessive pronoun
CD	Cardinal number	RB	Adverb
DT	Determiner	RBR	Adverb, comparative
EX	Existential there	RBS	Adverb, superlative
FW	Foreign word	RP	Particle
IN	Preposition or subordinating conjunction	SYM	Symbol
JJ	Adjective	TO	to
JJR	Adjective, comparative	UH	Interjection
JJS	Adjective, superlative	VB	Verb, base form
LS	List item marker	VBD	Verb, past tense
MD	Modal	VBG	Verb, gerund or present participle
NN	Noun, singular or mass	VBN	Verb, past participle
NNS	Noun, plural	VBP	Verb, non-3rd person singular present
NNP	Proper noun, singular	VBZ	Verb, 3rd person singular present
NNPS	Proper noun, plural	WDT	Wh-determiner
PDT	Predeterminer	WP	Wh-pronoun
POS	Possessive ending	WP$	Possessive wh-pronoun
PRP	Personal pronoun	WRB	Wh-adverb

39

40 Text Processing Panel - Anaphora Resolution





Get Name List from a file : Pressing the button brings up a dialog box from where a filename can be entered. This file will contain a list of predefined human names which the Anaphora Resolution will use to recognize whom relate to the pronouns. This feature is optional. If not used then Automap will use the Parts of Speech function to recognize words related to pronouns.

Anaphora is a linguistic instance of a grammatical expression referring to another. 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.

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

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 (he refers to Larry).

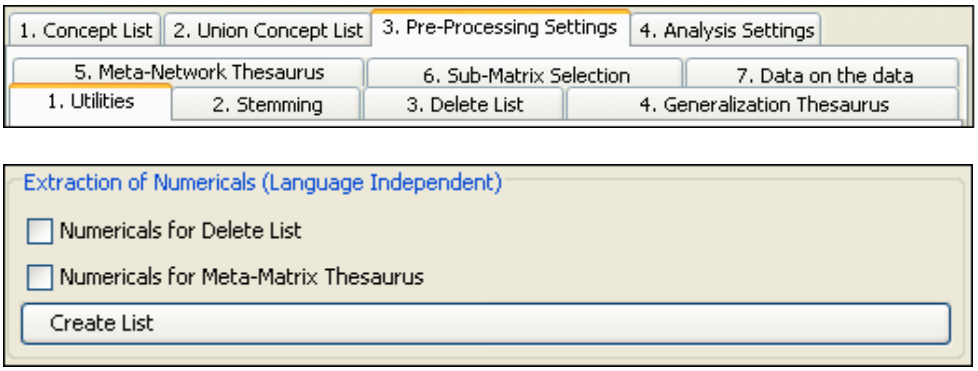
Moreover, think of a sentence being taken out of context and the missing information required to fully explain the meaning.

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.

The results can be viewed in *Text Display Panel => Tab 9. Texts after Anaphora Resolution*.

41 Text Processing Panel - Extraction of Numerals



41.1 Text Example

On March 22, 2002 the Ren'al of the Tok'ra arrived @ the SGC to discuss their plan. 5 days later, on March 27, 2002, SG-1 and SG-17 arrive @ Revenna. The summit meeting with the 6 main System Lords* will be happening in 10 days. Daniel is faced with the question, "Will this prevent the System Lords from overthrowing the galaxy?"

*Gou'alds

41.2 Extraction of Numerals for Delete List

To extract the numerals from a text place a checkmark in the box labeled Numerals for Delete List then click the Create List button. Unless you've changed the destination in the Output Storage Manager AutoMap will place it in the root directory of AutoMap. In the Output Storage Manager you have the opportunity to change the name for the file. The extraction file contains the following numerals:

22
2002
5
27
2002
1
17
6
10

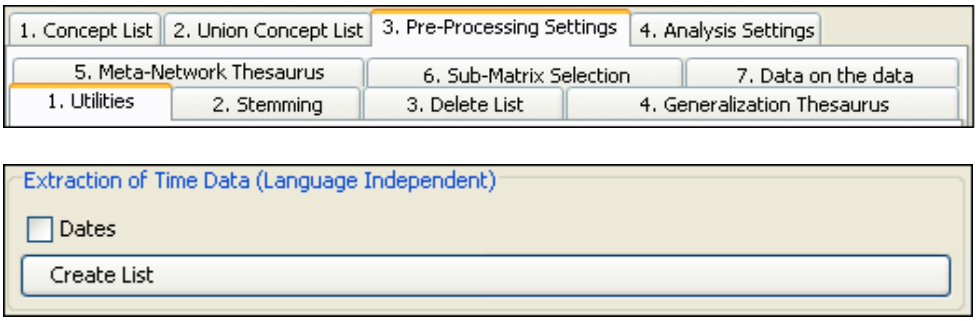
The file is saved as a .txt file which can be opened with a text editor.

41.3 Extraction of Numerals for Meta Network Thesaurus

The extraction of Numerals for Meta Network Thesaurus works in the same manner as above. Only the name of the file and the destination directory will be different as designated in the Output Storage Manager.

The file is saved as a .csv file which can be opened in Excel.

42 Text Processing Panel - Extraction of Time Data



42.1 Text Example for Extraction of Date

On March 22, 2002 the Ren'al of the Tok'ra arrived @ the SGC to discuss their plan. 5 days later, on March 27, 2002, SG-1 and SG-17 arrive @ Revenna. The summit meeting with the 6 main System Lords* will be happening in 10 days. Daniel is faced with the question, "Will this prevent the System Lords from overthrowing the galaxy?"

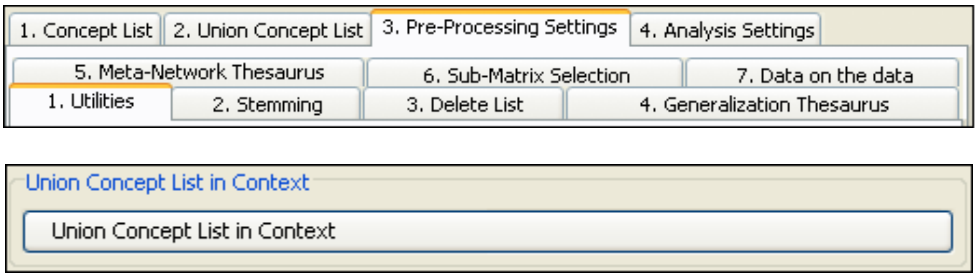
*Gou'alds

AutoMap creates a .csv file in the AutoMap directory which can be opened in Excel. The final location and filename can be altered in the Output Storage Manager

March 22, 2002

March 27, 2002

43 Text Processing Panel - Union Concept List in Context



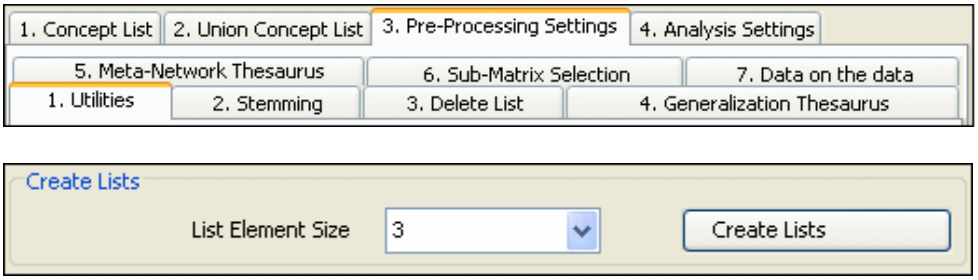
This function requires loading multiple files. Form the main menu select *File => Open multiple text files*. Navigate to the directory that contains the files and select [OK].

From the main menu select *File => Run Create Union Concept List.*

From the *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 1. Utilities* scroll down to the Union Concept List in Context section. Press the [Union Concept List in Context] button to create the list.

This function gives the concept with # of words in front and in back.

44 Text Processing Panel - Create Lists



This functions recognizes lists in a text. A list is a sequence of a set of words divided by a separator (i.e. a comma) followed by a key word.

This bag contains apples, strawberries, yogurt, and a gallon of milk.

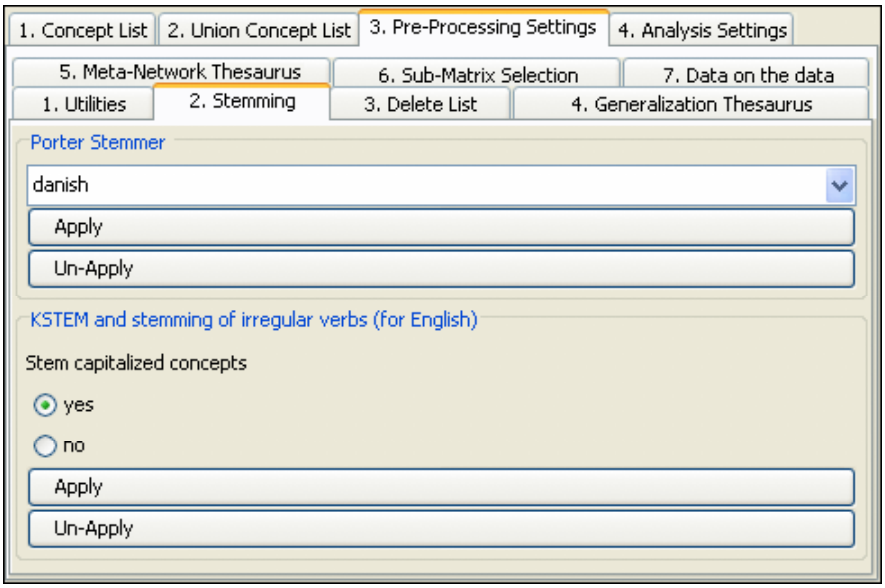
The list would be apples | strawberries | yogurt | a gallon of milk

The lists found affect the output DynetML file in the following ways. For a given list, all list members have links among each other. the "span" Create List feature must be run AFTER Parts-of-Speech processing but BEFORE Clean up texts.

List Element Size allows the choice of the maximum size (in words) of the list member. This is set to distinguish the lists from complete sentences.

The list can be found under *Text Display Panel => Tab 10. Texts after List tagging.*

45 Text Processing Panel - 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 two 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. 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 pair wise 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.

46 Stemming Example

46.1 Original Text

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

46.2 Text After Porter Stemming

Synopsi: The Tok'ra be plan an attack which be kill all of the System Lord Gou'ald in on fall swoop by inflitr a summit and poison them all simultan. Thei present their plan to the US militari of the planet Earth. With the help of the SGC and the Unite State Militari Daniel be abl to infiltr the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stai on Revenna. The plan be go fine until a new emmisarri present some new inform and Revenna be attack. All plan be now consid the worst thing possibl.

46.3 Text After KSTEM

Select YES to Stem capitalized concepts)

The Bolded words in the following two examples shows the difference between Stemming, or not stemming, capitalized words.

Using YES to Stem capitalized concepts

Synopsi: The Tok'ra be plan a attack which be kill all of the System Lord Gou'ald in one fall swoop by infiltrating a summit and poison them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the Unite State Military Daniel be able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan be go fine until a new emmisarry present some new information and Revenna be attack. All plan be now consider the worst thing possible.

Using NO to Stem capitalized concepts

Synopsis: The Tok'ra be plan a attack which be kill all of the System Lord Gou'alds in one fall swoop by infiltrating a summit and poison them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel be able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan be go fine until a new emmisarry present some new information and Revenna be attack. All plan be now consider the worst thing possible.

47 Text Processing Panel - Deletion List

Deletion List 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).

48 Text Processing Panel - Deletion List

- 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.

48.1 Four different Delete List can be opened form the menu.

- Open from file: The Opens dialog box opens and allows you to navigate to the directory with your delete list.
- Open from DB:
- Open small predefined Delete List: The small predefined Delete List is compiled of the most common words that occur in English: a, an, and, some, many, this, that, these, those, the, all, one, every.
- Open extensive predefined Delete List: An Extensive predefined Delete List contains a broader list of common words which includes the small defined delete list. a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself.

The Delete List will appear in the *Text Processing Panel*. This list is editable.

48.2 Format of the Delete List

1. The structure of a Delete List is one single Concept per line.
2. Avoid empty lines.
3. The Delete List is NOT case sensitive.
4. You can edit the Delete List within AutoMap if you wish.

48.3 Editing the Delete List

1. Editing within AutoMap: The structure of a Delete List is one single concept per line. You add concepts by typing one concept per line. Hit enter after entering a concept. Avoid empty lines.
2. Editing outside of AutoMap: Use a text editor to create a Delete List.

48.4 Save an applied Delete List

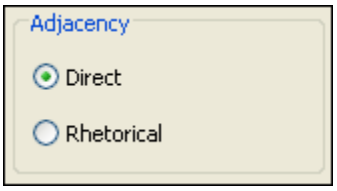
From the main menu select *File => Save Delete List As…* The Save dialog box will appear. Navigate to the directory you wish to save your delete file and select the OK button.

49 Deletion List Example

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

Extensive Delete List: a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself

50 Adjacency Options



The Adjacency section contains two radio buttons. This section is only active if the Thesaurus Content only is chosen. The default is Direct Adjacency

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

50.1 Direct Adjacency:

The concepts in the delete list are totally removed the text.

Synopsis: Tok'ra are planning attack which kill all System Lord Gou'alds in one fell swoop by infiltrating summit poisoning all simultaneously. present plan military planet Earth. With help SGC United States Military Daniel is able infiltrate summit with Jacob Carter/Selmak backup. O'Neill, rest SG-1, stay on Revenna. plan is going fine until new emmisarry presents some new information Revenna is attacked. All plans are now considered worst things possible.

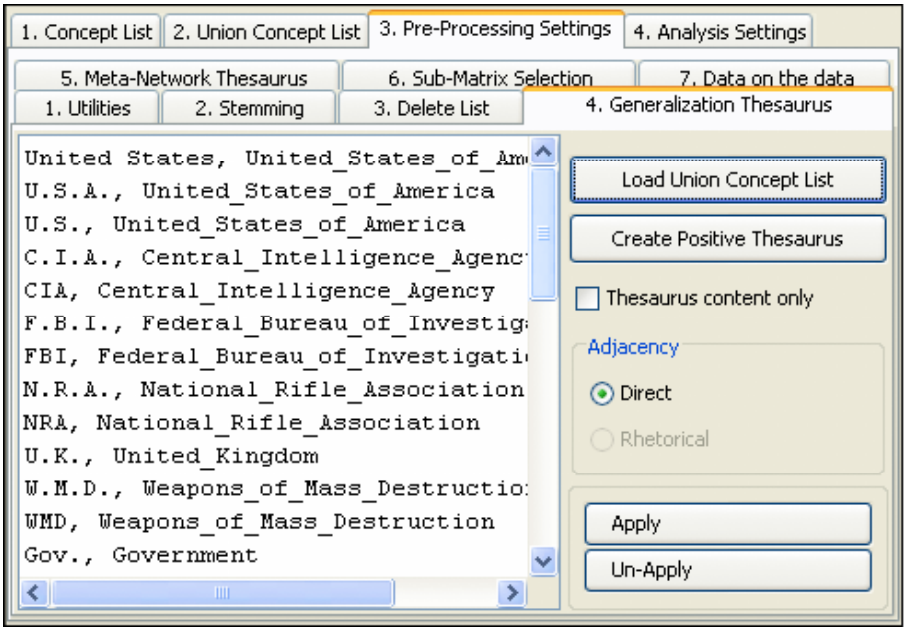
50.2 Rhetorical Adjacency:

The concepts in the delete list are replaced by xxx.

Synopsis: xxx Tok'ra are planning xxx attack which xxx kill all xxx xxx System Lord Gou'alds in one fell swoop by infiltrating xxx summit xxx poisoning xxx all simultaneously. xxx present xxx plan xxx xxx xxx military xxx xxx planet Earth. With xxx help xxx xxx SGC xxx xxx United States Military Daniel is able xxx infiltrate xxx summit with Jacob Carter/Selmak xxx backup. O'Neill, xxx xxx rest xxx SG-1, stay on Revenna. xxx plan is going fine until xxx new emmisarry presents some new information xxx Revenna is attacked. All plans are now considered xxx worst things possible.

51 Text Processing Panel - Generalization Thesaurus
52 Open a Generalization Thesaurus

From the main menu select *File => Open Generalization Thesaurus*. A file chooser will appear. Navigate to the thesaurus you want select [Open]. The thesaurus will be displayed on *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 4. Generalization Thesaurus.*



53 Editing a Thesaurus

The structure of a Thesaurus follows the five points below:

1. Every line contains Concept, Key Concept or in other words what it is in the file, new_set_of_words_to_use.
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 comma.

US Military, United_States_Military
SGC, Stargate_Command
Daniel, Doctor_Daniel_Jackson
O'Neill, Colonel_Jack_O'neill

53.1.1 NOTE : The Thesaurus is NOT case sensitive.

54 Creating a General Thesaurus

A Thesaurus can be created typing is within AutoMap, typing is in a text editor, or loading a union of concepts file. The concept list loaded into AutoMap can be refined by applying Named-Entity Recognition and Deletion prior to Generalization.

To illustrate the multi-step pre-processing technique copy the text passage below and save it as a TXT files with the names "stargateSynopsis.txt".

Text I.txt

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. It's going fine until new emmisarry present so new information and Revenna is attacked. All plans are now considered the worst things possible. .

54.1 Create Named-Entity List

From *Text Processing Panel => Tab 3. Pre-Processing Settings => SubTab 1. Utilities* scroll down to Named Entity Recognition. Select the [Create List] button.

54.1.1 NOTE : Remember to go to the Output Storage Manager to set the location of this file.

O'Neill
US
Synopsis: The Tok'ra
Earth. With
System Lord Gou'alds
All
United States
SG-1
System Lord
They
Revenna. It's
United States Military Daniel
Revenna
It's
SGC
Jacob Carter/Selmak

54.2 Build a Generalization Thesaurus

From the main menu select *File => Load Generalization Thesaurus ⇒ Open from file thes*. Navigate to the file created with the Named-Entity List. This will be the basis for a new thesaurus.

From this list build a Generalization Thesaurus by deleting unwanted concepts and editing existing concepts. After the comma place the concept name.

O'Neill, colonel_jack_o'neill
US, united_states
Earth, tauri_homeworld_earth
System Lord Gou'alds
United States, united_states
Revenna, tok'ra_homeworld
United States Military, united_states_military
Daniel, doctor_daniel_jackson
Revenna, tok'ra_homeworld_revenna
SGC, stargate_command
jacob, Jacob Carter/Selmak

54.3 Add further words

Further words can be added if desired.

summit, gou'ald_summit

55 Applying a Generalization Thesaurus

To apply a Delete List and a Generalization Thesaurus make sure to apply the Delete List first. Then execute the following process:

1. Decide whether to use the Thesaurus content only option or not. If selected do not select the Thesaurus content only option this setting will not be applied.
2. If the Thesaurus content is chosen the only option you can choose an adjacency option which can be either direct (default) or rhetorical.
3. To apply a Generalization Thesaurus with the settings specified select the Apply Thesaurus button. 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.
4. The pre-processed texts on the *Text Display Panel => Tab 5. Texts after Generalization*. 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.

55.1 Apply a Generalization Thesaurus

Once the Generalization Thesaurus is ready select the [Apply] button in *Text Processing Panel => Tab 3, Pre-Processing Settings => 4. Generalization Thesaurus*. The text with the Thesaurus applied can be viewed in *Text Display Panel => Tab 5. Texts after Generalization.*

55.2 Un-Apply a Generalization Thesaurus

To un-apply a Generalization Thesaurus that was applied to the data, go to *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 4. Generalization Thesaurus* and hit the Un-Apply button. The *Text Display Panel => Tab 5. Texts after Generalization* will be cleared.

56 Save a Generalization Thesaurus

From the main menu select *File => Output Storage Manager*. Find the line for Applied Generalization Thesaurus. Select the [Save file as…] button and select the folder to save the new thesaurus to. After giving it a name, select the [Open] button. Close the Output Storage Manager. From the main menu select *File => Save applied Generalization Thesaurus*.

The Thesaurus just created will be saved to the location designated in the Output Storage Manager.

57 Generalization Thesaurus Example

57.1 Original Text

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

57.2 Thesaurus Content Only - Off

synopsis: the tok'ra are planning an attack which would kill all of the system lord gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. they present their plan to the united_states_military of the planet earth. with the help of the stargate_command and the united states military doctor_daniel_jackson is able to infiltrate the summit with jacob carter/selmak as backup. colonel_jack_o'neill, and the rest of sg-1, stay on revenna. the plan is going fine until a new emmisarry presents some new information and revenna is attacked. all plans are now considered the worst things possible.

57.3 Thesaurus Content Only - On – Direct Adjacency

:. united_states_military. stargate_command doctor_daniel_jackson. colonel_jack_o'neill,,...

57.4 Thesaurus Content Only - On – Rhetorical Adjacency

xxx: xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx united_states_military xxx xxx xxx xxx. xxx xxx xxx xxx xxx stargate_command xxx xxx xxx xxx xxx doctor_daniel_jackson xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx. colonel_jack_o'neill, xxx xxx xxx xxx xxx, xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx.

58 Create Positive Thesaurus

A positive thesaurus translates every term into itself, such that only terms contained in the generalization thesaurus are being maintained. If the positive thesaurus is applied, the user needs to make a decision about adjacency as well.

59 Load Union Concept

AutoMap supports 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 *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 4. Generalization Thesaurus*.

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

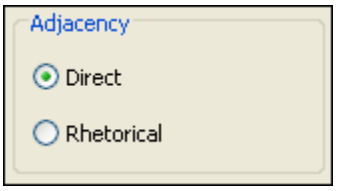
59.1 Creating and Using "Union of Concepts"

From the main menu select *File => Create or refresh the Union Concept List*. This will create a file in the directory specified in the Output Storage Manager.

In *Text Processing Panel => Tab 3. Pre-Processing Settings => SubTab 4. Generalization Thesaurus* select the [Load Union Concept List] button.

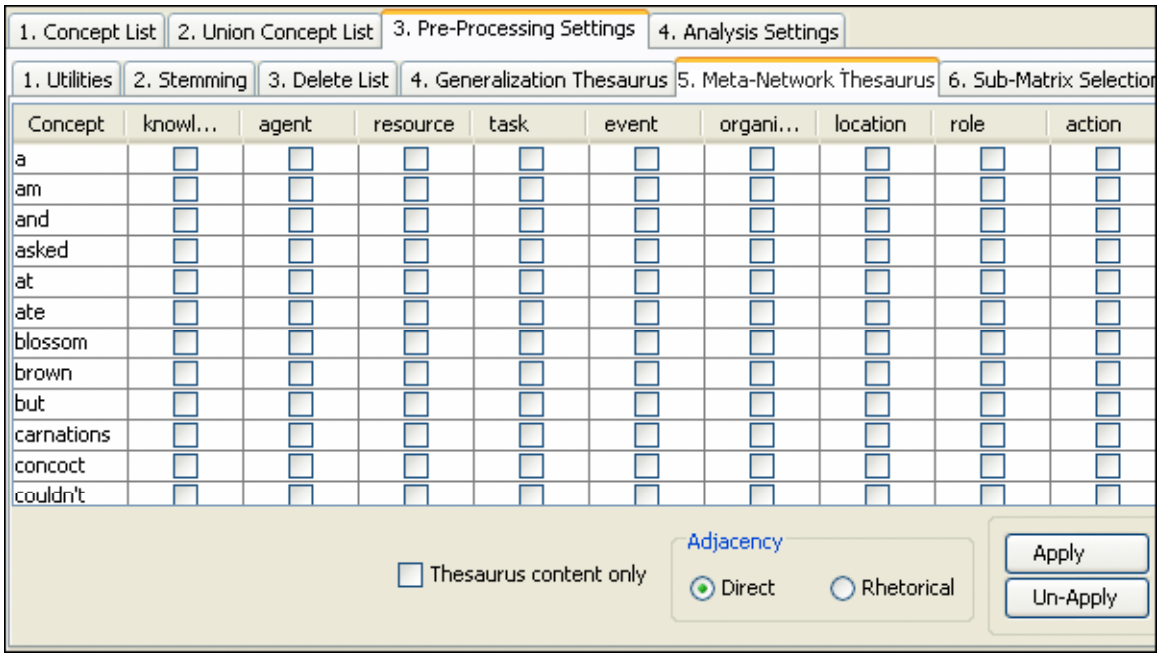
This concept list loaded into AutoMap can be refined by applying processing tools such as Named-Entity Recognition and Deletion prior to Generalization.

60 Adjacency Options



The Adjacency section contains two radio buttons. This section is only active if the Thesaurus Content only is chosen. The default is Direct Adjacency

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



A Meta Network Thesaurus associates text-level concepts with Meta Network categories. Since one concept might need to be translated into several Meta Network categories, a Meta Network 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.

62.1 Related type of Analysis: Meta Network Analysis.

For the Meta Network thesaurus, column headers start with concept knowledge. The order AND naming of column headers of the Meta Network thesaurus can be changed in the XML file.

63 Open a Meta Network Thesaurus

From the main menu select *File => Open Meta Network Thesaurus => 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 in the *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 4. Generalization Thesaurus*

63.1.1 Note : The Meta Network Thesaurus can be edited from *Panel 2*.

If you have a pre-defined Meta Network Thesaurus available that matches (some of) the concepts contained in the loaded Meta Network Thesaurus you can open this file. From the main menu select *File => Open Meta Network Thesaurus => Open from file*.
If a concept contained in the pre-defined file matches a concept in the currently opened Meta Network Thesaurus the Meta Network categories assigned to this concept in the pre-defined file will be automatically assigned to the concept in the currently opened Meta Network Thesaurus.

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

64 Edit Meta-Network Thesaurus

Each of the concepts can be assigned one or more Meta Network categories. This is done by placing a checkmark in the corresponding checkbox. To remove a category from a concept remove the checkmark in the category column.

64.1 User-defined Category

The far-right column is the User-defined category. Double-click in the cell and type in the Meta Network category for the concept. The number of new categories is not limited.

64.2 Multiple User-Defined Categories

Multiple User-defined categories can be assigned to the same concept by creating separate rows for the number of new concepts to be assigned. 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 Network categories.

65 Build a Meta-Network outside of AutoMap

You can build a Meta Network Thesaurus outside of AutoMap by using a text editor.
Format: (Key) Concept/Meta Network category/Meta Network category/....

65.1 Rules for Building a Thesaurus

A (key) concept can be assigned to as many Meta Network categories as possible. If a concept that appears in the User-Defined Meta Network Thesaurus does not appear in the Concepts column of the Meta Network Thesaurus this will not result in a conflict.

AutoMap automatically searches for concepts contained in the Concepts column of the Meta Network 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.

Avoid empty lines.

Make sure to separate the words by a slash.

The Meta Network Thesaurus is NOT case sensitive.

66 Save an applied Meta Network Thesaurus

66.1 Save the Meta Network

To save a Meta Network Thesaurus that has been applied select form the main menu *File => Save Meta Network Thesaurus*. The file will be saved to the location designated in the Output Storage Manager.

66.2 Save the Applied text

From the main menu select *File => Save Text(s) after Meta Network Thesaurus applied*. All texts are automatically saved, by default, in a folder called preprocessed in the root directory of AutoMap. The filename will be "after_MMCatThes_NameOfYourText.txt". The folder these files are saved to can be changed in the Output Storage Manager.

67 Example of a Meta Network Thesaurus

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

A Meta Network Thesaurus was opened selected from the main menu *File => Open Meta Network Thesaurus => Open from highest level of pre-processing*.

The union of concepts from the highest level of pre-processing will be displayed in alphabetical order in the left window. As we did not pre-process the text the original input sentence is used for input for the Meta Network Thesaurus.

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

Tok'ra, agent
attack, task
kill, task
System Lords, organization
Gou'alds, agent
infiltrating, task
summit, event
US military, organization
Earth, location
SGC, organization
Daniel, agent
Jacob Carter, agent
O'Neill, agent
SG-1, organization
Revenna, location

Now from the main menu select *File => Open Meta Network Thesaurus => Open from file*. Navigate to the directory where the file is saved and select [OK]. AutoMap will search the opened Meta Network Thesaurus for the words contained in the prepared Thesaurus. When it found a match it assigned the words in the opened Thesaurus the Meta Network categories that were assigned to the same concept in the pre-defined file.

This list can now be edited. Two entries will be added: poisoning, task and presents, task.

Next select the [Apply] button to apply the Meta Network Thesaurus with the following settings:

Thesaurus content only NOT selected (no checkmark in box is the default)

synopsis : xxx < agent organization> be plan xxx < task> which be < task action> xxx xxx xxx system lord < agent organization> xxx one fall swoop xxx infiltrating xxx < event> xxx poison xxx xxx simultaneously . xxx present xxx plan xxx xxx xxx < organization> xxx xxx < location> < location> . with xxx help xxx xxx < organization> xxx xxx < organization location> < organization> < agent> be able xxx < task> xxx < event> with < agent> < agent> < agent> xxx backup . < agent> , xxx xxx rest xxx < organization> , stay xxx < location> . xxx plan be go fine xxx xxx new < agent> present xxx new information xxx < location> be < task> . xxx plan be now consider xxx worst thing possible .

Thesaurus content only with Direct Adjacency (default if Thesaurus content only is chosen)

< agent organization> < task> < task action> < agent organization> < event> . < organization> < location> < location> . < organization> < organization location> < organization> < agent> < task> < event> < agent> < agent> < agent> . < agent> < organization> < location> . < agent> < location> < task> . .

Thesaurus content only with Rhetorical Adjacency

xxx xxx < agent organization> xxx xxx xxx < task> xxx xxx < task action> xxx xxx xxx xxx xxx < agent organization> xxx xxx xxx xxx xxx xxx xxx < event> xxx xxx xxx xxx xxx . xxx xxx xxx xxx xxx xxx xxx < organization> xxx xxx < location> < location> . xxx xxx xxx xxx xxx < organization> xxx xxx < organization location> < organization> < agent> xxx xxx xxx < task> xxx < event> xxx < agent> < agent> < agent> xxx xxx . < agent> xxx xxx xxx xxx < organization> xxx xxx < location> . xxx xxx xxx xxx xxx xxx xxx xxx < agent> xxx xxx xxx xxx xxx < location> xxx < task> . xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx .

68 Apply Meta-Network Thesaurus

Meta Network pre-processing is a higher level of pre-processing than the application of a Delete List and a Generalization Thesaurus. A Delete List and / or a Generalization Thesaurus needs to be applied before these other pre-processing tools.

68.1 Procedure:

Select whether to use the Thesaurus content only option or not.

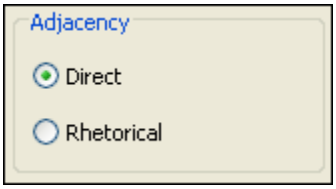
If Thesaurus content only option is chose either Adjacency options. Adjacency can be either direct (default) or rhetorical.

To activate the Thesaurus select the [Apply Thesaurus] button. 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 Network category.

68.1.1 NOTE : The Thesaurus is case insensitive.

If multiple Meta Network 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>).

69 Adjacency Options



The Adjacency section contains two radio buttons. This section is only active if the Thesaurus Content only is chosen. The default is Direct Adjacency

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

70 Thesaurus Content Only

70.1 Thesaurus Content only - chosen

AutoMap will search the text(s) for concepts specified in the thesaurus. It will then translate matches into key concepts.

It will 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.

As a result, all key concepts in the resulting text appear directly adjacent to each other.

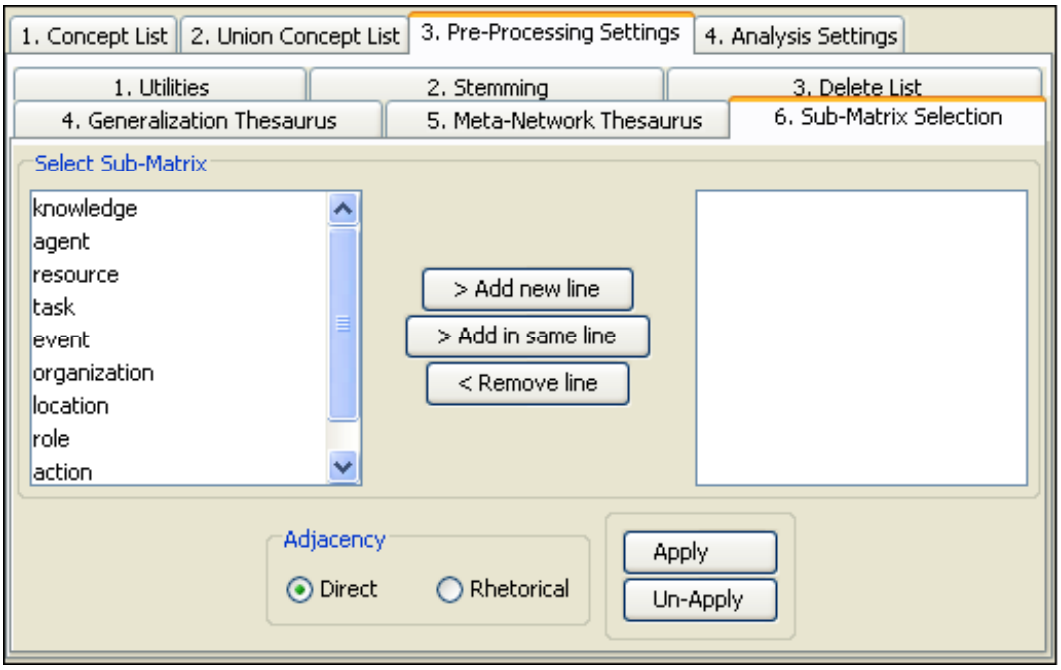
To select the Thesaurus content only option go to *Text Processing Panel => Tab 6. Generalization Thesaurus*. Select the radio button for Thesaurus Content Only. You can now choose to either use direct or rhetorical adjacency for the application of the Generalization Thesaurus. Then select the Apply button. To switch from not using the Thesaurus content only option uncheck the Thesaurus content only item on *Text Processing Panel => Tab 6. Generalization Thesaurus* and apply the Generalization Thesaurus again.

70.2 Thesaurus Content only - NOT chosen

The text will remain as is. 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.

If not checked AutoMap does not apply the Thesaurus content only option. To switch to using the Thesaurus content only option check the Thesaurus content only item then reapply the Generalization Thesaurus again.

71 Sub-Matrix Selection



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

71.1.1 NOTE : Sub-Matrix Selection can only be performed if Meta Network Thesaurus was applied.

71.2 Select Sub-Matrix

Left Panel : Lists the concepts that are available.

Middle Buttons : Allow you to add or remove concepts to a Thesaurus.

[> Add new line] : Creates a new line and moves selected concept to right window.

[> Add in same line] : Attaches the concept selected in the left window to the currently selected line in the right window.

[< Remove line] : Removes the selected line in the right window.

Right Panel : Displays the your choices of concepts to apply.

71.3 Adjacency

You can also apply Adjacency in the forms of Direct and Rhetorical

71.4 Apply / Un-Apply buttons

As the name implies Apply of Remove the selected Sub-Matrix.

72 Applying a Sub-Matrix Selection

72.1 Apply a Sub-Matrix

1. Select an Adjacency option. Adjacency can be either direct (default) or rhetorical. Direct adjacency is the default.
2. Apply the options by selecting the Apply Sub-Matrix Selection button. AutoMap will then search the text(s) for Meta Network tags. If a match is found it will be translated into the related concept.

   The Thesaurus IS case insensitive.

3. The text can be viewed in the *Text Display Panel => Tab 1. Original texts*

72.2 Un-Apply a Sub-Matrix

To un-apply a Sub-Matrix Selection that was applied to the data, go to the *Text Processing Panel => Tab 3. Pre-processing Settings => Sub-Tab 6. Sub-Matrix Selection* and select the [Un-Apply] button. The text in the *Text Display Panel => Tab 7. Texts after Sub-Matrix Selection* will be cleared.

72.3 Direct Adjacency

Direct adjacency means that original distances of concepts that represent Meta Network categories will neither be visualized nor considered for analysis.

If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis.

As a result, only concepts that represent Meta Network categories are displayed on the *Text Display Panel => Tab 7. Texts after Sub-Matrix Selection* and will be considered for analysis. All concepts in the resulting text appear directly adjacent to each other.

72.4 Rhetorical Adjacency

As a result, concepts that represent Meta Network categories are displayed on the *Text Display Panel => Tab 7. Texts after Sub-Matrix Selection* and will be considered for analysis. Original distances of concepts that represent Meta Network categories will be visually symbolized by placeholders (xxx) and considered for analysis.

73 Example of a Sub-Matrix

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

Open a Sub-Matrix Selection.

Apply the Sub-Matrix Selection with the following settings:

73.1 Select:

Agent	/	Location	/	Action
Location	/	Agent	/	Action
Action	/	Location	/	Agent

Direct Adjacency (default)

Output:

. dog.. dog. dog.. dog.... dog,.. .,... prof_john_darren.. dog.....,,,.

73.2 Select:

Agent / Location / Action
Location / Agent / Action
Action / Location / Agent
Rhetorical Adjacency

Output:

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. xxx, xxx. xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx. prof_john_darren xxx xxx. xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx dog. xxx. xxx xxx xxx xxx. xxx xxx. xxx xxx. xxx xxx xxx, xxx, xxx, xxx xxx.

73.3 Select:

Full Meta Network
Direct Adjacency (default)

Output:

. dog eating lotus.. dog. yelling dog eating flowers weeds.. dog... roses. dog roses,.. .,.. met. prof_john_darren.. dog..... lotus, carnations, daffodils, roses.

73.4 Select:

Full Meta Network Rhetorical Adjacency (default)

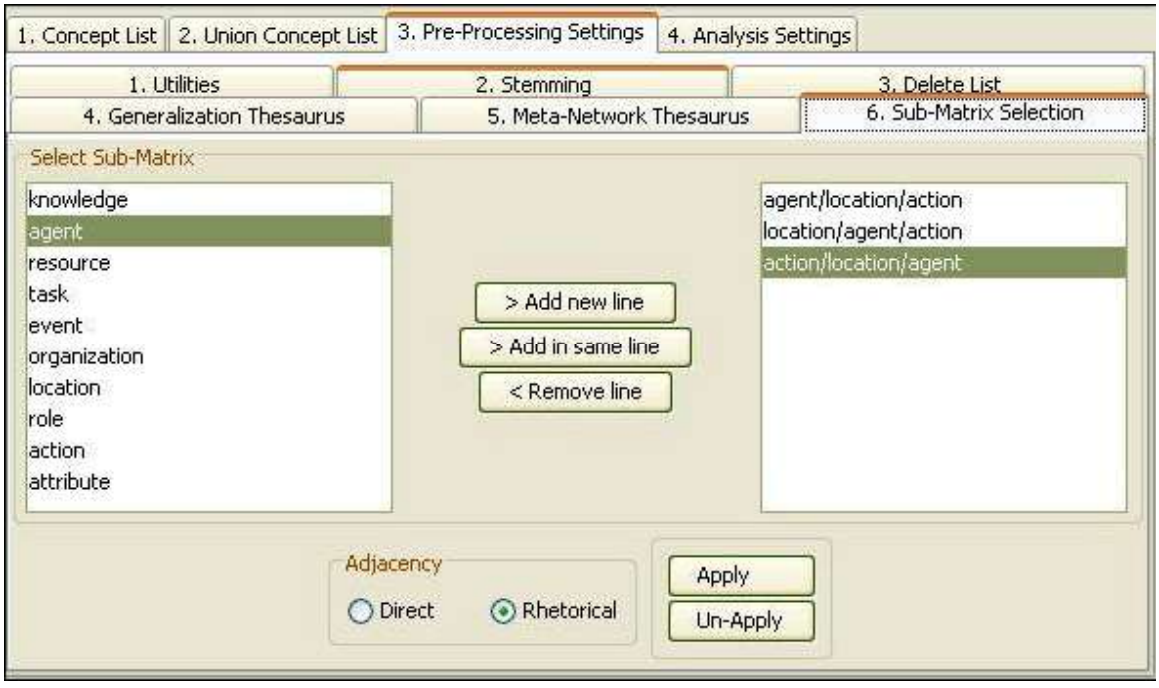
Output:

xxx. xxx xxx dog eating xxx lotus xxx xxx xxx xxx. xxx. xxx xxx xxx xxx xxx dog. xxx yelling xxx xxx xxx xxx xxx xxx xxx xxx dog xxx eating xxx flowers xxx weeds. xxx xxx xxx. xxx xxx xxx xxx dog. xxx xxx. xxx. xxx xxx roses xxx xxx xxx xxx. xxx xxx dog xxx xxx xxx xxx roses xxx xxx xxx xxx xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx

xxx

xxx. xxx, xxx. xxx xxx xxx. xxx met xxx xxx xxx xxx xxx xxx xxx. prof_john_darren xxx xxx. xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx dog. xxx. xxx xxx xxx xxx. xxx xxx. xxx xxx. xxx xxx lotus, carnations, daffodils, xxx roses.

Detailed example for Sub-Matrix Selection:



74 Saving a Sub-Matrix

From the main menu select *File => Save Text(s) after Sub-Matrix Selection*. All texts are automatically saved in a folder called preprocessed in the root directory of AutoMap. The filename will be after_SubMatrixSelection_NameOfYourText.txt".

75 Selecting Sub-Matrix Categories

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

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

Create and modify a sub matrix selection.
Go to *Text Processing Panel => Tab 3. Pre-Processing settings => Sub-Tab 6. Sub-Matrix Selection*. You have three options:

1. You can select as many Sub-Matrix categories per row as you wish by clicking on a category in the left window and selecting the [> Add in same line] button to move it to the right window.
2. You can select as many rows of sub matrix selections as you wish by selecting a category in the left window on and move it to the right window by selecting the [> Add in new line] button.
3. To unselect a selected Sub-Matrix category, click on the row in the right window and hit the [< Remove line] button. The row will disappear from the right window.

75.1.1 NOTE : If you wish to analyze any relation between 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].

75.1.2 NOTE : 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].

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

From the main menu select *File => Open Sub-Matrix Selection => Select Full Meta Network*. AutoMap dynamically generates all combinations of Meta Network categories as specified in the Meta Network thesaurus, including user-defined categories, that represent all cells of the Meta Network 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 Network:

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

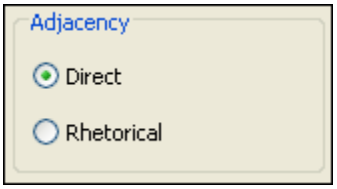
75.3 If you have a sub matrix selection file available

From the main menu and select *File => Open Sub-Matrix Selection => Open from file*.
You can modify this sub matrix selections.

75.4 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.
From the main menu select *File menu => Open Sub-Matrix Selection => Open from file*. You can modify this sub matrix selections.

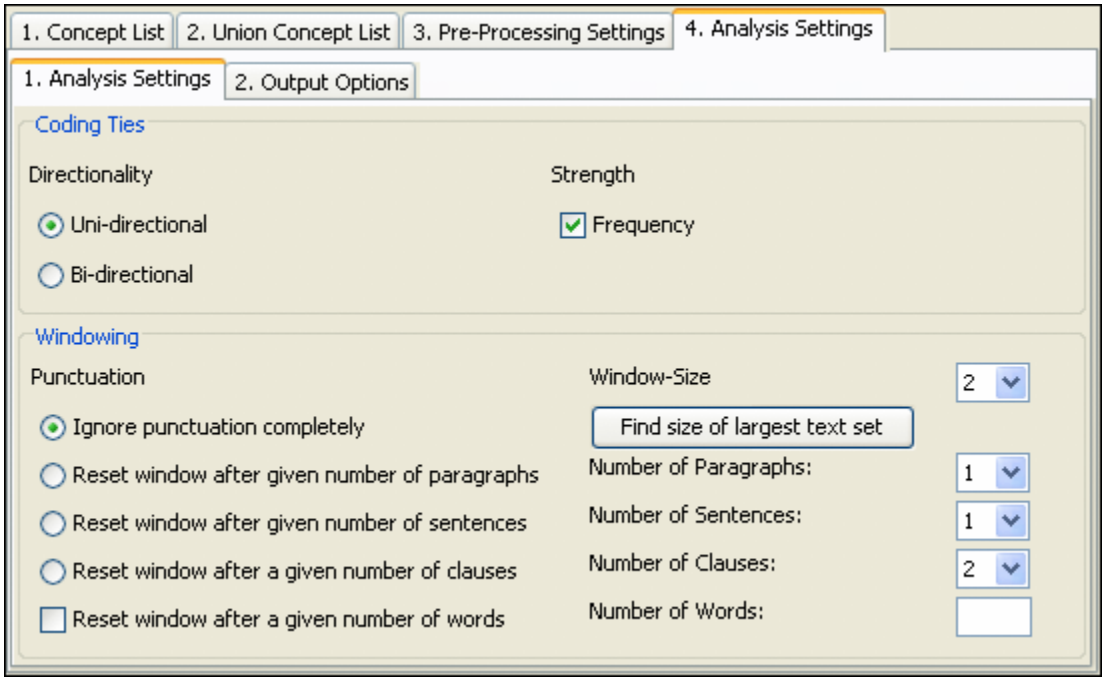
76 Adjacency Options



The Adjacency section contains two radio buttons. This section is only active if the Thesaurus Content only is chosen. The default is Direct Adjacency

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

77 Text Processing Panel - Analysis Setting



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.

77.1 Analysis Settings

Coding Ties : Specify the way statements are counted.

Directionality : Select one of the following two possibilities by checking the button.

Uni-Directional : (When coding a link, only 1st⇒2nd concept should be noted)

Bi-Directional : (When coding a link, both 1st⇔2nd concept shall be noted)

Strength : Strength will be printed out by default. To not print out Strength uncheck the item.

Frequency : If checked counts the cumulative frequency of every existing statement. If not checked the existence of frequency will be printed out (binary result).

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

Punctuation : Select one option by checking the radio button.

The radio buttons [Ignore punctuation completely and the Reset window after given number of / paragraphs / sentences / clauses] : allow for selecting the size of the text for analysis.

Save your analysis in DyNetML format: Reset window after given number of words checkbox can be used with any of the above options.

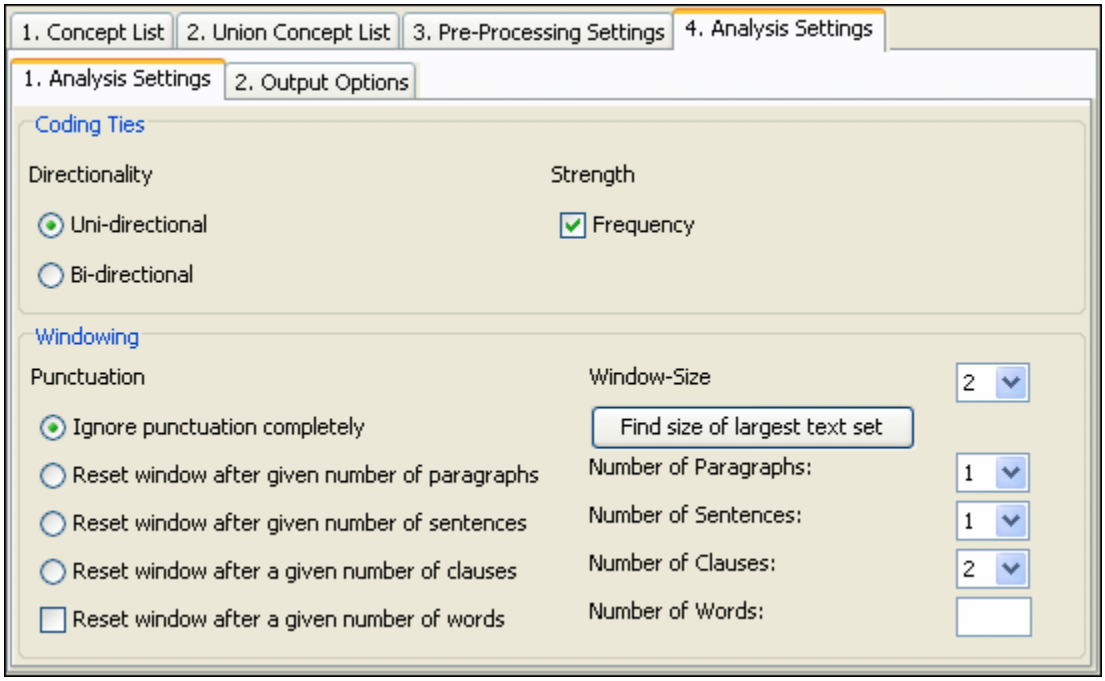
Window Size : Select one window size by using the number chooser.

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.

78 Text Processing Panel - Analysis

Analysis Setting Tab.

79 Text Processing Panel - Analysis Setting



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.

79.1 Coding Ties

Specify the way statements are counted.

Directionality : Select one of the following two possibilities by checking the button.

Uni-Directional : (When coding a link, *only 1st⇒2nd concept* should be noted)

Bi-Directional : (When coding a link, *both 1st⇔2nd concept* shall be noted)

Strength : Strength will be printed out by default. To not print out Strength uncheck the item.

Frequency : If checked counts the cumulative frequency of every existing statement. If not checked the existence of frequency will be printed out (binary result).

79.2 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.

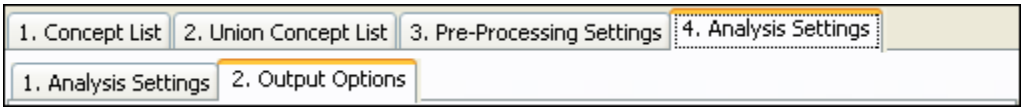
Reset window at end of paragraphs only : Statements will be placed only within every single paragraph.

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.

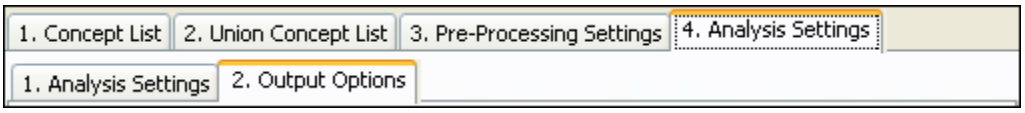
Window size between 2 and 100. : The Window Size defines how distant concepts can be and still have a relationship. Only concepts in same window can form statements.

80 Output Options



The map and statistic output generated by Automap are displayed in the *Semantic Network Panel and the Results Panel*, respectively. In addition, Automap offers further output options that are available in *Text Processing Panel => Tab 4. Analysis Settings => SubTab 2. Output Options*. All additional outputs are only generated after the analysis are run.

81 Term Distribution List



The map and the statistic output generated by AutoMap are displayed on the *Semantic Network Panel and the Results Panel*, respectively. In addition to that AutoMap offers further output options that can be chosen on *Text Processing Panel => Tab 4. Analysis Settings => Sub-Tab 2. Output Options* Scroll down Term Distribution Lists. All additional outputs are only generated after analyses were run.

81.1 Points To Consider

- No list will be generated by default.
- All lists 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 will be automatically saved in a folder called Term Distribution Lists under the root directory of AutoMap.
- The Term Distribution Lists folder is overwritten with every new analysis you run. So if you want to save the results of a current Term Distribution Lists folder just rename the folder. For each Term Distribution List checked two output lists are generated.

81.2 Types and content of Term Distribution Lists
81.3 Concepts analyzed

List of concept analyzed.csv - Concept, Text, Frequency

Statistics of concepts analyzed.csv - Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts

81.4 Concepts in Statements and Isolates

List of concept in statements.csv - Concept, Text, Frequency

Statistics of concept in statements.csv - Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts

List of isolates in statements.csv - Concept, Text, Frequency

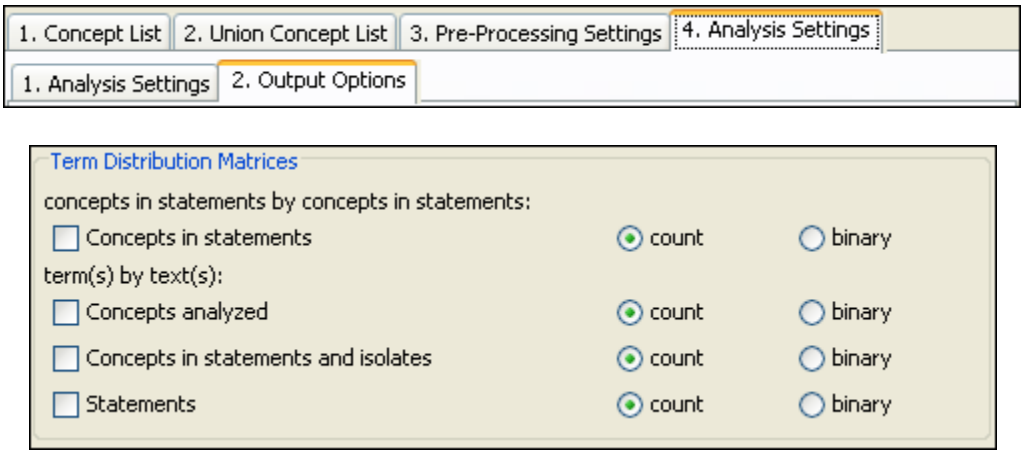
Statistics of isolates.csv - Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts

81.5 Statements

List of statements.csv - Statement, Text, Frequency

Statistics of statements.csv - Statement, Cumulated sum across text set, Number of text statement occurs in, Percentage of texts statement occurs in, Texts

82 Term Distribution Matrices



82.1 Concepts in statements by concepts in statements:

Concepts in statements: Matrix of concepts in statements.csv

Matrix of Concept that were linked into statements (first row) by Concept that were linked into statements (first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept

82.2 Term(s) by text(s):

Concepts analyzed: Matrix of concept analyzed.csv

Matrix of Concept (union of concepts listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept

Concepts in statements and isolates: Matrix of concept in statements.csv

Matrix of Concept (union of concepts listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept

Statements: Matrix of statements.csv

Matrix of Statements (union of statements listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of statement If binary was chosen, cells denote existence (1) or absence (0) of statement

83 Term Distribution List Example

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

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

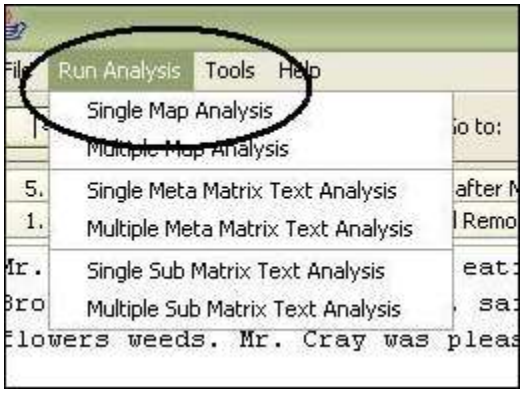
Our Text II.txt: Mr. Cray's brown dog stopped eating the lotus blossom at 12 pm the next day. Mrs. Brown was now happy with the dog. She said "You good dog!" The dog no longer ate the flowers and weeds. Mr. Cray was pleased too. Mrs. Brown watered the roses and fertilized the garden on June 13, 1880.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

Then AutoMap's Extensive Delete List, Direct Adjacency was applied to both texts. The Delete List was extended by further non-content bearing words that appeared in the sample texts (a an and as at awhile but for from happening he her 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.

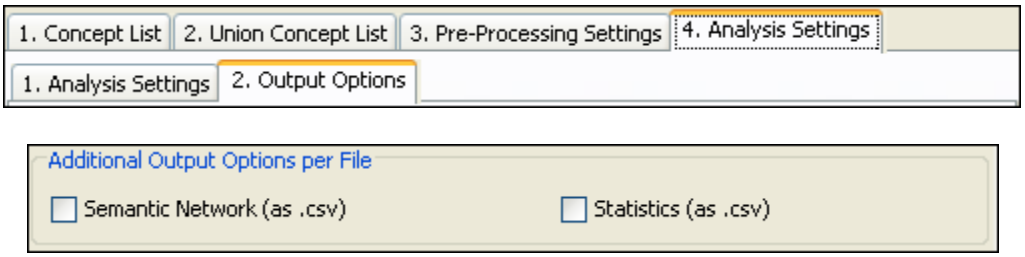
Our Text II.txt: Mr. Cray's brown dog stopped eating lotus blossom 12 pm next day. Mrs. Brown was now happy with dog. said "You good dog!" dog no longer ate flowers weeds. Mr. Cray was pleased too. Mrs. Brown watered roses fertilized garden on June 13, 1880. Prof. Darren, Mrs. Brown & Mr. Cray met over 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):



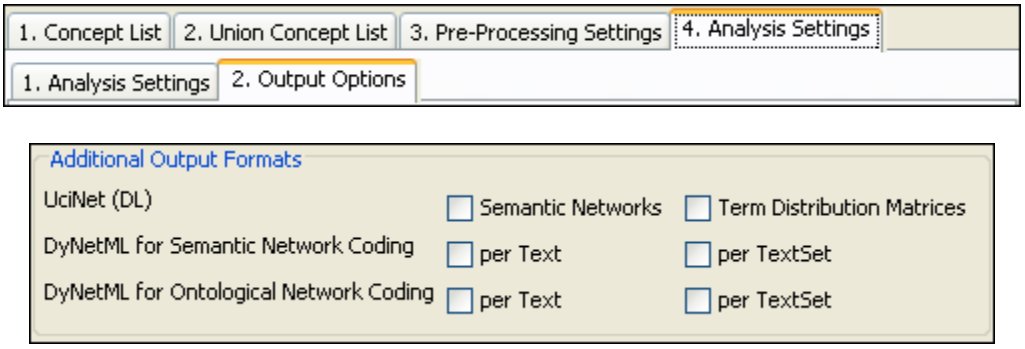
All outputs provided on the Output Options panel were checked. Again, the default settings were used. All output lists are saved in this folder (Term_Distribution_Lists_and_Matrices.zip).

84 Additional Output Options



The these two options controls whether AutoMap writes these two files as output.

85 Additional Output Formats



Three additional data formats are offered:

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

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

85.2 UciNet (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
- 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.

85.2.1 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.

85.3 DyNetML

Results of Map Analysis and Sub-Matrix Text Analysis can be output in DyNetML format. The resulting DyNetML files will be stored as NameOfText.xml in the xml folder under the root directory of AutoMap.

85.4 After Map Analysis:

See Concepts not forming edges

85.5 Run Sub-Matrix Text Analysis.

Two options for creating DyNetML files are offered. Either one or both options can be selected per analysis. These are found on *Text Processing Panel => Tab 4. Analysis Settings => Sub-Tab 2. Output Options* under the heading Concepts not forming edges.

1. per Text: The DyNetML files will be stored as NameOfText.xml and in a folder called xml under the root directory of AutoMap.
2. per TextSet: The DyNetML file will be stored as consolidated_map.xml in a folder called xml under the root directory of AutoMap.

85.6 Examples for additional Output Formats

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

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

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

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

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

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

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

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

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

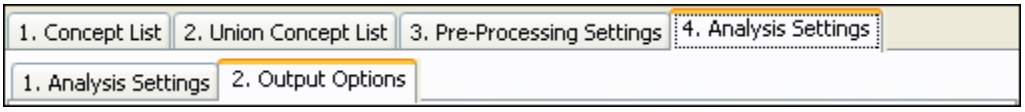
Prof. Darren, Mrs. Brown & Mr. Cray met over dinner discussed how plan had worked. John Darren Mrs. Brown take down scarecrow following week. thought too scary dog. Mr. Craye painted fence. Then Mrs. Brown watered lotus, carnations, daffodils, roses.

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

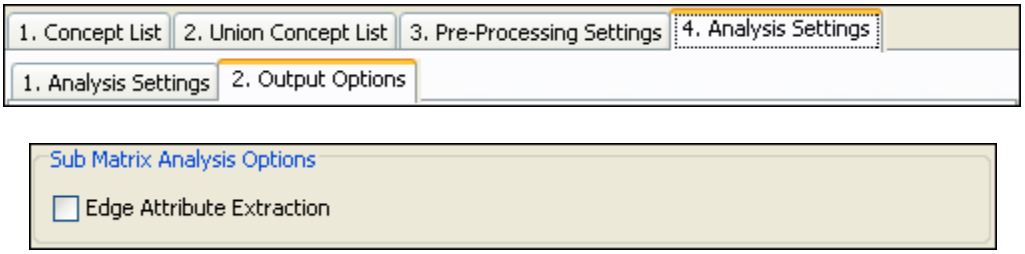
Semantic network of current Text

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

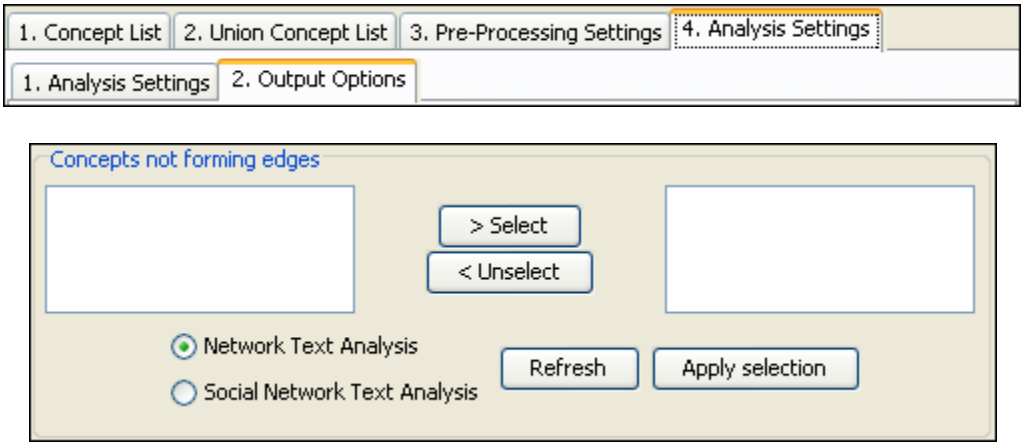
86 Parts of Speech Tagging Options



87 Sub Matrix Analysis Options



88 Concepts not Forming Edges



By default, all entities in the applied ontology are considered as nodes, and all statements between nodes in the applied ontology (either in anterior or posterior or both positions) are represented as links 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:

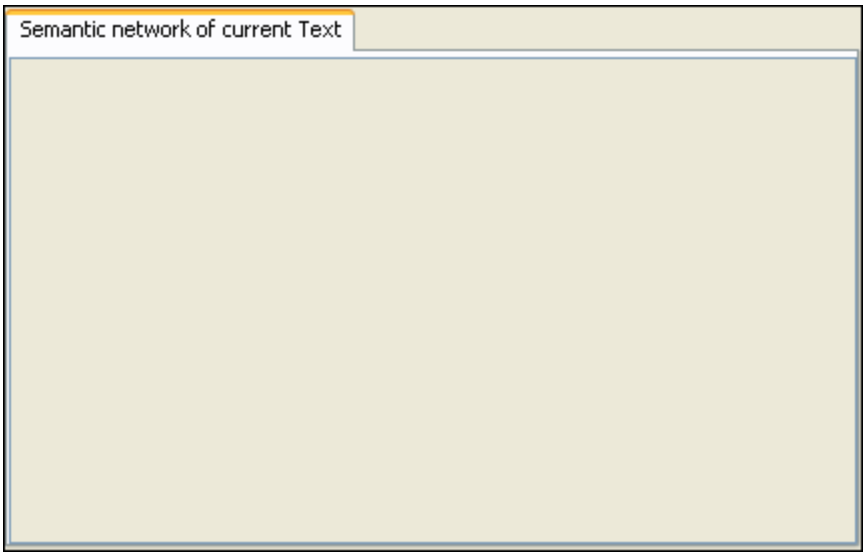
1. On *Text Processing Panel => Tab 4. Analysis Settings => Sub-Tab 2. Output Options* in the section Concepts not forming edges the Refresh button to automatically generate a list of all entities in the applied ontology.
2. Use the [>] and [<] buttons to move entities from one window to the other.
3. 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 (the default) (NTA) or Social Network Text Analysis (SNTA). To select a type use the radio buttons. Both types are a form of Sub-Matrix Text Analysis, but differ in what they measure:
    1. Network Text Analysis: If one measures a 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. Social Network Text Analysis: 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.

88.1.1 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.

89 Semantic Networks of Current Text Panel



After running an analysis, the semantic network will be displayed in the *Semantic Network Panel*. The semantic network contains one coded statement per line.

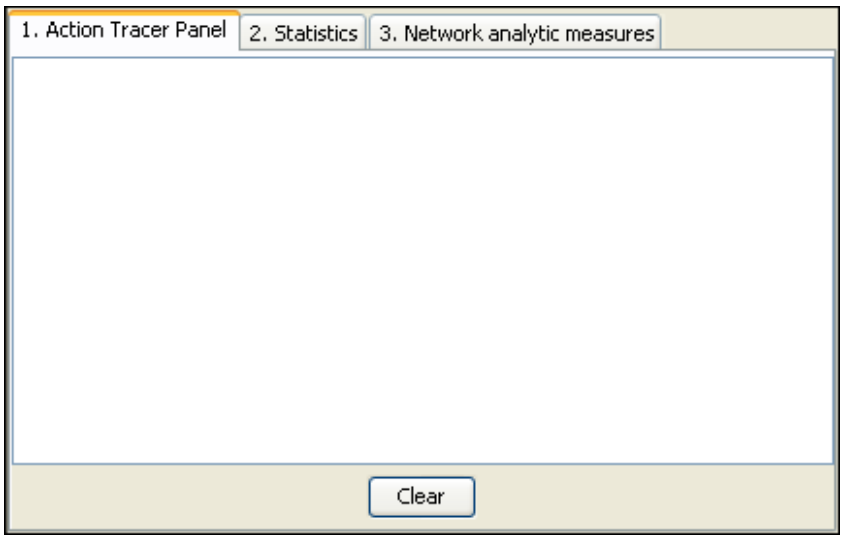
If the Frequency item was checked in the Analysis Settings the first column of the semantic network indicates the frequency of every displayed statement.

Each semantic network generated is automatically saved in a folder named “Stat Output” in the root directory of AutoMap. 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.

90 Results Panel

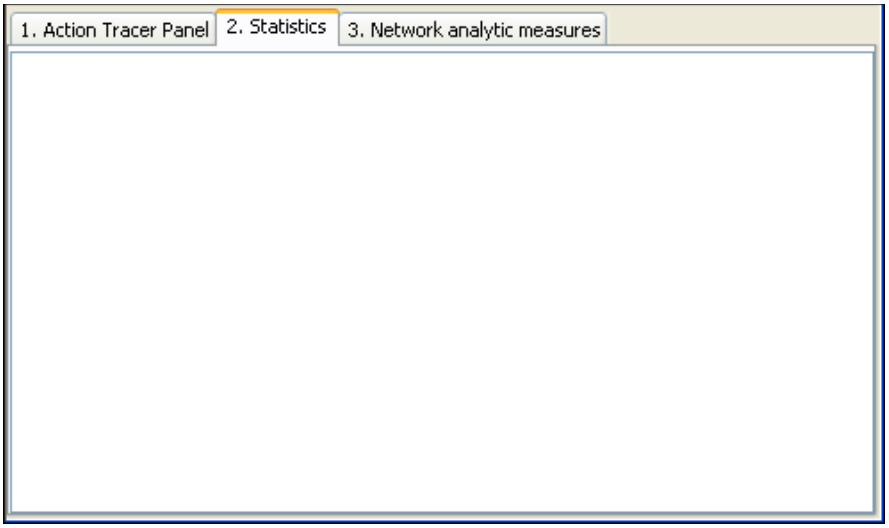
Panel 4 keeps a running record of your actions. This includes the *Action Tracer tab*, the *Statistics tab*, and the *Analytic tab.*

91 Results Panel — Action Tracer



The *Results Panel => Tab 1. Action Tracer* records the actions taken on the text files you work on.

92 Results Panel ⇒ Statistics



After running an analysis of a file, the Statistics will be displayed in the *Results Panel => Tab 2. Statistics*.

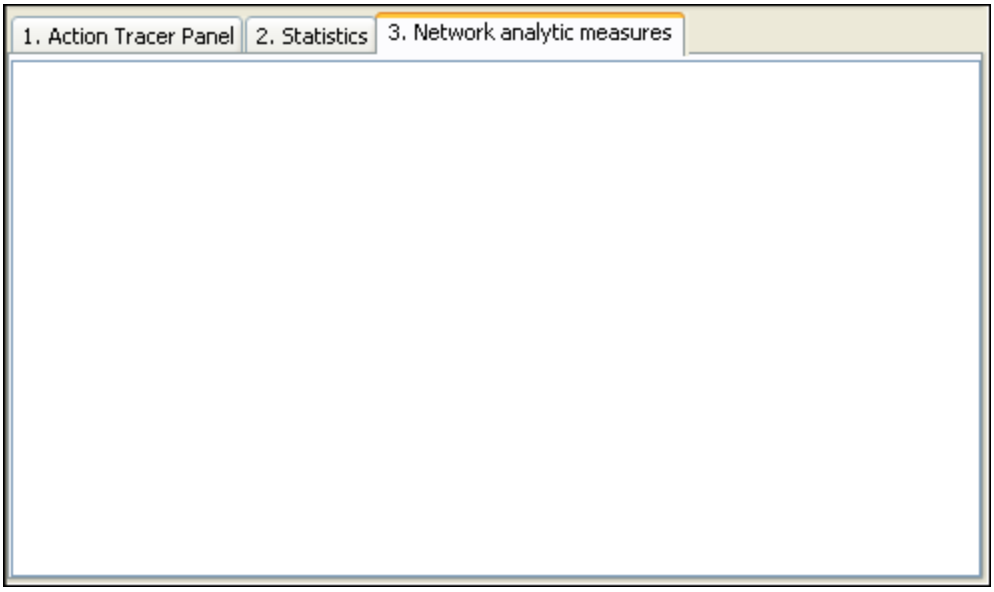
Each stat file generated is automatically saved in a folder named Stat Output in the AutoMap main directory. This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as (stat_output.xls) file a that contains the stats of all texts.

Entries in the stat output and explanation:

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

93

94 Network Analytic Measures



AutoMap supports the computation of Network Analytic Measures per map or network and per concept or node.

94.1.1 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: (steps need redone)

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

The following Table explains the measures that can be computed:

Name of measure	Calculation of measure	Name of and reference for corresponding SNA measure	Corresponding name of measure in Cube
Concept (node) level measures, direct connectivity			
Concept outdegree	Total number of statements with concept in anterior position	Outdegree, expansiveness, actor degree centrality (Wasserman & Faust 1994); Prestige, Influence (Mrvar)	Local imageability
Concept indegree	Total number of statements with concept in posterior position	Indegree, receptivity, popularity, actor degree prestige (Wasserman &	Local evokability

		Faust 1994); Prestige, Support (Mrvar)	
Concept outdegree centrality	Total number of statements with concept in anterior position/ Number of unique concepts in text	Outdegree Centrality (Wasserman & Faust 1994); Relative Influence (Mrvar)	None
Concept indegree centrality	Total number of statements with concept in posterior position/ Number of unique concepts per text	Indegree Centrality (Wasserman & Faust 1994); Relative Support (Mrvar)	None
Total degree	Concept indegree + concept outdegree	-	Local density
	Map (graph) level measures, direct connectivity		
Text outdegree centrality	Sum (largest observed outdegree - outdegree of concepts)/(Number of unique concepts in text) 2	Group outdegree centralization (Wasserman & Faust, 1994)	None
Mean concept outdegree centrality	Sum (outdegree)/ Number of unique concepts in text	Mean outdegree (= Mean indegree) (Wasserman & Faust 1994)	None
Variance of concept outdegree centrality	Sum(sum outdegree - mean outdegree) 2 / Number of unique concepts in text	Variance of outdegree (Wasserman & Faust 1994, p.127-128)	None
Text indegree centrality	Sum (largest observed indegree - indegree of concepts)/(Number of unique concepts in text) 2	Group indegree centralization (Wasserman & Faust, 1994)	None
Mean concept indegree centrality	Sum (indegree)/ Number of unique concepts in text	Mean indegree (=Mean outdegree)	None
Variance of concept indegree centrality	Sum (sum indegree - mean indegree) 2 / Number of unique concepts in text	Variance of indegree (Wasserman & Faust 1994, p.127-128)	None
Density	Number of statements/ Possible number of statements	Density (Wasserman & Faust 1994, p.129, Scott 1991, p.74) Wasserman and Faust use g(g-1) as denominator:, we use (g*g), because unique concept can form statement	None

		with same unique concept (e.g. agent-agent)	
Concept (node) level measures, indirect connectivity			
Concept closeness centrality	Minimum possible total distance from node i to all other nodes/ Sum of all geodesics between node i and all other nodes	Closeness (Wasserman & Faust, Mrvar) According to Wasserman and Faust (1994, p.200) group level closeness centrality is not computed	None
Concept betweenness centrality	sum ((Number of all geodesics between all nodes that go through node i)/(Number of geodesics between node i and all other nodes))/((Number of unique concepts in text -1)(Number of unique concepts in text -2))	Betweenness (Gould 1987, Mrvar)	None
Concept proximity prestige	Number of concepts directly or indirectly adjacent to node i	Proximity Prestige (Wasserman & Faust 1994, Mrvar)	None
Map (graph) level measures, indirect connectivity			
Text Proximity Prestige	Sum (Proximity Prestige (all unique concepts in text)/ Number of unique concepts in text	Group level proximity prestige (Wasserman & Faust 1994)	None

95 Open Input

95.1 Open a single text

From the main menu select *File => Open => Open single file*. From the Open dialog box navigate to the directory and highlight the file. Then select [OK]. The text will appear in *Text Display Panel => Tab 1. Original Text*.

95.2 Open a Multiple Files

To open multiple files place them all in the same directory with no other files. Then from the main menu select *File => Open multiple files*. The Open dialog will appear. Navigate to the directory where the files are stored. The folder will only appear empty. The files will all be loaded into Automap as shown in the menu bar designating which file is currently active.

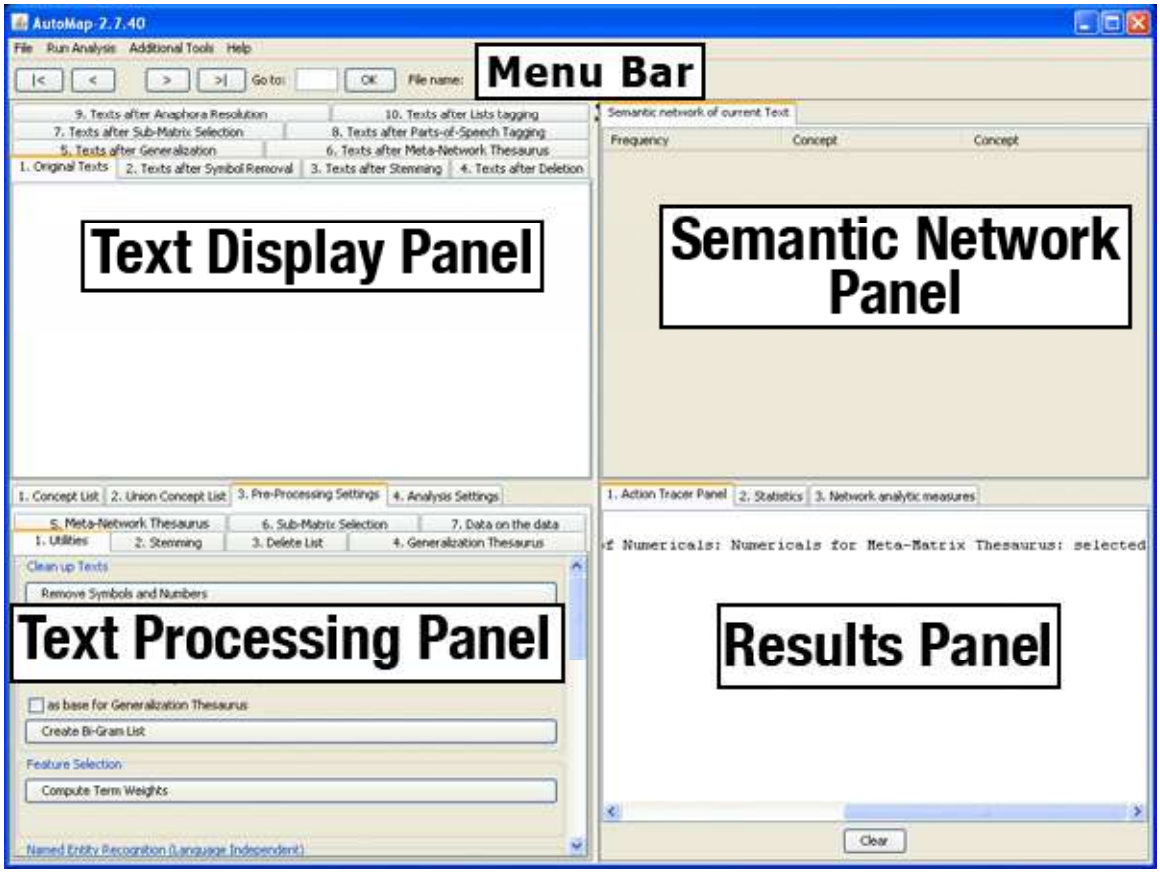
96 GUI Overview

AutoMap's graphical user interface (GUI) is divided into four primary panels. In the images below you can see the references: *Text Display Panel* (top left), *Text Processing Panel* (bottom left), *Semantic Network Panel* (top right) and *Results Panel* (bottom right).

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

The panels contain tabs which provide an 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:



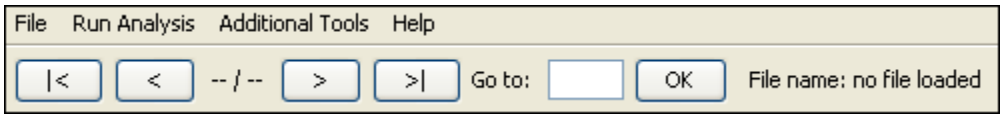
96.1.1 NOTE : The Text Display Panel is the only panel that *can* be edited. The other panels (Text Display Panel, Semantic Network Panel, & Results Panel) display information regarding the currently active file. All four panels always relates to each other. The Text Browse Menu relates to all panels 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 the *Results Panel* will log preprocessing utilities applied to your text. This is a handy way to keep track of changes and actions relating to your text.

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.

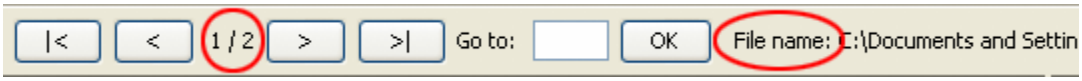
97 AutoMap Menus - The MenuBar



[| <] Go to first text file.
[<] Go to previous text file.
--/-- denotes # of file currently active / total # of files.
[>] Go to next text file
[> |] Go to last text file
Go to: Input the number of the text file you want to bring up.
[OK]
File name: xxx : Display the currently active text file. When AutoMap is first started this displays no file loaded

97.1 Browsing through texts

This function enables you to quickly jump from text example to another. All panels are synchronized in the Browse Menu.
Using the [>] and [<] buttons allow you move back and forth between the currently opened text files. The [|<] and [>|] will send you to the first or last file opened. The numbers in the middle 1 / 2 represent which number file you are viewing / total number of files open.



To go to a specific text enter the text number in text field right next to the Go to command and select [OK].
The name of the currently selected text is displayed to the right of File name:.

98 Concept List

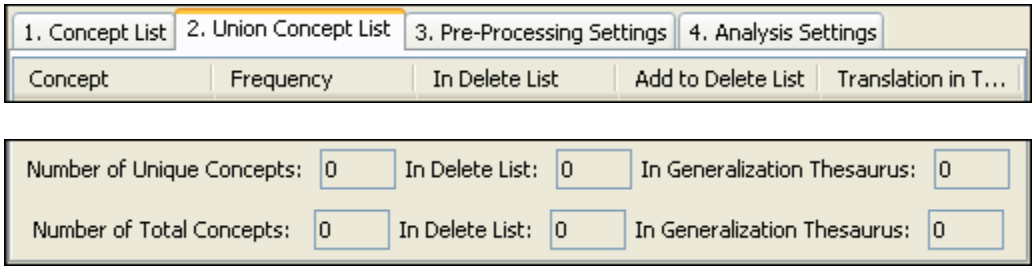
After loading text a Concept List is automatically generated and can be viewed in *Text Processing Panel => Tab 1. Concept List*. The list defaults to the order of Frequency and Concept. You can change the order by clicking the either of those headers. The concept list tells you several important facts about your text:

- The number of concepts found in the text displayed in *Text Display Panel => Tab 1. Original Text*.
- Related frequencies of those concepts.
- Cumulated unique concepts and total concepts contained in the data set.

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.

99 Union Concept List

Union Concept List results are displayed in *Text Processing Panel => Tab 2. Union Concept List*.



99.1 Create and refresh Union Concept List

The Union Concept List differs from the Concept List in that it considers concepts across all the texts currently loaded in Automap, rather than only the currently selected 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.

The union concept list must be refreshed by selected from the main menu *File => Refresh Union Concept List* before viewing your results. 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. Below is an example of the text and union concept list.

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.

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

99.2 Example for Union Concept List

File 1

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

File 2

On March 22, 2002 the Ren'al of the Tok'ra arrived @ the SGC to discuss their plan. 5 days later, on March 27, 2002, SG-1 and SG-17 arrive @ Revenna. The summit meeting with the 6 main System Lords* will be happening in 10 days. Daniel is faced with the question, "Will this prevent the System Lords from overthrowing the galaxy?"

*Gou'alds

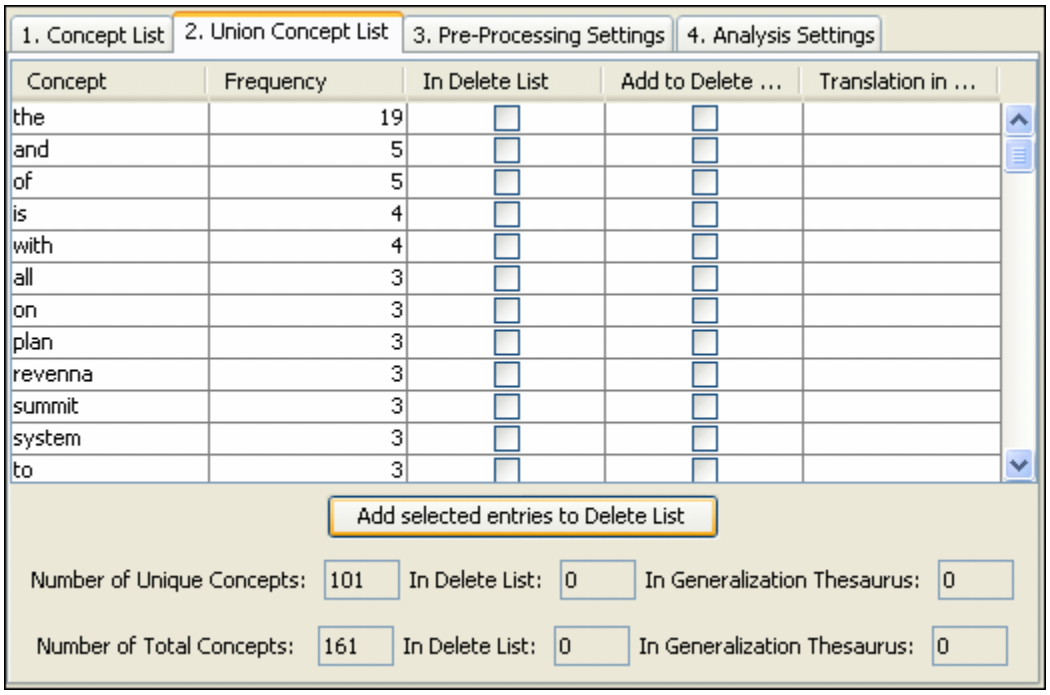
99.3 Creating a Union Concept List

The first step is to load in your text files. This can be done individually or using the *Open Multiple Files* function. In order to use the *Open Multiple Files* move all your files into a directory which contains no other text files.

From the main menu select *File => Open Multiple Files* and navigate to the folder in which you placed your text files. Then select [Open].

To run the Create and Refresh Union Concept Utility based on multiple texts select *File => Create and Refresh Union Concept List*

The result is shown in the screen shot below:

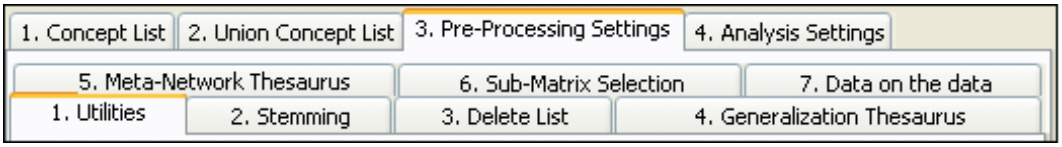


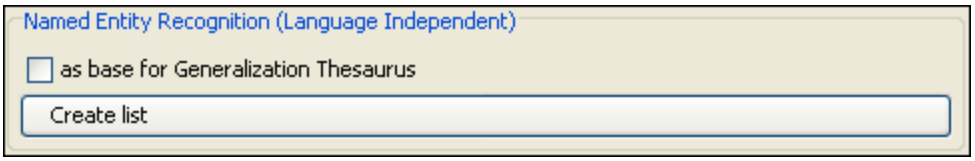
The table is ordered by Frequency but can be changed by clicking in any of the headers. The concept list contains more entries than the interface can display. The displayed Union Concept List indicates there are 101 unique concepts and 161 total concepts in the data set.

99.4 Save Union Concept List

From the main menu select *File => Save Union Concept List*. A file chooser will pop up. Navigate to a directory and save your file.

100Named Entity Recognition





Named-Entity Recognition 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.

100.1 Text Example

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

In *Text Processing Panel => Tab 3. Pre-Processing settings => SubTab 1. Utilities* under Named Entity Recognition select the Create List button.
This will create a Named Entity list which can be found in the directory specified in the Output Storage Manager.

100.2 NamedEntities.csv

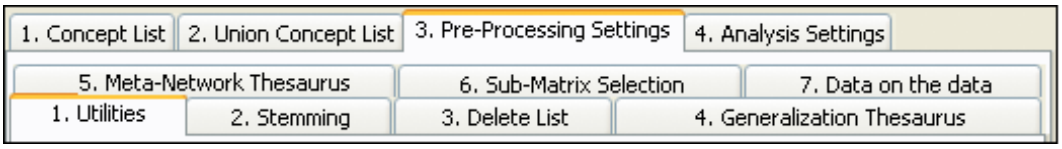
A .csv file is created. The csv file can be edited in a text editor.
Jacob CarterSelmak
US
SG
System Lord Goualds
Earth. With
Synopsis The Tokra
ONeill
All
United States
Revenna. The
Jacob CarterSelmak
System Lord
They
United States Military Daniel
Revenna
SGC

100.3 Used as a base for Generalization Thesaurus

If a checkmark is placed in the box before the button is selected the NamedEntities.csv file takes on a different format as used in the Generalization Thesaurus.
Jacob CarterSelmak,,Jacob CarterSelmak/ Jacob CarterSelmak
US,,US/ US
SG,,SG/ SG
System Lord Goualds,,System Lord Goualds/ System Lord Goualds
Earth. With,,Earth. With/ Earth. With
Synopsis The Tokra,,Synopsis The Tokra/ Synopsis The Tokra
ONeill,,ONeill/ ONeill
All,,All/ All
United States ,,United States / United States
Revenna. The,,Revenna. The/ Revenna. The
Jacob CarterSelmak ,,Jacob CarterSelmak / Jacob CarterSelmak
System Lord ,,System Lord / System Lord
They,,They/ They

United States Military Daniel,,United States Military Daniel/ United States Military Daniel
Revenna,,Revenna/ Revenna
SGC,,SGC/ SGC

101 Text Processing Panel - Bi-Gram Detection



101.1 N-gram Identification: Bigrams

Creates a list of all bigrams* that are contained in the data set opened. The resulting list will be automatically saved under CorrelationList.csv in the root directory of AutoMap.

101.2 Text Example

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

101.3 Results after Bi-Gram list created

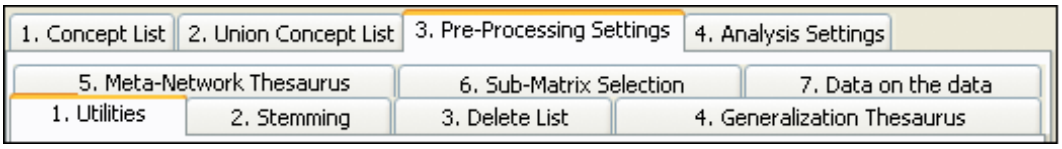
After pressing the Create Bi-Gram List button a file in the AutoMap directory will be created like below.

Frequency,Correlation,Concept,Concept,
3,1.0,of,the
2,0.6666667,and,the
1,0.33333334,a,new
1,0.33333334,a,summit
1,0.33333334,able,to
1,0.33333334,all,of
1,0.33333334,all,plans
1,0.33333334,all,simultanously
1,0.33333334,an,attack
1,0.33333334,and,poisoning
1,0.33333334,and,revenna
1,0.33333334,are,now
1,0.33333334,are,planning
1,0.33333334,as,backup
1,0.33333334,attack,which
1,0.33333334,attacked,all
1,0.33333334,backup,o'neill
1,0.33333334,by,inflitrating
1,0.33333334,carterselmak,as
1,0.33333334,considered,the
1,0.33333334,daniel,is
1,0.33333334,earth,with
1,0.33333334,emmisarry,presents
1,0.33333334,fell,swoop
1,0.33333334,fine,until
1,0.33333334,going,fine
1,0.33333334,gou'alds,in
1,0.33333334,help,of
1,0.33333334,in,one
1,0.33333334,infiltrate,the
1,0.33333334,inflitrating,a
1,0.33333334,information,and
1,0.33333334,is,able
1,0.33333334,is,attacked
1,0.33333334,is,going
1,0.33333334,jacob,carterselmak

1,0.33333334,kill,all
1,0.33333334,lord,gou'alds
1,0.33333334,military,daniel
1,0.33333334,military,of
1,0.33333334,new,emmisarry
1,0.33333334,new,information
1,0.33333334,now,considered
1,0.33333334,o'neill,and
1,0.33333334,of,sg-1
1,0.33333334,on,revenna
1,0.33333334,one,fell
1,0.33333334,plan,is
1,0.33333334,plan,to
1,0.33333334,planet,earth
1,0.33333334,planning,an
1,0.33333334,plans,are
1,0.33333334,poisoning,them
1,0.33333334,present,their
1,0.33333334,presents,some
1,0.33333334,rest,of
1,0.33333334,revenna,is
1,0.33333334,revenna,the
1,0.33333334,sg-1,stay
1,0.33333334,sgc,and
1,0.33333334,simultanously,they
1,0.33333334,some,new
1,0.33333334,states,military
1,0.33333334,stay,on
1,0.33333334,summit,and
1,0.33333334,summit,with
1,0.33333334,swoop,by
1,0.33333334,synopsis,the
1,0.33333334,system,lord
1,0.33333334,the,help
1,0.33333334,the,plan
1,0.33333334,the,planet
1,0.33333334,the,rest
1,0.33333334,the,sgc
1,0.33333334,the,summit
1,0.33333334,the,system
1,0.33333334,the,tok'ra

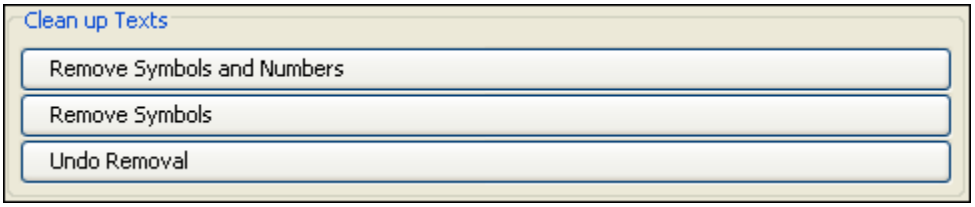
1,0.33333334,the,united
1,0.33333334,the,us
1,0.33333334,the,worst
1,0.33333334,their,plan
1,0.33333334,them,all
1,0.33333334,they,present
1,0.33333334,things,possible
1,0.33333334,to,infiltrate
1,0.33333334,to,the
1,0.33333334,tok'ra,are
1,0.33333334,united,states
1,0.33333334,until,a
1,0.33333334,us,military
1,0.33333334,which,would
1,0.33333334,with,jacob
1,0.33333334,with,the
1,0.33333334,worst,things
1,0.33333334,would,kill

102 Clean-up Text



This function removes all characters that are neither a letter nor a number depending on the option chosen. It maintains sentence marks converting 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.

Clean up Texts contains three buttons.



1. Remove Symbols and Number : Removes both numbers and sentence marks.
2. Remove Symbols : Removes sentence marks but retains the numbers
3. Undo Remove : Restores Symbols and/or numbers to the text.

The results of Clean up Text are found in *Panel 1 => Tab 2. Text after Symbol Removal*.

102.1 Text before Removal of Symbols and Number

On March 22, 2002 the Ren'al of the Tok'ra arrived @ the SGC to discuss their plan. 5 days later SG-1 and SG-17 arrive @ Revenna. The summit meeting with the 6 main System Lords* will be happening in 10 days. Daniel is faced with the question, "Will this prevent the System Lords from overthrowing the galaxy?"

*Gou'alds

102.2 Text after Removal of Symbols and Number

On March the Renal of the Tokra arrived the SGC to discuss their plan. days later SG and SG arrive Revenna. The summit meeting with the main System Lords will be happening in days. Daniel is faced with the question Will this prevent the System Lords from overthrowing the galaxy.

Goualds

This function removed all the numbers 22, 2002, 5, 1, 17, & 6. If also removed the punctuation commas, hyphens, & quote marks and the symbols @, and *

102.2.1 NOTE : Question marks are changed to periods

102.3 Text after Removal of Symbols

On March 22 2002 the Renal of the Tokra arrived the SGC to discuss their plan. 5 days later SG1 and SG17 arrive Revenna. The summit meeting with the 6 main System Lords will be happening in 10 days. Daniel is faced with the question Will this prevent the System Lords from overthrowing the galaxy.

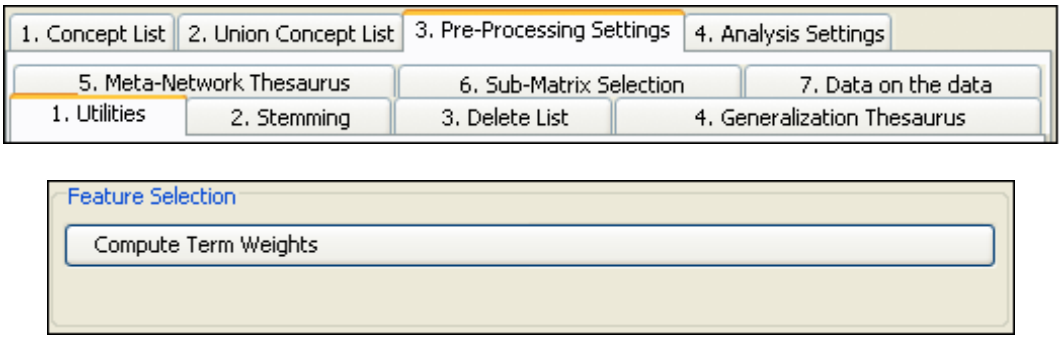
Goualds

102.4 N-gram Identification: Bigrams

More information about Collocation Identification in AutoMap.

To create a list of all bigrams that are contained in the data set opened, go to the *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 3. Utilities*. Scroll down to Create Bigram (Correlation) List section. The resulting list created will be automatically saved under CorrelationList.csv in the root directory of AutoMap.

103 Text Processing Panel - Feature Selection



From the main menu load a file using *File => Open single text file*.

From the main menu select *Create Union Concept List*.

From the main menu select *Output Storage Manager*. Locate the line for TFIDF computation.

103.1.1 NOTE : In Windows the simplest method for viewing the file is to highlight the pathway in the Output Storage Manager and copy it with [Ctrl-C]. Go to an explorer window and put this pathway in the address bar and press [Enter]. The file will automatically open

103.2 Opening the file in Excel

The text file will be in a comma-delimited format.

Open up Excel with a new spreadsheet.

In AutoMap highlight and copy the entire file. Paste this into Excel.

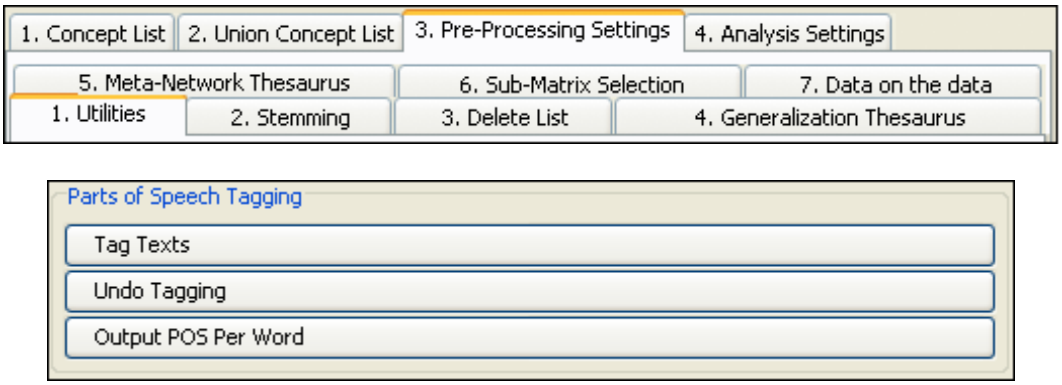
Highlight the column which contains the data. From the main menu select *Tools => Text to Columns….*

1. Select the Delimited radio button and hit [Next].
2. Make sure that only the Comma checkbox is checked. Make sure in the view area that the information is separated correctly. Press [Next]
3. Press [Finish]

The data can now be sorted by TF*IDF (Term frequency by Inverse Document Frequency) descending order.

103.2.1 This is an Alternative Method to find most important terms or features in text or text set.

104 Text Processing Panel - Parts of Speech Tagging



This feature will show you what part of speech each word is. This routine associates every word after the highest level of pre-processing applied so far with its Parts of Speech. 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.

- The Tag Texts option will take the text document and define each part of text.
- The Undo Tagging button removes any tags in the text
- The Output POS Per Word.

104.1 Text Example

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

Click the Tag Text button and view the results in *Panel 1 => Tab 8. Texts after Parts-of-Speech Tagging*.

Synopsis/VERB The/NOISE Tokra/NN are/VERB planning/VERB an/NOISE attack/NN which/NOISE would/MODAL kill/VERB all/NOISE of/NOISE the/NOISE System/NNP Lord/NNP Goualds/VERB in/NOISE one/NUM fell/VERB swoop/NOISE by/NOISE/NN a/NOISE summit/NN and/NOISE poisoning/VERB them/ANA all/NOISE./NOISE ./. They/ANA present/VERB their/NOISE plan/NN to/NOISE the/NOISE US/NOISE military/ADJ of/NOISE the/NOISE planet/NN Earth./NOISE ./. With/NOISE the/NOISE help/VERB of/NOISE the/NOISE SGC/NN and/NOISE the/NOISE United/NNP States/NNPS Military/NNP Daniel/NNP is/VERB able/ADJ to/NOISE infiltrate/NN the/NOISE summit/NN with/NOISE Jacob/NNP CarterSelmak/VERB as/NOISE backup./NOISE ./. ONeill/VERB and/NOISE the/NOISE rest/NN of/NOISE SG/NN stay/VERB on/NOISE Revenna./NOISE ./. The/NOISE plan/NN is/VERB going/VERB fine/ADJ until/NOISE a/NOISE new/ADJ emmisarry/NN presents/VERB some/NOISE new/ADJ information/NN and/NOISE Revenna/NN is/VERB attacked./NOISE ./. All/NOISE plans/NN are/VERB now/NOISE considered/VERB the/NOISE worst/ADJ things/NN possible./NOISE ./.

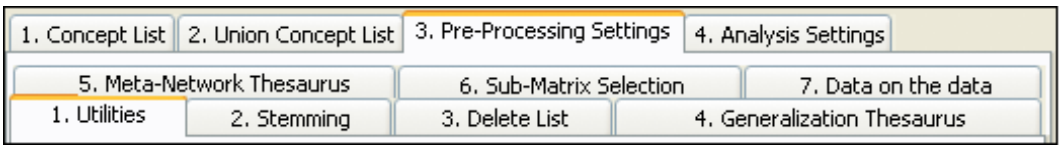
To remove the tags select the Undo Tagging button.

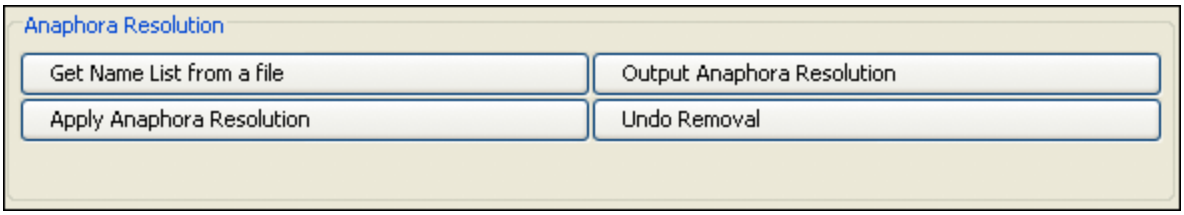
To save the Parts-of-Speech as a .csv file use the Output POS Per Word button. This file can be opened in Excel.

Below is a table of the Parts of Speech used

Code	Part of Speech	Code	Part of Speech
CC	Coordinating conjunction	PRP$	Possessive pronoun
CD	Cardinal number	RB	Adverb
DT	Determiner	RBR	Adverb, comparative
EX	Existential there	RBS	Adverb, superlative
FW	Foreign word	RP	Particle
IN	Preposition or subordinating conjunction	SYM	Symbol
JJ	Adjective	TO	to
JJR	Adjective, comparative	UH	Interjection
JJS	Adjective, superlative	VB	Verb, base form
LS	List item marker	VBD	Verb, past tense
MD	Modal	VBG	Verb, gerund or present participle
NN	Noun, singular or mass	VBN	Verb, past participle
NNS	Noun, plural	VBP	Verb, non-3rd person singular present
NNP	Proper noun, singular	VBZ	Verb, 3rd person singular present
NNPS	Proper noun, plural	WDT	Wh-determiner
PDT	Predeterminer	WP	Wh-pronoun
POS	Possessive ending	WP$	Possessive wh-pronoun
PRP	Personal pronoun	WRB	Wh-adverb

105Text Processing Panel - Anaphora Resolution





Get Name List from a file : Pressing the button brings up a dialog box from where a filename can be entered. This file will contain a list of predefined human names which the Anaphora Resolution will use to recognize whom relate to the pronouns. This feature is optional. If not used then Automap will use the Parts of Speech function to recognize words related to pronouns.

Anaphora is a linguistic instance of a grammatical expression referring to another. 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.

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

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).

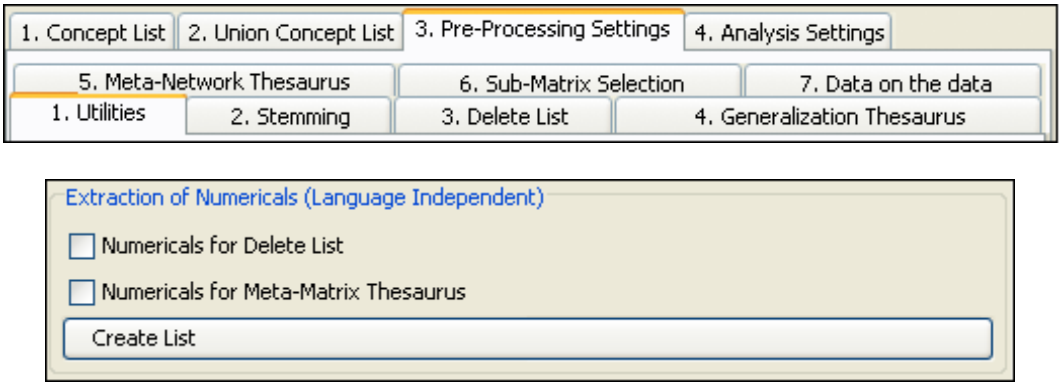
Larry called off work because he felt sick. he is anaphoric (he refers to Larry).

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.
The results can be viewed in *Text Display Panel => Tab 9. Texts after Anaphora Resolution*.

106 Text Processing Panel - Extraction of Numerals



106.1 Text Example

On March 22, 2002 the Ren'al of the Tok'ra arrived @ the SGC to discuss their plan. 5 days later, on March 27, 2002, SG-1 and SG-17 arrive @ Revenna. The summit meeting with the 6 main System Lords* will be happening in 10 days. Daniel is faced with the question, "Will this prevent the System Lords from overthrowing the galaxy?"

*Gou'alds

106.2 Extraction of Numerals for Delete List

To extract the numerals from a text place a checkmark in the box labeled Numerals for Delete List then click the Create List button. Unless you've changed the destination in the Output Storage Manager AutoMap will place it in the root directory of AutoMap. In the Output Storage Manager you have the opportunity to change the name for the file. The extraction file contains the following numerals:

22
2002
5
27
2002
1
17
6
10

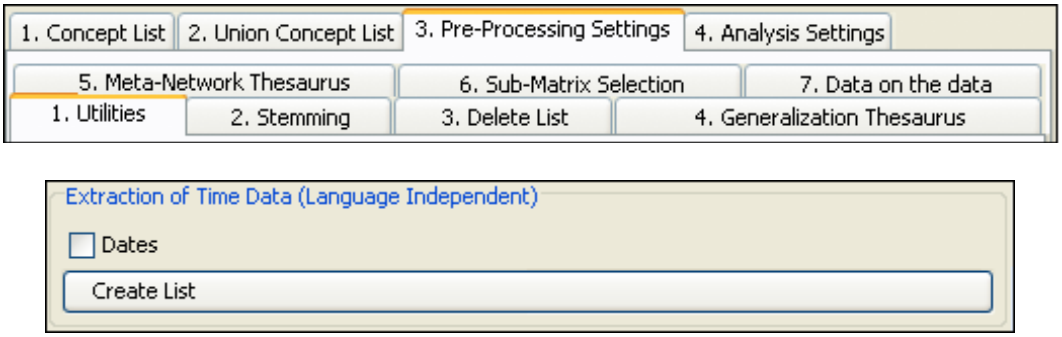
The file is saved as a .txt file which can be opened with a text editor.

106.3 Extraction of Numerals for Meta Network Thesaurus

The extraction of Numerals for Meta Network Thesaurus works in the same manner as above. Only the name of the file and the destination directory will be different as designated in the Output Storage Manager.

The file is saved as a .csv file which can be opened in Excel.

107 Text Processing Panel - Extraction of Time Data



107.1 Text Example for Extraction of Date

On March 22, 2002 the Ren'al of the Tok'ra arrived @ the SGC to discuss their plan. 5 days later, on March 27, 2002, SG-1 and SG-17 arrive @ Revenna. The summit meeting with the 6 main System Lords* will be happening in 10 days. Daniel is faced with the question, "Will this prevent the System Lords from overthrowing the galaxy?"

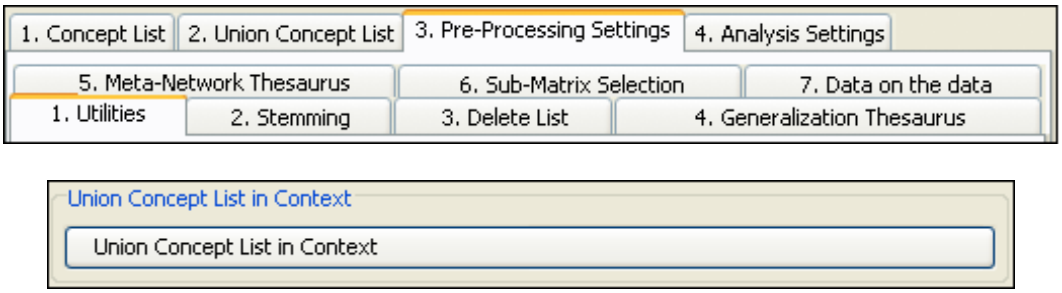
*Gou'alds

AutoMap creates a .csv file in the AutoMap directory which can be opened in Excel. The final location and filename can be altered in the Output Storage Manager

March 22, 2002

March 27, 2002

108Text Processing Panel - Union Concept List in Context



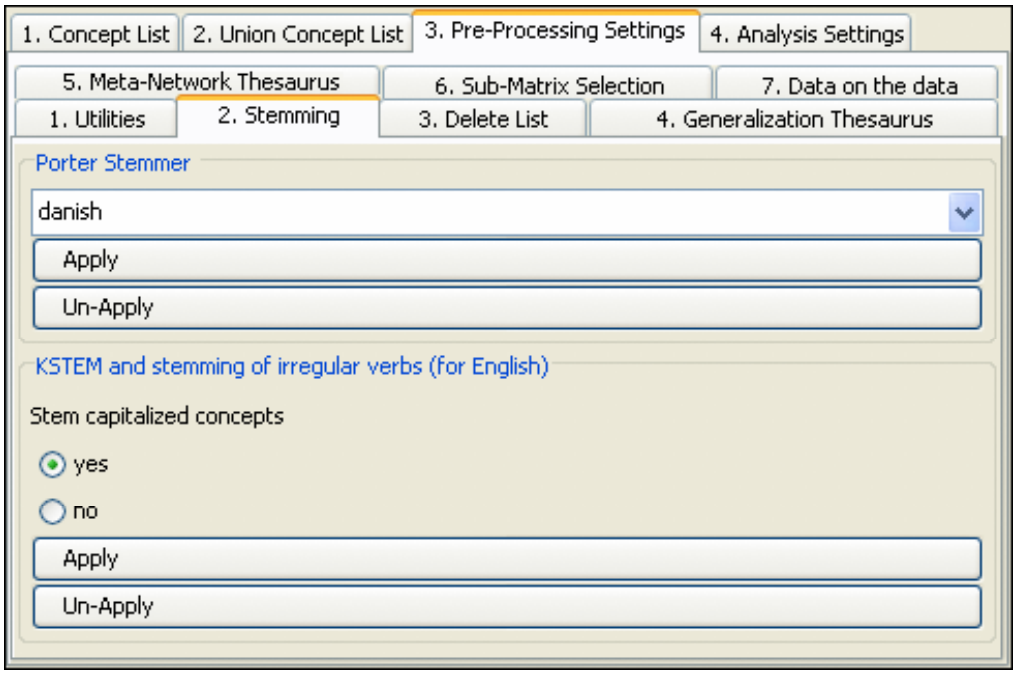
This function requires loading multiple files. Form the main menu select *files => Open multiple text files*. Navigate to the directory that contains the files and select [OK].

From the main menu select *File => Run Create Union Concept List.*

From the *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 1. Utilities* scroll down to the Union Concept List in Context section. Press the [Union Concept List in Context] button to create the list.

This function gives the concept with # of words in front and in back.

109Stemming



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. 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 pair wise 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.

110 Stemming Example

110.1 Original Text

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

110.2 Text After Porter Stemming

Synopsi: The Tok'ra be plan an attack which be kill all of the System Lord Gou'ald in on fall swoop by inflitr a summit and poison them all simultan. Thei present their plan to the US militari of the planet Earth. With the help of the SGC and the Unite State Militari Daniel be abl to infiltr the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stai on Revenna. The plan be go fine until a new emmisarri present some new inform and Revenna be attack. All plan be now consid the worst thing possibl.

110.3 Text After KSTEM

YES to Stem capitalized concepts)
The Bolded words in the following two examples shows the difference between Stemming, or not stemming, capitalized words.

Using YES to Stem capitalized concepts

Synopsi: The Tok'ra be plan a attack which be kill all of the System Lord Gou'ald in one fall swoop by infiltrating a summit and poison them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the Unite State Military Daniel be able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan be go fine until a new emmisarry present some new information and Revenna be attack. All plan be now consider the worst thing possible.

Using NO to Stem capitalized concepts

Synopsis: The Tok'ra be plan a attack which be kill all of the System Lord Gou'alds in one fall swoop by infiltrating a summit and poison them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel be able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan be go fine until a new emmisarry present some new information and Revenna be attack. All plan be now consider the worst thing possible.

111 Deletion List

- 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.

111.1 The Two main Delete Lists can be opened form the menu.

- Small predefined Delete List: The small predefined Delete List is compiled of the most common words that occur in English: a, an, and, some, many, this, that, these, those, the, all, one, every.
- Extensive predefined Delete List: An Extensive predefined Delete List contains a broader list of common words which includes the small defined delete list. a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself.

The Delete List will appear in Panel 2. This list is editable.

111.2 Format of the Delete List

1. The structure of a Delete List is one single Concept per line.

2. Avoid empty lines.
3. The Delete List is NOT case sensitive.
4. You can edit the Delete List within AutoMap if you wish.

111.3 Editing the Delete List

1. Editing within AutoMap: The structure of a Delete List is one single concept per line. You add concepts by typing one concept per line. Hit enter after entering a concept. Avoid empty lines.
2. Editing outside of AutoMap: Use a text editor to create a Delete List.

111.4 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.

   Delete matches from the text(s).

2. 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.

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.

111.5 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.

111.6 Save an applied Delete List

From the main menu select *File => Save Delete List As…* The Save dialog box will appear. Navigate to the directory you wish to save your delete file and select the OK button.

112 Deletion List Example

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

Extensive Delete List: a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself

112.1 Direct Adjacency:

The concepts in the delete list are totally removed the text.

Synopsis: Tok'ra are planning attack which kill all System Lord Gou'alds in one fell swoop by infiltrating summit poisoning all simultaneously. present plan military planet Earth. With help SGC United States Military Daniel is able infiltrate summit with Jacob Carter/Selmak backup. O'Neill, rest SG-1, stay on Revenna. plan is going fine until new emmisarry presents some new information Revenna is attacked. All plans are now considered worst things possible.

112.2 Rhetorical Adjacency:

The concepts in the delete list are replaced by xxx.

Synopsis: xxx Tok'ra are planning xxx attack which xxx kill all xxx xxx System Lord Gou'alds in one fell swoop by infiltrating xxx summit xxx poisoning xxx all simultaneously. xxx present xxx plan xxx xxx xxx military xxx xxx planet Earth. With xxx help xxx xxx SGC xxx xxx United States Military Daniel is able xxx infiltrate xxx summit with Jacob Carter/Selmak xxx backup. O'Neill, xxx xxx rest xxx SG-1, stay on Revenna. xxx plan is going fine until xxx new emmisarry presents some new information xxx Revenna is attacked. All plans are now considered xxx worst things possible.

The following concepts were added to the Delete List: by, all, in, is, on, until, some, are. Then the Delete List was applied again.

112.3 Direct Adjacency:

The concepts in the delete list are totally removed the text.

Synopsis: Tok'ra planning attack which kill System Lord Gou'alds one fell swoop infiltrating summit poisoning simultaneously. present plan military planet Earth. With help SGC United States Military Daniel able infiltrate summit with Jacob Carter/Selmak backup. O'Neill, rest SG-1, stay Revenna. plan going fine new emmisarry presents new information Revenna attacked. plans now considered worst things possible.

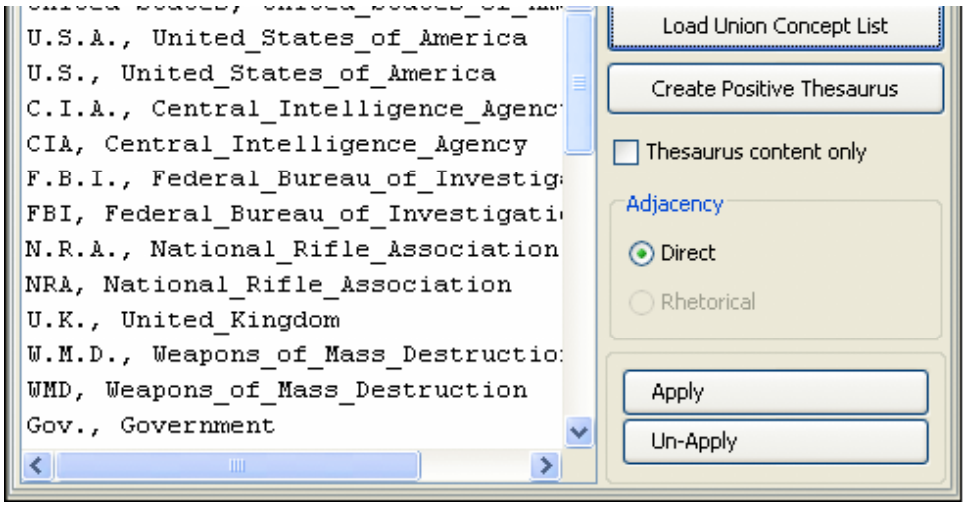
112.4 Rhetorical Adjacency:

The concepts in the delete list are replaced by xxx.

Synopsis: xxx Tok'ra xxx planning xxx attack which xxx kill xxx xxx xxx System Lord Gou'alds xxx one fell swoop xxx infiltrating xxx summit xxx poisoning xxx xxx simultaneously. xxx present xxx plan xxx xxx xxx military xxx xxx planet Earth. With xxx help xxx xxx SGC xxx xxx United States Military Daniel xxx able xxx infiltrate xxx summit with Jacob Carter/Selmak xxx backup. O'Neill, xxx xxx rest xxx SG-1, stay xxx Revenna. xxx plan xxx going fine xxx xxx new emmisarry presents xxx new information xxx Revenna xxx attacked. xxx plans xxx now considered xxx worst things possible.

113 Generalization Thesaurus

From the main menu select *File => Open Generalization Thesaurus*. A file chooser will appear. Navigate to the thesaurus you want select [Open]. The thesaurus will be displayed on *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 4. Generalization Thesaurus.*



113.1 Creating a General Thesaurus

A Thesaurus can be created by typing is within AutoMap, typing is in a text editor, or loading a union of concepts file. The concept list loaded into AutoMap can be refined by applying Named-Entity Recognition and Deletion prior to Generalization.

From the main menu select File ⇒ Create or refresh Union Concept List. Then select the [Load Union Concept List] button. The concept list is loaded into Automap and can be refined with Named-Entity Recognition and Deletion.

To illustrate the multi-step pre-processing technique copy the text passage below and save it as a TXT files with the names "stargateSynopsis.txt".

Text I.txt

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. It's going fine until new emmisarry present so new information and Revenna is attacked. All plans are now considered the worst things possible. .

113.2 Create Named-Entity List

From *Text Processing Panel => Tab 3. Pre-Processing Settings => SubTab 1. Utilities* scroll down to Named Entity Recognition. Select the [Create List] button.

113.2.1 NOTE : Remember to go to the Output Storage Manager to set the location of this file.

O'Neill
US
Synopsis: The Tok'ra
Earth. With
System Lord Gou'alds
All
United States
SG-1
System Lord
They
Revenna. It's
United States Military Daniel
Revenna
It's
SGC
Jacob Carter/Selmak

113.3 Build a Generalization Thesaurus

From the main menu select *File => Load Generalization Thesaurus => Open from file thes*. Navigate to the file created with the Named-Entity List. This will be the basis for a new thesaurus.

From this list build a Generalization Thesaurus by deleting unwanted concepts and editing existing concepts. After the comma place the concept name.

O'Neill, colonel_jack_o'neill
US, united_states
Earth, tauri_homeworld_earth
System Lord Gou'alds
United States, united_states
Revenna, tok'ra_homeworld
United States Military, united_states_military
Daniel, doctor_daniel_jackson
Revenna, tok'ra_homeworld_revenna
SGC, stargate_command
jacob, Jacob Carter/Selmak

113.4 Add further words

Further words can be added if desired such as summit, gou'ald_summit

113.5 Applying a Generalization Thesaurus

If you wish to apply a Delete List and a Generalization Thesaurus make sure to use the Delete List first. Then execute the following process:
Once the Generalization Thesaurus is ready select the [Apply] button in *Panel 2 => Tab 3, Pre-Processing Settings => SubTab 4. Generalization Thesaurus*. The text with the Thesaurus applied can be viewed in *Text Display Panel => Tab 5. Texts after Generalization.*

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 which can be either direct (default) or rhetorical.
3. To apply your Generalization Thesaurus with the settings you have specified click the Apply Thesaurus button. 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.
4. The pre-processed texts on the *Text Display Panel => Tab 5. Texts after Generalization*. 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.

113.6 Un-Apply a Generalization Thesaurus

To un-apply a Generalization Thesaurus that was applied to the data, go to *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 4. Generalization Thesaurus* and hit the Un-Apply button. The *Text Display Panel => Tab 5. Texts after Generalization* will be cleared.
Deletion: Rhetorical Adjacency
Create Union Concept List
Load Union Concept List

113.7 Editing a Thesaurus

The structure of a Thesaurus follows the five points below:

1. Every line contains Concept, Key Concept or in other words what it is in the file, new_set_of_words_to_use.
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 comma.

US Military, United_States_Military
SGC, Stargate_Command
Daniel, Doctor_Daniel_Jackson
O'Neill, Colonel_Jack_O'neill

5. The Thesaurus is NOT case sensitive.

113.8 Save a Generalization Thesaurus

From the main menu select *File ⇒ Output Storage Manager*. Find the line for Applied Generalization Thesaurus. Select the [Save file as…] button and select the folder to save the new thesaurus to. After giving it a name, select the [Open] button. Close the Output Storage Manager. From the main menu select *File => Save applied Generalization Thesaurus*.
The Thesaurus just created will be saved to the location designated in the Output Storage Manager.

114 Generalization Thesaurus Example

114.1 Original Text

Synopsis: The Tok'ra are planning an attack which would kill all of the System Lord Gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. They present their plan to the US military of the planet Earth. With the help of the SGC and the United States Military Daniel is able to infiltrate the summit with Jacob Carter/Selmak as backup. O'Neill, and the rest of SG-1, stay on Revenna. The plan is going fine until a new emmisarry presents some new information and Revenna is attacked. All plans are now considered the worst things possible.

114.2 Thesaurus Content Only - Off

synopsis: the tok'ra are planning an attack which would kill all of the system lord gou'alds in one fell swoop by infiltrating a summit and poisoning them all simultaneously. they present their plan to the united_states_military of the planet earth. with the help of the stargate_command and the united states military doctor_daniel_jackson is able to infiltrate the summit with jacob carter/selmak as backup. colonel_jack_o'neill, and the rest of sg-1, stay on revenna. the plan is going fine until a new emmisarry presents some new information and revenna is attacked. all plans are now considered the worst things possible.

114.3 Thesaurus Content Only - On – Direct Adjacency

:. united_states_military. stargate_command doctor_daniel_jackson. colonel_jack_o'neill,,...

114.4 Thesaurus Content Only - On – Rhetorical Adjacency

xxx: xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx united_states_military xxx xxx xxx xxx. xxx xxx xxx xxx xxx stargate_command xxx xxx xxx xxx xxx doctor_daniel_jackson xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx. colonel_jack_o'neill, xxx xxx xxx xxx xxx, xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx.

115 One Mode Network

Semantic network is the same as one-mode network. It is generated it by selecting from the main menu *Run Analysis => Map Analysis* (single or multiple, both)

Information can be found in the *Semantic Network Panel*.

116 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.

For a single text, select from the main menu *Run Analysis => Single Map Analysis*. To analyze a set of texts select from the main menu *Run Analysis => Multiple Map Analysis*. The Results will be displayed on the *Semantic Network Panel* and on *Results Panel => Tab 2. Statistics*.

If additional outputs were requested, 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.

117Semantic Network

After running analysis, the semantic network will be displayed on the *Semantic Network Panel*.

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.

117.1 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).

117.2 Example for Semantic Network Analysis

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

Text 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.

Text 2

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.

AutoMap's customized extensive Delete List:

a, an, and, as, at, but, for, he, her, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself

Direct Adjacency (default):

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

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

Rhetorical Adjacency:

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

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

Next we run Map Analysis on both texts using AutoMap's default Analysis Settings. These are the Map and Statistics outputs for the first text (direct adjacency): These are the Map and Statistics outputs for the second text (rhetorical adjacency):

Text 1

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

Stat:
# of concepts analyzed:
unique: 62
total: 88
# of concepts in statements:
unique: 54
total: 60
# of isolated concepts:
unique: 12
total: 13
# of statements:
unique: 54
total: 60
Density (based on Statements):
unique: 0.87
total: 0.97

Text 2

2: mr | 2: brown | 1: you | 1: worked | 1: weeds | 1: watered | 1: was | 1: too | 1: then | 1: take | 1: stopped | 1: said | 1: prof | 1: pleased | 1: plan | 1: over | 1: on | 1: now | 1: no | 1: next | 1: mr | 1: met | 1: lotus | 1:

longer | 1: june | 1: john | 1: happy | 1: had | 1: good | 1: garden | 1: following | 1: fence | 1: dog | 1: discussed | 1: day | 1: darren | 1: craye | 1: cray's | 1: cray | 1: carnations | 1: brown | 1: 1880 | 1: 13 | 1: 12

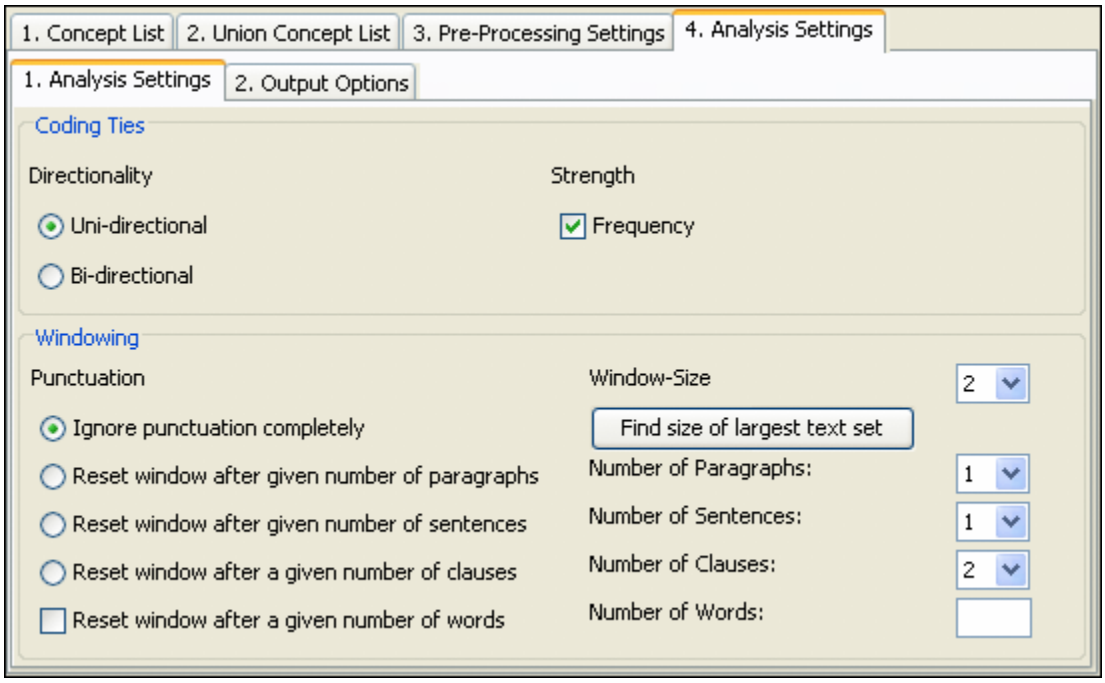
Stat:
# of concepts analyzed:
unique: 60
total: 84
# of concepts in statements:
unique: 55
total: 61
# of isolated concepts:
unique: 6
total: 7
# of statements:
unique: 55
total: 61
Density (based on Statements):
unique: 0.92
total: 1.02

117.2.1 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).

118Multi Mode Network

This section contains information regarding general settings and the Meta Network Thesaurus and the Sub-Matrix Thesaurus.
These are found in the *Text Processing Panel*

119 Analysis Setting



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.

119.1 Coding Ties

Specify the way statements are counted.

Directionality : Select one of the following two possibilities by checking the button.

Uni-Directional : (When coding a tie, *only 1st⇒2nd concept* should be noted)

Bi-Directional : (When coding a tie, *both 1st⇔2nd concept* shall be noted)

Strength : Strength will be printed out by default. To not print out Strength uncheck the item.

Frequency : If checked counts the cumulative frequency of every existing statement. If not checked the existence of frequency will be printed out (binary result).

119.2 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.

120 Meta Network Thesaurus

A Meta Network Thesaurus has to be applied if Meta Network Analysis should be performed.
A Meta Network Thesaurus associates concepts with Meta Network 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 Network Thesaurus, AutoMap searches the text(s) for the entries specified in the Meta Network Thesaurus and translates matches into related Meta Network 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 Network Thesaurus. The Meta Network Thesaurus is NOT case sensitive.
You might also see the Meta Network model as implemented in AutoMap to better understand the Meta Network.

120.1 Open a Meta Network Thesaurus

Select *File menu ⇒Open Meta Network Thesaurus ⇒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 in Text Processing Panel => Tab 3. Utilities => Sub-Tab 5. Meta Network Thesaurus.

120.1.1 NOTE : The Meta Network Thesaurus can be edited.

If you have a pre-defined Meta Network Thesaurus available that matches (some of) the concepts contained in the loaded Meta Network Thesaurus you can open this file.
Select *File => Open Meta Network Thesaurus => Open from file*.
If a concept contained in the pre-defined file matches a concept in the currently opened Meta Network Thesaurus the Meta Network categories assigned to this concept in the pre-defined file will be automatically assigned to the concept in the currently opened Meta Network Thesaurus.

120.1.2 NOTE : The pre-assigned Meta Network Thesaurus can be edited.

120.2 Edit a Meta Network Thesaurus

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

- One or multiple Meta Network categories to a concept by checking these categories. To unselect an assignment, uncheck the Meta Network categories.
- Self-defined Meta Network 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 Network 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 Network categories.
- No Meta Network category to a concept. In this case, this concept will not be considered for Meta Network pre-processing and Meta Network Analysis.

120.3 Build a Meta Network Thesaurus

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

1. Every line contains (Key) Concept / Meta Network category / Meta Network category/.... You can assign a (key) concept to as many Meta Network categories as you want to. If a concept that appears in the Meta Network Thesaurus written by you does not appear in the Concepts column of the Meta Network Thesaurus this will not result in a conflict. AutoMap automatically searches for concepts contained in the Concepts column of the Meta Network 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.

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

120.4 Apply a Meta Network Thesaurus

Meta Network 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 Network 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 Network Thesaurus with the settings you have specified click the Apply Thesaurus button on the *Text Processing Panel => Tab 3. Utilities => Sub-Tab 5. Meta Network Thesaurus*.
   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 Network category. The Thesaurus is case insensitive.
5. See the pre-processed texts on the *Text Display Panel => Tab 6. Texts after Meta Network Thesaurus*. Meta Network categories are confined by tags (e.g., <agent>). If multiple Meta Network 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>).

120.5 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 Network Thesaurus.
2. Translate matches into related Meta Network categories.
3. Maintain only Meta Network 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 in the *Text Processing Panel => Tab 3. Utilities => Sub-Tab 5. Meta Network Thesaurus*. You can now choose to either direct (default) or rhetorical adjacency. Then apply the Meta Network 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 Network 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 Network categories will be maintained. This rule does not apply if a concept consisting of more than one word is translated into a Meta Network category.

AutoMap by default does not select the Thesaurus content only option. Therefore, the Thesaurus content only item on *Text Processing Panel => Tab 3. Utilities => Sub-Tab 5. Meta Network Thesaurus* by default is not checked. Just apply the Meta Network Thesaurus. In order to switch to using the Thesaurus content only option check the Thesaurus content only item on *Text Processing Panel => Tab 3. Utilities => Sub-Tab 5. Meta Network Thesaurus* and then apply the Thesaurus again.

120.6 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 the *Text Processing Panel => Tab 3. Utilities => Sub-Tab 5. Meta Network Thesaurus*. Then apply the Meta Network Thesaurus. If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis. As a result, only Meta Network categories are displayed on *Text Display Panel => 6. Texts after Meta Network Thesaurus* will be considered for analysis. All Meta Network tags in the resulting text appear directly adjacent to each other.

120.7 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 *Text Processing Panel => Tab 3. Utilities => Sub-Tab 5. Meta Network Thesaurus*. Then apply the Meta Network Thesaurus. If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis. As a result, the Meta Network tags and the rest of the text are displayed on *Text Display Panel => 6. Texts after Meta Network Thesaurus* and will be considered for analysis. Original distances of Meta Network tags that represent the key concepts will be visually symbolized by placeholders (xxx) and considered for analysis.

120.8 Un-Apply a Meta Network Thesaurus

To un-apply a Meta Network Thesaurus that was applied to the data, go to *Text Processing Panel => Tab 3. Utilities => Sub-Tab 5. Meta Network Thesaurus* and hit the Un-Apply button. *Text Display Panel => 6. Texts after Meta Network Thesaurus* will be cleared.

120.9 Save an applied Meta Network Thesaurus

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

120.10 Save text(s) after application of Meta Network Thesaurus

To save the text(s) after the application of the Meta Network Thesaurus, click the File menu, select Save Text(s) after Meta Network 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.

121 Example of Meta Network Thesaurus

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

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

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

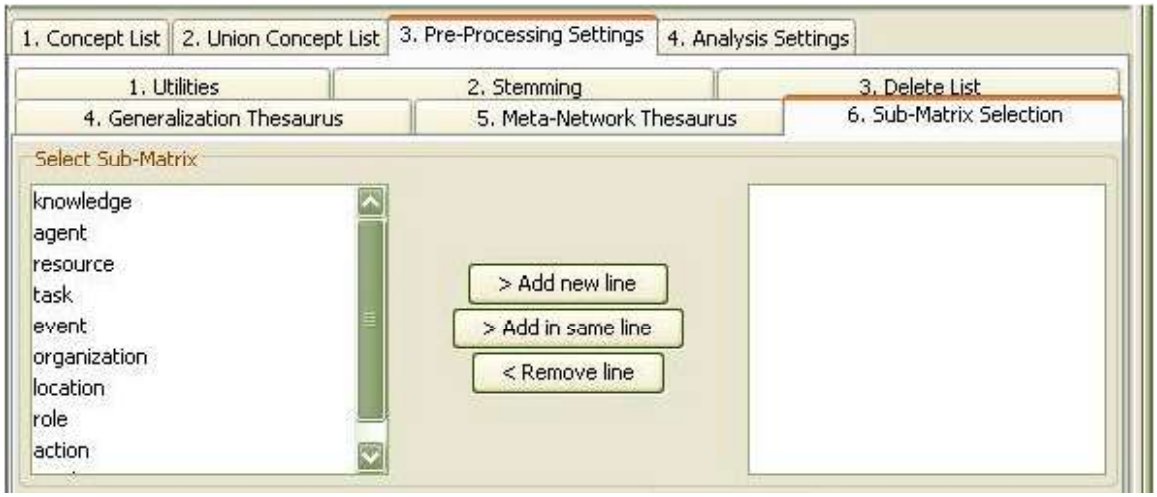
The customized extensive Delete List was applied to this text. The resulting text looks like this:

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

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

Then from the main menu select *File menu => Open Meta Network Thesaurus => Open from highest level of pre-processing*.

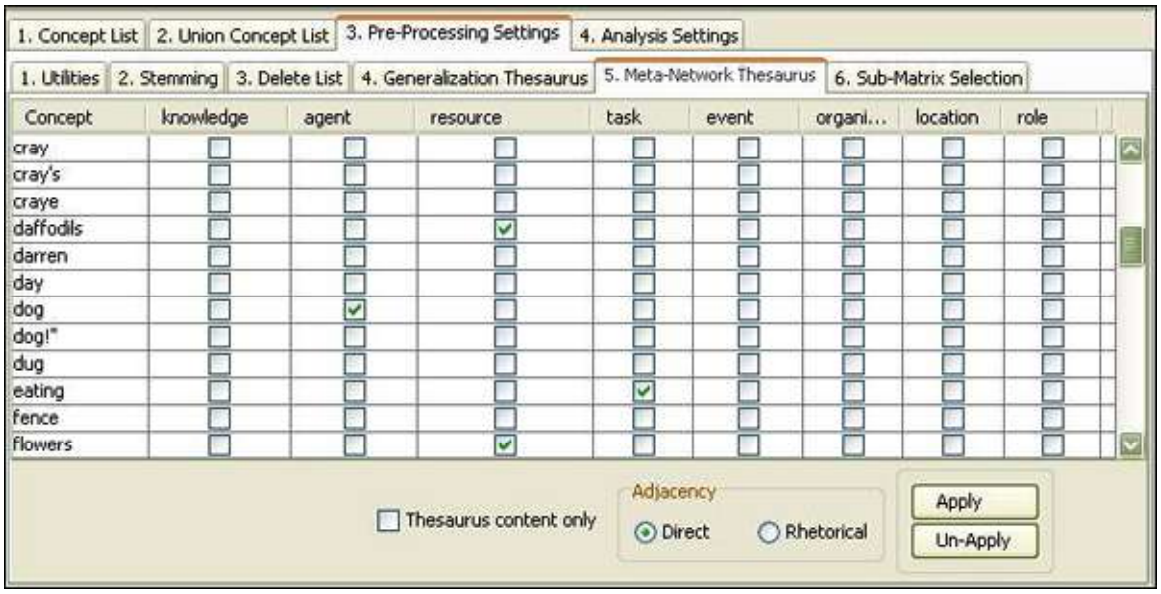
The union of concepts from the highest level of pre-processing will be displayed in alphabetical order on the Text Processing Panel ⇒ Tab 3. Pre-Processing Settings ⇒ Sub-Tab 5. Meta Network Thesaurus. As we did not pre-process the text the original input sentence is used for input for the Meta Network Thesaurus.



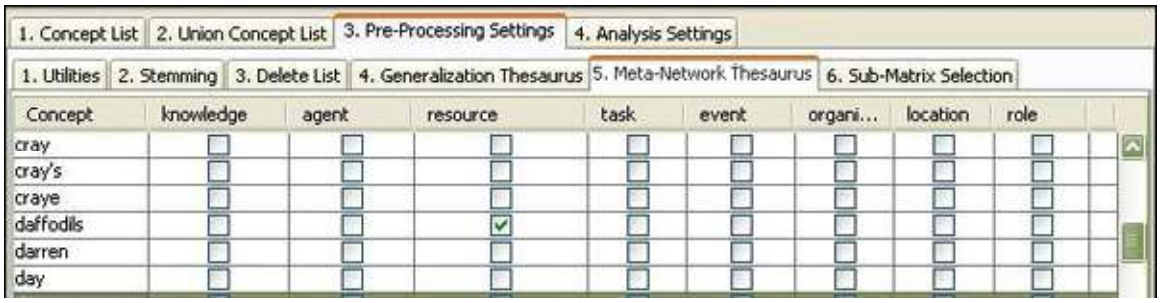
Furthermore, we have prepared a Meta Network Thesaurus that we stored on our machine. This file looks like this:

- 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 Network Thesaurus and chose Open from file. AutoMap searched the opened Meta Network Thesaurus for the words contained in the prepared Thesaurus. When it found a match it assigned the words in the opened Thesaurus the Meta Network 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 Network categories yet (e.g., contact, copenhagen). Not all concepts were associated with Meta Network categories (e.g., mobile).



Thesaurus content only NOT checked

mr . cray's brown < agent> < task> the < resource> blossom at 10 am . mrs . brown was unhappy with the < agent> . she < task> at it saying "you impossible dog!" but the < agent> kept < task> the < resource> and < task> . she asked mr . cray to stop the < agent> . he couldn't . mrs . brown planted < resource> and weeded the garden . the silly < agent> % dug up the < resource> looking for a vole on june 12 , 1880 . weeding was no longer needed . prof . darren , mrs . brown & mr . cray < task> the next day to concoct a plan . < agent> and mrs . brown put up a scarecrow . she thought it would scare the < agent> . mr . craye put up a fence . problem solved . then mrs . brown planted < resource> , < resource> , < resource> , and < resource> .

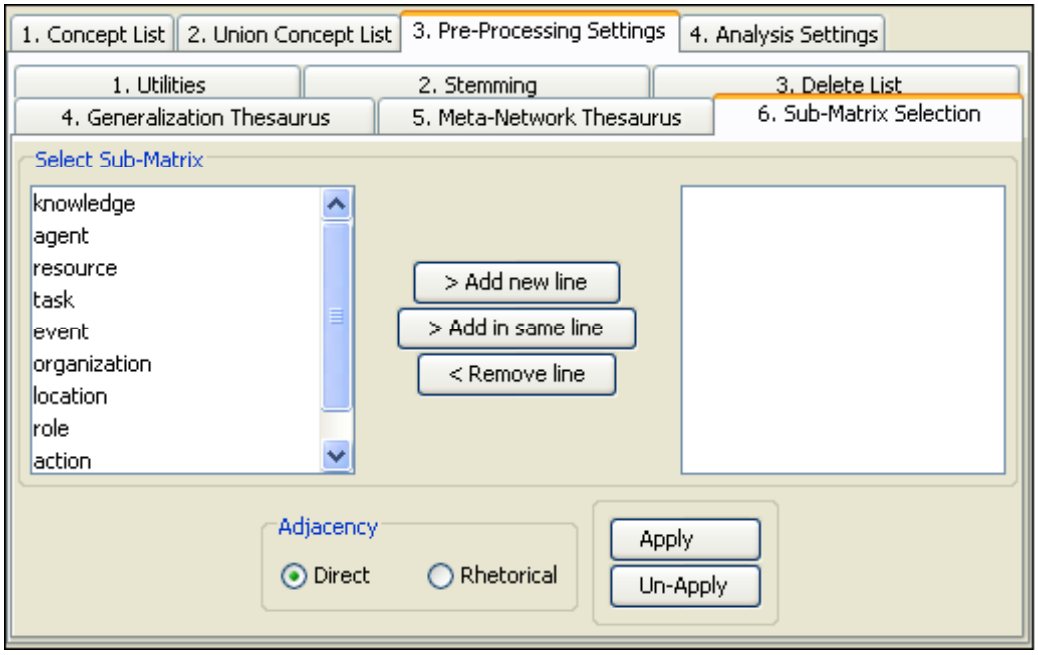
Thesaurus content only, Direct Adjacency

. < agent> < task> < resource> . . < agent> . < task> < agent> < task> < resource> < task> . . < agent> . . . < resource> . < agent> < resource> . . . . . < task> . < agent> . . < agent> . . . . . < resource> < resource> < resource> < resource> .

Thesaurus content only, Rhetorical Adjacency

xxx . xxx xxx < agent> < task> xxx < resource> xxx xxx xxx xxx . xxx . xxx xxx xxx xxx xxx < agent> . xxx < task> xxx xxx xxx xxx xxx xxx xxx xxx < agent> xxx < task> xxx < resource> xxx < task> . xxx xxx xxx . xxx xxx xxx xxx < agent> . xxx xxx . xxx . xxx xxx < resource> xxx xxx xxx xxx . xxx xxx < agent> xxx xxx xxx xxx < resource> xxx xxx xxx xxx xxx xxx xxx xxx . xxx xxx xxx xxx xxx . xxx xxx xxx . xxx xxx . xxx xxx xxx . xxx < task> xxx xxx xxx xxx xxx xxx xxx . < agent> xxx xxx . xxx xxx xxx xxx xxx . xxx xxx xxx xxx xxx xxx < agent> . xxx . xxx xxx xxx xxx . xxx xxx . xxx xxx . xxx xxx < resource> < resource> < resource> xxx < resource> .

122Sub-Matrix Selection



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

122.1.1 NOTE : Sub-Matrix Selection can only be performed if Meta Network Thesaurus was applied.

122.2 Select Sub-Matrix

Left Panel : Lists the concepts that are available.

Middle Buttons : Allow you to add or remove concepts to a Thesaurus.

[> Add new line] : Creates a new line and moves selected concept to right window.

[> Add in same line] : Attaches the concept selected in the left window to the currently selected line in the right window.

[< Remove line] : Removes the selected line in the right window.

Right Panel : Displays the your choices of concepts to apply.

122.3 Adjacency

You can also apply Adjacency in the forms of Direct and Rhetorical

123 Selecting Sub-Matrix Categories

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

If you do not have a sub matrix selection file available

Create and modify a sub matrix selection.

Go to *Text Processing Panel => Tab 3. Pre-Processing settings => Sub-Tab 6. Sub-Matrix Selection*. You have three options:

1. You can select as many Sub-Matrix categories per row as you wish by clicking on a category in the left window and selecting the [> Add in same line] button to move it to the right window.
2. You can select as many rows of sub matrix selections as you wish by selecting a category in the left window on and move it to the right window by selecting the [> Add in new line] button.
3. To unselect a selected Sub-Matrix category, click on the row in the right window and hit the [< Remove line] button. The row will disappear from the right window.

123.3 If you have a sub matrix selection file available

From the main menu and select *File => Open Sub-Matrix Selection => Open from file*.
You can modify this sub matrix selections.

123.4 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.
From the main menu select *File menu => Open Sub-Matrix Selection => Open from file*. You can modify this sub matrix selections.

123.5 Create Sub-Matrix Selection

Go to *Text Processing Panel => Tab 3. Utilities => Sub-Tab 6. Sub-Matrix Selection* and select a Sub-Matrix category. Then select the > Add in same line button. The selected category appears in the right text field on the *Text Processing Panel => Tab 3. Utilities => Sub-Tab 6. Sub-Matrix Selection*.

Multiple Sub-Matrix categories can be added to each row from left window in *Text Processing Panel => Tab 3. Utilities => Sub-Tab 6. Sub-Matrix Selection* and moved to the right window by selecting the > Add in same line button. Multiple rows can be added by selecting a category in the left window on and moving it to the right window by selecting 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 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. The row will disappear from the right window.

123.5.1 NOTE : 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.

123.5.2 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/.

123.6 Applying a Sub-Matrix Selection

1. Select an Adjacency option. Adjacency can be either direct (default) or rhetorical. Direct adjacency is the default.
2. Apply the options by selecting the Apply Sub-Matrix Selection button. AutoMap will then search the text(s) for Meta Network tags. If a match is found it will be translated into the related concept.

   The Thesaurus IS case insensitive.

3. The text can be viewed in the *Text Display Panel => Tab 1. Original Texts*.

123.7 Un-Apply a Sub-Matrix

To un-apply a Sub-Matrix Selection that was applied to the data, go to *Text Processing Panel => Tab 3. Pre-processing Settings => Sub-Tab 6. Sub-Matrix Selection* and select the [Un-Apply] button. The text in *Text Display Panel => Tab 7. Texts after Sub-Matrix Selection* will be cleared.

123.8 Direct Adjacency

Direct adjacency means that original distances of concepts that represent Meta Network categories will neither be visualized nor considered for analysis.

If the user does not change the adjacency option, AutoMap uses direct adjacency for analysis.

As a result, only concepts that represent Meta Network categories are displayed on *Text Display Panel => Tab 7. Texts after Sub-Matrix Selection* and will be considered for analysis. All concepts in the resulting text appear directly adjacent to each other.

123.9 Rhetorical Adjacency

As a result, concepts that represent Meta Network categories are displayed on *Text Display Panel => Tab 7. Texts after Sub-Matrix Selection* and will be considered for analysis. Original distances of concepts that represent Meta Network categories will be visually symbolized by placeholders (xxx) and considered for analysis.

123.10 Save Sub-Matrix Selection

Apply the Sub-Matrix Selection before you save it. In *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 6. Sub-Matrix Selection*, click the File menu, select Save applied Sub-Matrix Selection as applied.

123.11 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.

124 Example of a Sub-Matrix

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

Open a Sub-Matrix Selection.

Apply the Sub-Matrix Selection with the following settings:

124.1 Select:

Agent / Location / Action
Location / Agent / Action
Action / Location / Agent
Direct Adjacency (default)

Output:

. dog.. dog. dog.. dog.... dog,.. .,... prof_john_darren.. dog.....,,,.

124.2 Select:

Agent / Location / Action
Location / Agent / Action
Action / Location / Agent
Rhetorical Adjacency

Output:

xxx. xxx xxx dog xxx xxx xxx xxx xxx xxx xxx. xxx. xxx xxx xxx xxx xxx dog. xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx dog xxx xxx xxx xxx xxx xxx. xxx xxx xxx. xxx xxx xxx xxx dog. xxx xxx. xxx. xxx xxx xxx xxx xxx xxx xxx. xxx xxx dog xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx

xxx

xxx. xxx, xxx. xxx xxx xxx. xxx xxx xxx xxx xxx xxx xxx xxx xxx. prof_john_darren xxx xxx. xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx dog. xxx. xxx xxx xxx xxx. xxx xxx. xxx xxx. xxx xxx xxx, xxx, xxx, xxx xxx.

124.3 Select:

Full Meta Network
Direct Adjacency (default)

Output:

. dog eating lotus.. dog. yelling dog eating flowers weeds.. dog... roses. dog roses,.. .,.. met. prof_john_darren.. dog..... lotus, carnations, daffodils, roses.

124.4 Select:

Full Meta Network
Rhetorical Adjacency (default)

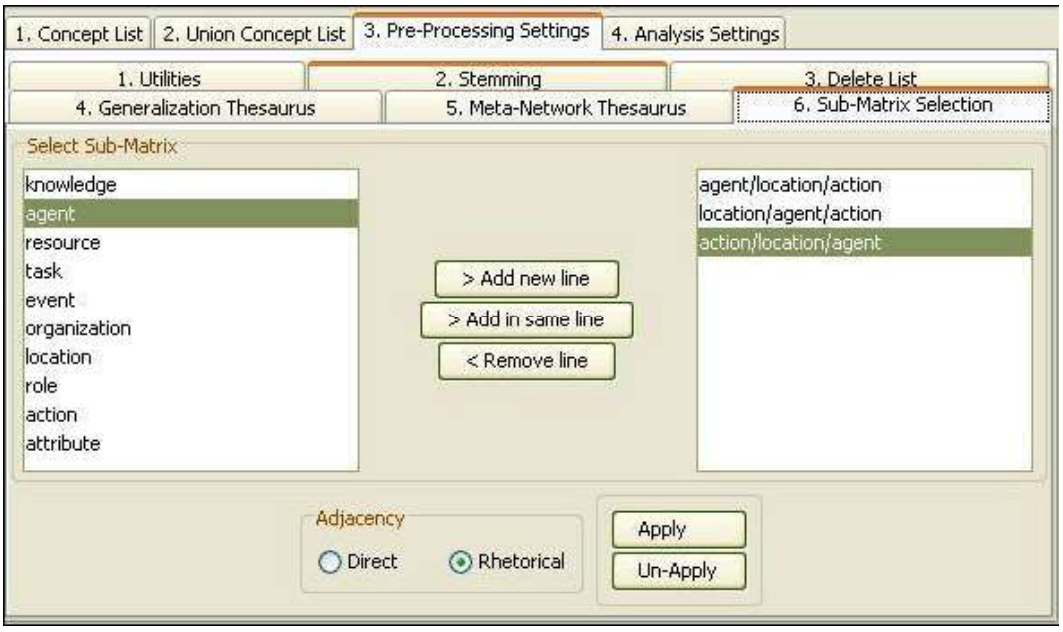
Output:

xxx. xxx xxx dog eating xxx lotus xxx xxx xxx xxx. xxx. xxx xxx xxx xxx xxx dog. xxx yelling xxx xxx xxx xxx xxx xxx xxx xxx dog xxx eating xxx flowers xxx weeds. xxx xxx xxx. xxx xxx xxx xxx dog. xxx xxx. xxx. xxx xxx roses xxx xxx xxx xxx. xxx xxx dog xxx xxx xxx xxx roses xxx xxx xxx xxx xxx xxx xxx, xxx. xxx xxx xxx xxx xxx. xxx

xxx

xxx. xxx, xxx. xxx xxx xxx. xxx met xxx xxx xxx xxx xxx xxx xxx. prof_john_darren xxx xxx. xxx xxx xxx xxx xxx. xxx xxx xxx xxx xxx xxx dog. xxx. xxx xxx xxx xxx. xxx xxx. xxx xxx. xxx xxx lotus, carnations, daffodils, xxx roses.

Detailed example for Sub-Matrix Selection:



125Saving a Sub-Matrix

From the main menu select *File => Save Text(s) after Sub-Matrix Selection*. All texts are automatically saved in a folder called preprocessed in the root directory of AutoMap. The filename will be after_SubMatrixSelection_NameOfYourText.txt".

126Meta Network Text Analysis

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

126.1 Steps to take before running the Meta Network analysis:

- Run any Text Pre-Processing with a Meta Network Thesaurus. Also is a Delete List and/or a Generalization Thesaurus is used make sure to use them before running the Meta Network Thesaurus.
- Specify the analysis settings. If no specific settings are made, the default settings will be applied.
- Specify any additional output options.

126.2 Analyzing Text(s)

To analyze a single text select from the main menu *Run Analysis => Single Meta Network Text Analysis*. To analyze a set of texts, select *Run Analysis => Multiple Meta Network Text Analysis*.
The Results will be displayed on *Panel 3* and on the *Results Panel => Tab 3. Statistics*. If additional outputs were requested they will be generated and stored under the directories specified in the Output Storage Manager.
If multiple texts were analyzed they can be browsed through the results using the arrows on the menubar.

126.3 Results

Results of multiple analyses are automatically saved in a folder named output under AutoMap's default directory. This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed. In addition there is a stat_output.xls file a that contains the stats of all texts.
The Stat Output folder is overwritten with every new analysis you run. If this needs to be saved either rename the folder or move the files.

126.4 Example for Meta Network 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 Network-Thesaurus. This are the resulting texts:

Thesaurus content only (default) : mr . < agent> < agent> dog ate lotus < event> < attribute> am . mrs . < agent> was unhappy with dog . yelled saying < agent resource> impossible dog!" dog kept eating flowers weeds . asked mr . cray stop dog . couldn't . mrs . < agent> planted roses weeded garden . silly dog <

attribute> dug up roses looking vole on june < attribute> , < attribute> . weeding was no longer needed . prof . < agent> , mrs . < agent> < attribute> mr . cray met next day < task> plan . john < agent> mrs . < agent> put up scarecrow . thought scare dog . mr . < agent> put up fence . problem solved . then mrs . < agent> planted lotus , < resource> , < resource> , roses .

Thesaurus content only, Direct Adjacency : . < agent> < agent> < event> < attribute> . . < agent> . < agent resource> . . . . . < agent> . < attribute> < attribute> < attribute> . . . < agent> . < agent> < attribute> . < task> . < agent> . < agent> . . . < agent> . . . < agent> < resource> < resource> .

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

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

Not Selected - Thesaurus content only (default)

1 10 am | 1 12 1880 | 1 1880 weeding | 1 am mrs | 1 asked mr | 1 ate lotus | 1 blossom 10 | 1 brown was | 1 carnations daffodils | 1 concoct plan | 1 couldn't mrs | 1 cray stop | 1 cray's brown | 1 craye put up | 1 daffodils roses | 1 day concoct | 1 dog yelled | 1 dug up | 1 eating flowers | 1 fence problem | 1 flowers weeds | 1 garden silly | 1 impossible dog | 1 john darren | 1 june 12 | 1 kept eating | 1 longer needed | 1 looking vole | 1 lotus carnations | 1 met next | 1 mr craye | 1 needed prof | 1 next day | 1 no longer | 1 on june | 1 plan john | 1 planted roses | 1 problem solved | 1 prof darren | 1 put up | 1 put up fence | 1 roses weeded | 1 saying you | 1 scare dog | 1 scarecrow thought | 1 silly dog | 1 solved then | 1 stop dog | 1 then mrs | 1 thought scare | 1 unhappy with | 1 up scarecrow | 1 vole on | 1 was unhappy | 1 weeded garden | 1 weeding was | 1 weeds asked | 1 with dog | 1 yelled saying | 1 you impossible | 2 brown planted | 2 darren mrs | 2 mr cray

# of concepts analyzed:
unique: 62
total: 88
# of concepts in statements:
unique: 80
total: 87
# of isolated concepts:
unique: 0
total: 0
# of statements:
unique: 80
total: 87
Density (based on Statements):
unique: 1.29
total: 1.4
Thesaurus content only, Direct Adjacency

1 10 am | 1 12 1880 | 1 1880 weeding | 1 am mrs | 1 asked mr | 1 ate lotus | 1 blossom 10 | 1 brown was | 1 carnations daffodils | 1 concoct plan | 1 couldn't mrs | 1 cray stop | 1 cray's brown | 1 craye put up | 1 daffodils roses | 1 day concoct | 1 dog yelled | 1 dug up | 1 eating flowers | 1 fence problem | 1 flowers weeds | 1 garden silly | 1 impossible dog | 1 john darren | 1 june 12 | 1 kept eating | 1 longer needed | 1 looking vole | 1 lotus carnations | 1 met next | 1 mr craye | 1 needed prof | 1 next day | 1 no longer | 1 on june | 1 plan john | 1 planted roses | 1 problem solved | 1 prof darren | 1 put up | 1 put up fence | 1 roses weeded | 1 saying you | 1 scare dog | 1 scarecrow thought | 1 silly dog | 1 solved then | 1 stop dog | 1 then mrs | 1 thought scare | 1 unhappy with | 1 up scarecrow | 1 vole on | 1 was unhappy | 1 weeded garden | 1 weeding was | 1 weeds asked | 1 with dog | 1 yelled saying | 1 you impossible | 2 brown planted | 2 darren mrs | 2 mr cray | 5 mrs brown

# of concepts analyzed:
unique: 62
total: 88
# of concepts in statements:
unique: 80
total: 87
# of isolated concepts:
unique: 0
total: 0
# of statements:
unique: 80
total: 87
Density (based on Statements):
unique: 1.29
total: 1.4
Thesaurus content only, Rhetorical Adjacency

1 blossom 10 | 1 brown was | 1 carnations daffodils | 1 concoct plan | 1 couldn't mrs | 1 cray stop | 1 cray's brown | 1 craye put up | 1 daffodils roses | 1 day concoct | 1 dog yelled | 1 dug up | 1 eating flowers | 1 fence problem | 1 flowers weeds | 1 garden silly | 1 impossible dog | 1 john darren | 1 june 12 | 1 kept eating | 1 longer needed | 1 looking vole | 1 lotus carnations | 1 met next | 1 mr craye | 1 needed prof | 1 next day | 1 no longer | 1 on june | 1 plan john | 1 planted roses | 1 problem solved | 1 prof darren | 1 put up | 1 put up fence | 1 roses weeded | 1 saying you | 1 scare dog | 1 scarecrow thought | 1 silly dog | 1 solved then | 1 stop dog | 1 then mrs | 1 thought scare | 1 unhappy with | 1 up scarecrow | 1 vole on | 1 was unhappy | 1 weeded garden | 1 weeding was | 1 weeds asked | 1 with dog | 1 yelled saying | 1 you impossible | 2 brown planted | 2 darren mrs | 2 mr cray | 5 mrs brown

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

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

127 Sub-Matrix Text Analysis

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

127.1 Steps to take before running the Sub-Matrix analysis:

- Text Pre-processing with a Meta Network Thesaurus. If a Delete List and/or a Generalization Thesaurus is also to be used make sure these tools are used before using the Meta Network Thesaurus.
- Select the Sub-Matrix Categories.
- Specify analysis settings. If no settings are selected then the defaults will be used.
- Specify additional output options. If no additional options are selected then the default outputs will be generated.

127.2 Exclude Entities

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

127.3 The use of ontologies

Go to *Text Processing Panel => Tab 4. Analysis Settings => Sub-Tab 2. Output Options*. Scroll down to Concepts not forming edges. Press the [Refresh] button to automatically generate a list of all entities in the applied ontology. Use the < Select and > Unselect buttons to move entities from one window to the other. When finished press the [Apply Selection] button.

127.4 Network or Social Network Text Analysis

Before running Sub-Matrix Text Analysis decide whether or not to perform Network Text Analysis (NTA) or Social Network Text Analysis (SNTA). Choose one of the radio buttons in the Concepts not forming edges section. Both types are a form of Sub-Matrix Text Analysis, but differ in what they measure:

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).

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.

127.5 Run the Sub-Matrix Text Analysis.

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

The Results will be displayed on *Panel 3* and on *Panel 4 => Tab 2. Statistics*. If additional outputs were requested they 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 in the AutoMap default directory. This folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as stat_output.xls file a that contains the stats of all texts.

The Stat Output folder is overwritten with every new analysis you run. So if you want to save the results of a current Stat Output folder just rename the folder.

127.6 Example for Sub Matrix Text Analysis

This example is based on the example for Sub Matrix Selection. Our Text I.txt. was used as input:

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

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

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

Select: Knowledge/Agent/Resource

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

Select: Knowledge/Agent/Resource

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

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

1 lotus carnations | 1 met the | 1 mr craye | 1 needed prof darren | 1 next day | 1 no longer | 1 on june | 1 plan john | 1 planted roses | 1 problem solved | 1 put up | 1 put up a | 1 roses looking | 1 saying you | 1 scare the | 1 scarecrow she | 1 she yelled | 1 silly dog | 1 solved then | 1 stop the | 1 the silly | 1 then mrs | 1 thought it | 1 to stop | 1 unhappy with | 1 up the | 1 vole on | 1 was unhappy | 1 weeded the | 1 weeding was | 1 weeds she | 1 with the | 1 would scare | 1 yelled at | 1 you impossible | 2 brown planted | 2 mr cray | 4 the dog | 5 mrs brown

1 12 pm | 1 13 1880 | 1 1880 prof | 1 and weeds | 1 at 12 | 1 ate the | 1 blossom at | 1 brown would | 1 carnations daffodils | 1 cray was | 1 cray's brown | 1 craye painted | 1 daffodils and | 1 darren mrs | 1 day

mrs | 1 dinner and | 1 discussed how | 1 dog the | 1 down the | 1 eating the | 1 fence then | 1 fertilized the | 1 flowers and | 1 following week | 1 for the | 1 garden on | 1 good dog | 1 had worked | 1 happy with | 1 his fence | 1 how the | 1 it was | 1 john darren | 1 june 13 | 1 longer ate | 1 lotus carnations | 1 met over | 1 mr craye | 1 next day | 1 no longer | 1 now happy | 1 on june | 1 over dinner | 1 painted his | 1 plan had | 1 pleased too | 1 pm the | 1 prof darren | 1 roses and | 1 said you | 1 scarecrow the | 1 scary for | 1 she thought | 1 stopped eating | 1 take down | 1 the scarecrow | 1 then mrs | 1 thought it | 1 too scary | 1 was too | 1 watered the | 1 weeds mr | 1 week she | 1 with the | 1 worked john | 1 would take | 1 you good | 2 brown watered | 2 mr cray | 3 the dog | 5 mrs brown

unique: 72
total: 118
# of concepts in statements:
unique: 108
total: 117
# of isolated concepts:
unique: 0
total: 0
# of statements:
unique: 108
total: 117
Density (based on Statements):
unique: 1.5
total: 1.63
# of concepts analyzed:
unique: 68
total: 107
# of concepts in statements:
unique: 98
total: 106
# of isolated concepts:
unique: 0
total: 0
# of statements:
unique: 98
total: 106
Density (based on Statements):
unique: 1.44
total: 1.56

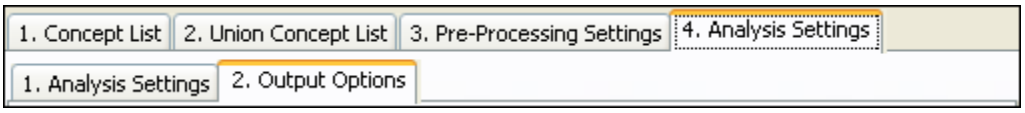
128Output

Using the menu item *File => Output Storage Manager* allows the viewing of the pathways for output results. These pathways can be changed by the user.

Controls for outputs are found in *Text Processing Panel => Tab 4. Analysis Settings => Sub-Tab 2. Output Options*.

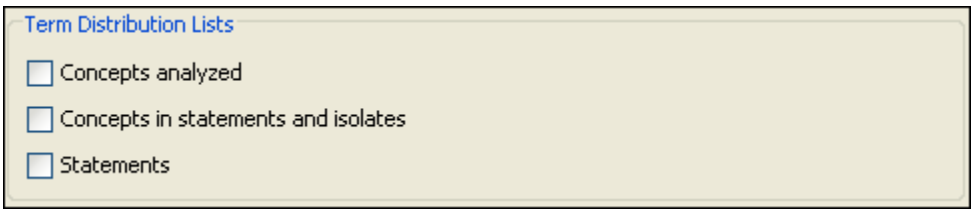
The *Results Panel => Tab 2. Statistics* contains information regarding the processed texts.

129 Output Options



The map and statistic output generated by Automap are displayed in *Panels 3 and 4*, respectively. In addition, Automap offers further output options that are available in *Text Processing Panel => Tab 4. Analysis Settings => SubTab 2. Output Options*. All additional outputs are only generated after the analysis are run.

129.1 Term Distribution List



The map and the statistic output generated by AutoMap are displayed on the *Semantic Network Panel and the Results Panel*, respectively. In addition to that AutoMap offers further output options that can be chosen on *Text Processing Panel => Tab 4. Analysis Settings => Sub-Tab 2. Output Options* Scroll down Term Distribution Lists. All additional outputs are only generated after analyses were run.

129.2 Points To Consider

- No list will be generated by default.
- All lists 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 will be automatically saved in a folder called Term Distribution Lists under the root directory of AutoMap.
- The Term Distribution Lists folder is overwritten with every new analysis you run. So if you want to save the results of a current Term Distribution Lists folder just rename the folder. For each Term Distribution List checked two output lists are generated.

129.3 Types and content of Term Distribution Lists
129.4 Concepts analyzed

List of concept analyzed.csv - Concept, Text, Frequency

Statistics of concepts analyzed.csv - Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts

129.5 Concepts in Statements and Isolates

List of concept in statements.csv - Concept, Text, Frequency

Statistics of concept in statements.csv - Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts

List of isolates in statements.csv - Concept, Text, Frequency

Statistics of isolates.csv - Concept, Cumulated sum across text set, Number of text concept occurs in, Percentage of texts concept occurs in, Texts

129.6 Statements

List of statements.csv - Statement, Text, Frequency

Statistics of statements.csv - Statement, Cumulated sum across text set, Number of text statement occurs in, Percentage of texts statement occurs in, Texts

129.7 Term Distribution List Example

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

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

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

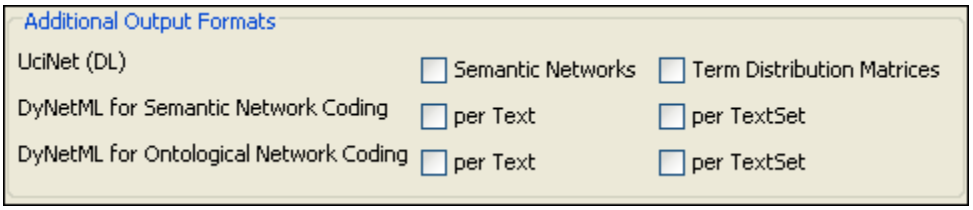
Then AutoMap's Extensive Delete List, Direct Adjacency was applied to both texts. The Delete List was extended by further non-content bearing words that appeared in the sample texts (a an and as at awhile but for from happening he her hers him his i in into it its me mine my nor of or our she so that 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. From the main menu select *Run Analysis => Single Map 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).

129.8 Additional Output Formats



Two additional data formats are offered:

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

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

129.9 UciNet (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
- 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.

129.9.1 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.

129.10 DyNetML

Results of Map Analysis and Sub-Matrix Text Analysis can be output in DyNetML format. The resulting DyNetML files will be stored as NameOfText.xml in the xml folder under the root directory of AutoMap.

129.11 After Map Analysis:

See Concepts not forming edges

129.12 Run Sub-Matrix Text Analysis.

Two options for creating DyNetML files are offered. Either one or both options can be selected per analysis. These are found on *Text Processing Panel => Tab 4. Analysis Settings ⇒Sub-Tab 2. Output Options* under the heading Concepts not forming edges.

1. per Text: The DyNetML files will be stored as NameOfText.xml and in a folder called xml under the root directory of AutoMap.
2. per TextSet: The DyNetML file will be stored as consolidated_map.xml in a folder called xml under the root directory of AutoMap.

129.13 Examples for additional Output Formats

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

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

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

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

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

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

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

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

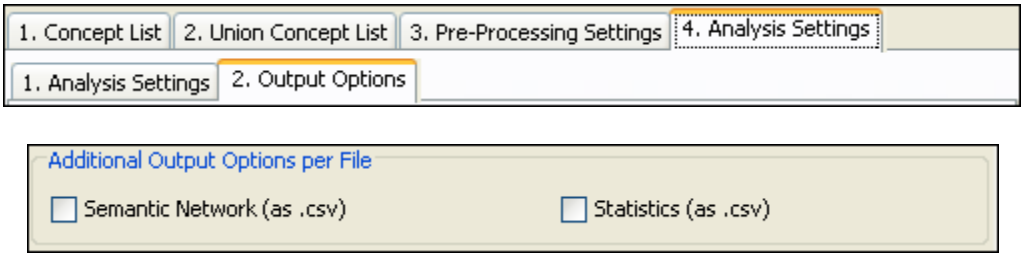
Prof. Darren, Mrs. Brown & Mr. Cray met over dinner discussed how plan had worked. John Darren Mrs. Brown take down scarecrow following week. thought too scary dog. Mr. Craye painted fence. Then Mrs. Brown watered lotus, carnations, daffodils, roses.

Then map analysis was run on both texts using the AutoMap's default settings. All term distribution matrices provided on the Output Options panel were checked. Again, the default settings were used. The screen shot below shows the resulting analysis that should be displayed in *Panel 3 - Semantic network of current Text*.

Semantic network of current Text

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

129.14 Additional Output Options



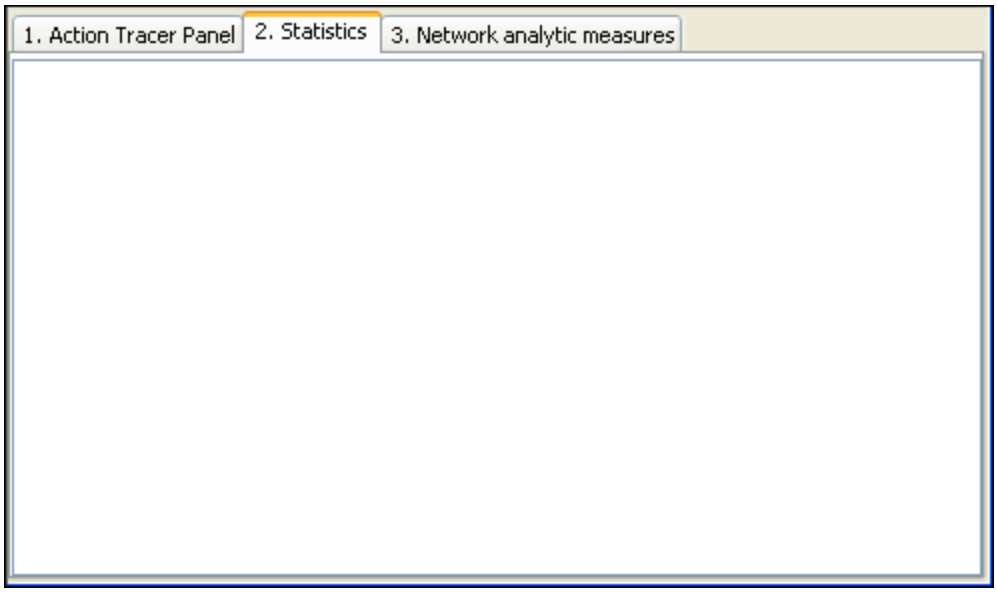
130 Outputs

Outputs for Map Analysis, Meta Network Text Analysis and Sub Matrix Text Analysis are displayed on The *Semantic Network Panel* and the *Results Panel*.

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.

131Statistics



After running an analysis of a file, the Statistics will be displayed in the *Results Panel => Tab 2. Statistics*.

Each stat file generated is automatically saved in a folder named Stat Output in the AutoMap main directory. This output folder contains a map file (nameOfText.map) and a stat file (nameOfText.stat) for each text analyzed as well as (stat_output.xls) file a that contains the stats of all texts.

Entries in the stat output and explanation:

Entry	Entry	Explanation
# of concepts analyzed:	unique: total:	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.
# of concepts in statements:	unique: total:	Only concepts are considered that linked into statements.
# of isolated concepts:	unique: total:	Only concepts are considered that did not link into statements.
# of statements:	unique: total:	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.
density	unique:	Unique density is the density of the resulting

(based on statements)	total:	network based on unique statements, total density respectively is the density of the resulting network based on the total number of statements.
	Punctuation:	The Punctuation option chosen by the user.
Analysis Settings	Window Size:	The Window Size chosen by the user.
	Directionality:	The Directionality option chosen by the user.

132 Support

132.1 How to cite AutoMap

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

132.2 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.

132.3 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.

133References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Popping, R., & Roberts, C.W. (1997). Network Approaches in Text Analysis. In R. Klar & O. Opitz (Eds.), Classification and Knowledge Organization: Proceedings of the 20th annual conference of the Gesellschaft fÃ¼r Klassifikation e.V., (pp. 381-898). University of Freiburg , Berlin , New York : Springer.

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

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

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

133.1 Where to learn more about Dynamic Network Analysis

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

Kathleen M. Carley, 2002, "Smart Agents and Organizations of the Future" The Handbook of New Media. Edited by Leah Lievrouw and Sonia Livingstone, Ch. 12, pp. 206-220, Thousand Oaks, CA, Sage.

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

Entropy references:

Theory:

W. Weaver & C. E. Shannon, The Mathematical Theory of Communication, Urbana, Illinois: University of Illinois Press, 1949

Empiric application:

Diesner, J., Carley, K.M., & Katzmair, H. (2007). The morphology of a breakdown. How the semantics and mechanics of communication networks from an organization in crises relate. XXVII Sunbelt Social Network Conference, Corfu, Greece, May 2007.

134Lessons

135Lesson 1

135.1 From content analysis to semantic networks

135.2 Load multiple text documents into AutoMap

Place all the documents for analysis into a single folder. Then from the main menu select *File => Open multiple files (point to folder)*. Navigate to the folder with the files and select [OK].

The first text will be displayed in the *Text Display Panel => Tab 1. Original Texts*. The arrows in the menubar can be used to view the other texts.

135.3 A simple content analysis

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

Examining the Concept List considers the 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 *Text Processing Panel => Tab 1. Concept List*. The Concept List will display important facts about each document in your set:

- The concepts found in the text displayed in *Text Display Panel => Tab 1. Original Text*
- The frequencies of each concept

135.3.1 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.

135.4 Create and refresh the Union Concept List:

The Union Concept List considers concepts across all texts in a set. It is located in *Text Processing Panel ⇒Tab 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

To view the Union Concept list select form the main menu, *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.

After refreshing your Union Concept List, it can viewed in *Text Processing Panel => Tab 2. Union Concept List*.

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.

135.4.1 NOTE : The number of unique concepts considers each concept only once. The number of total concepts also considers repetitions of concepts.

135.5 Save a Union Concept List

To specify the location to save this file select from the main menu *File => Output Storage Manager*. The Output Storage Manager window will appear. From this window designate the location for the file. Select the Save file as… option to the right of the default file pathway then navigate to the folder to save the file. At the bottom of the window select the [Save Settings] button. Finally, click the Close button at the top of the window.

To save the file select form the main menu *File => Save Union Concept List*.

135.6 View your csv file in Microsoft Excel

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

Open up Excel. From the Open dialog box navigate to the folder with the Union Concept List and select the file.

135.7 Perform a simple Semantic Network Analysis

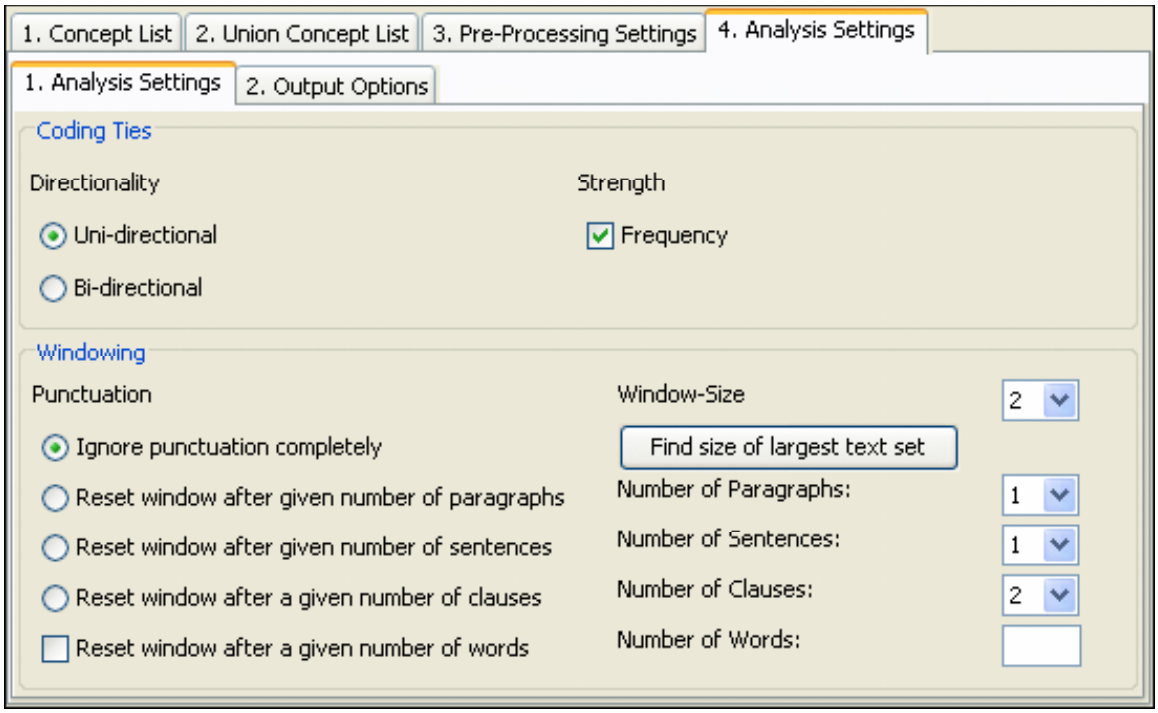
A Simple Network Analysis involves no thesauri and uses all the words. A Semantic Network Analysis can be run either on the original input texts or texts that have been pre-processed with a Delete List and/or a Generalization Thesaurus.

135.8 Before running the analysis make sure you have completed the following steps:

- Performed any necessary Pre-processing steps on the texts
- 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)

135.9 Analysis Settings:

Use the *Text Processing Panel => Tab 4. Analysis Settings to specify the analysis settings. Your settings will be automatically applied to the analysis (you do not need to confirm them). Below are the standard settings found in the Text Processing Panel => Tab 4. Analysis Settings.*



The following chart lists the possible analysis settings:
Coding Ties : Specify the way statements are counted.

Directionality : Select one of the following two possibilities by checking the button.

Uni-Directional : (When coding a tie, only 1st->2nd concept should be noted)

Bi-Directional : (When coding a tie, both 1st <-> 2nd concept shall be noted)

Strength : Strength will be printed out by default. To not print out Strength uncheck the item.

Frequency : (The cumulative frequency of every existing statement.)

Item not checked : Existence of frequency will be printed out (binary result).

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

Punctuation : Select one option by checking the radio button.

The radio buttons [Ignore punctuation completely and the Reset window after given number of / paragraphs / sentences / clauses] : allow for selecting the size of the text for analysis.

Reset window after given number of words checkbox can be used with any of the above options.

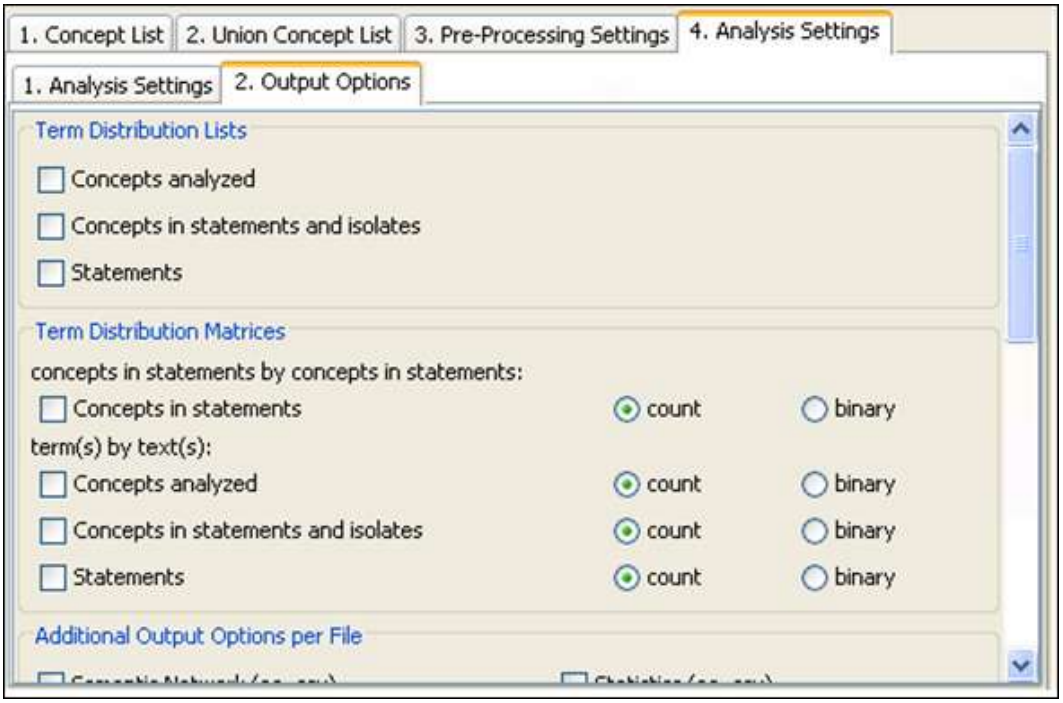
Window Size : Select one window size by using the number chooser.

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.

135.10 Output Options:

The map and the statistic output generated by AutoMap are displayed in the *Semantic Network Panel and the Results Panel*, respectively. In addition, AutoMap offers further output options that can be selected in the *Text Processing Panel => Tab 4. Analysis Settings => SubTab 2. Output Options*. Any other additional outputs are generated after analyses are run.



For all types of multiple analysis, term distribution lists and matrices can be selected as output options in the upper two fields of the *Text Processing Panel => Tab 4. Analysis Settings => SubTab 2. Output Options*.

135.11 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:

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

Concepts analyzed	Matrix of concept analyzed.csv	Matrix of Concept (union of concepts listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept
Concepts in statements and isolates	Matrix of concept in statements.csv	Matrix of Concept (union of concepts listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of concept If binary was chosen, cells denote existence (1) or absence (0) of concept
Statements	Matrix of statements.csv	Matrix of Statements (union of statements listed in first row) by texts (all text names listed in first column) If count was chosen, cells contain cumulated frequency of statement If binary was chosen, cells denote existence (1) or absence (0) of statement

135.12 Save your analysis in DyNetML format:

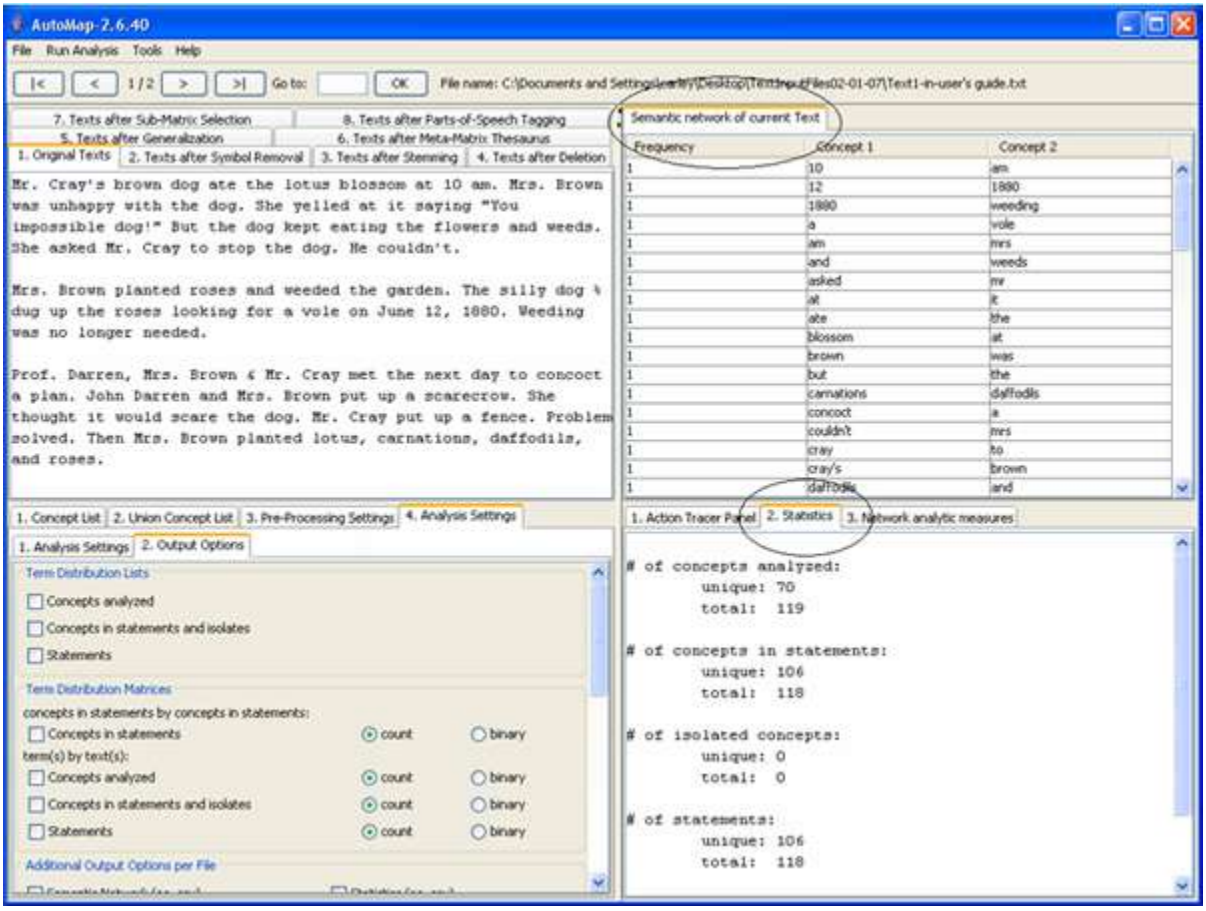
Before running a Multiple Map Analysis, it must be specified that the results will be in DyNetML format so that the results can be read into ORA.

In the Text Processing Panel => Tab 4. Analysis Settings => SubTab 2. Output Options, scroll down to the option Additional Output Formats and place a checkmark in the two checkboxes per Text / per TextSet next to either Semantic Network Coding or Ontological network Coding.

From the main menu select *File => Output Storage Manager*. Locate Text Analysis Output Directory. select 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. Click the [Open] button. Save these settings and close the Output Storage Manager.

135.13 Run the Multiple Map Analysis:

From the main menu select *Run Analysis => Multiple Map Analysis*. The results will be displayed on the *Semantic Network Panel* and on the statistics in the *Results Panel => Tab 2. Statistics*. Below is a screen capture showing the results of Multiple Map Analysis in the *Semantic Network Panel and the Results Panel*



If multiple texts have been analyzed multiple texts, the results can be browse through using the arrows in on the menubar. Results of multiple analyses are automatically saved in a folder titled output under the directory where Automap is installed.

135.13.1 NOTE : The output folders are overwritten with every new analysis run. If the results need to be saved either rename that folder or move the files.

136 ORA

136.1 Loading files into ORA

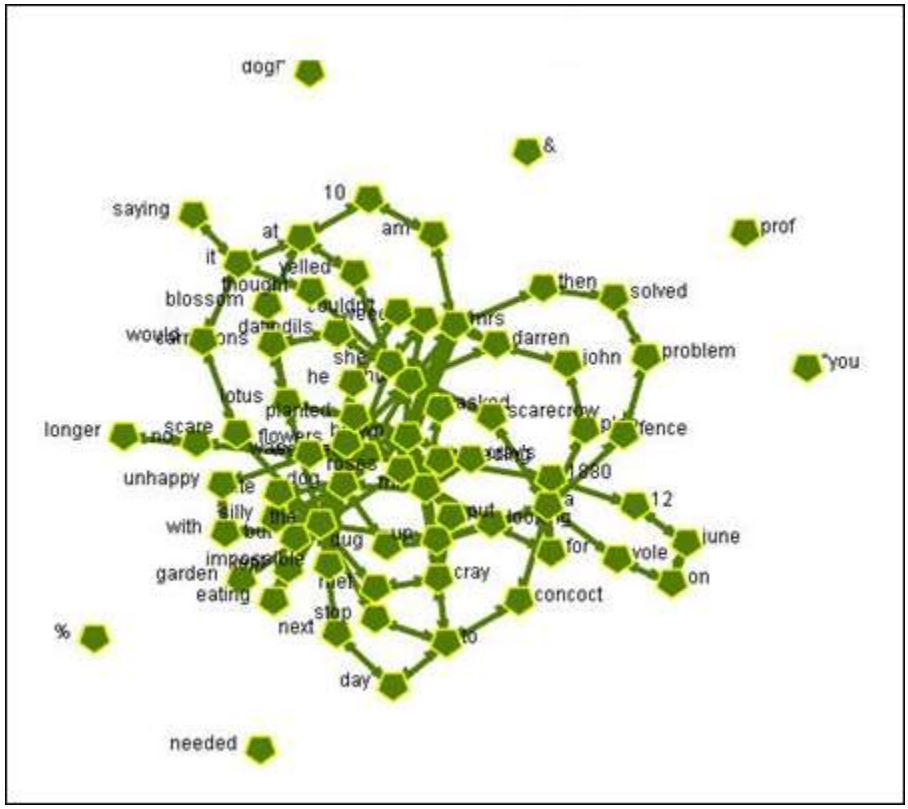
Use either the Load button of the menu item *File => Open File*. Navigate to the directory with the files created in Automap. Select a file(2) and press [OK]. Selecting multiple files can be done by holding down the [CTRL] key while clicking on each file..

136.1.1 NOTE : If there are already file(s) loaded into ORA, choose the Append As Additional Meta Network option.

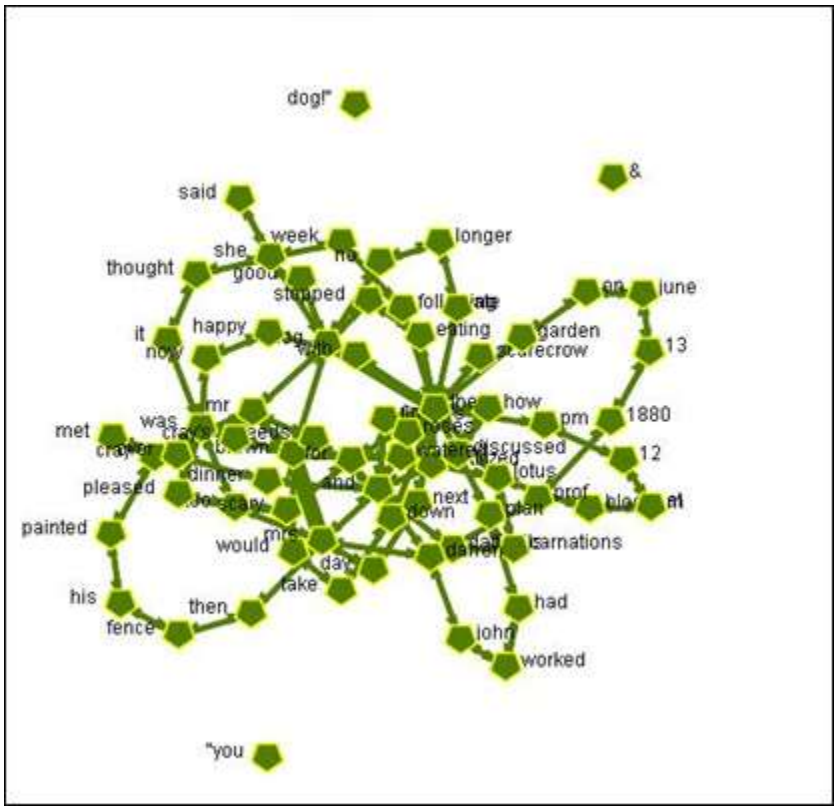
136.2 Visualize your meta-matrices

IN the left panel of ORA select the newly loaded Meta-Network. Then select the Visualize this Meta-Network.

Text1-in-user's guide.txt

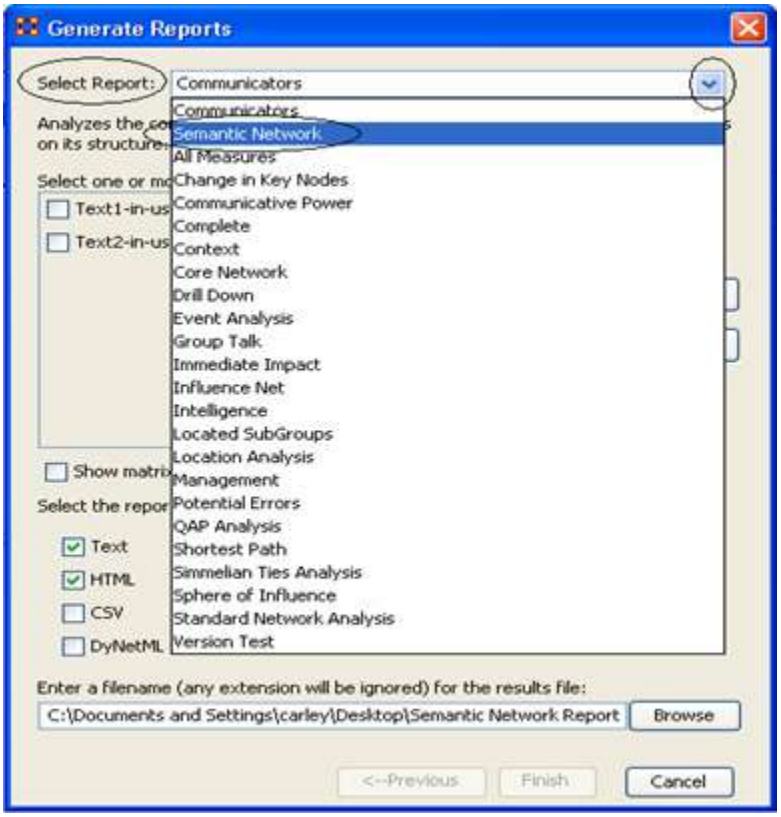


Text2-in-user's guide.txt



136.3 Run a Semantic Network Report

From ORA's main menu select *Analysis => Generate Reports*. The Select Report dialog box will appear. Click the dropdown menu and select Semantic Network from the list.



In the box titled Select one or more meta matrices place a checkmark next to the files to visualize. To specify a filename for the results, select the [Browse] button near the bottom right corner. Navigate to the desired directory and type in a filename. Select the [Next] button, then the [Finish]. ORA will run the report and an HTML file will be created to run in the browser.

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

136.4 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.

Network	Concepts	Statements	Density
Text1-in-user's guide.txt	74	200	0.0370233

Text2-in-user's guide.txt	69	184	0.0392157
Union	98	310	0.032611
Intersection	45	8	0.0040404

136.5 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.

Network	Concepts	Statements
Text1-in-user's guide.txt to Text2-in-user's guide.txt	29	192
Text2-in-user's guide.txt to Text1-in-user's guide.txt	24	176

137

138Lesson 2

138.1 Using the Delete List

Load a text file(s) into Automap. The file (or first file with multiple texts) will be displayed in *Text Display Panel => Tab 1. Original Text*. If multiple texts are loaded they can be viewed using the left and right arrows on the menubar.

138.2 Basics of delete lists

Deletion removes non-content bearing conjunctions and articles from texts. These are denoted in the Delete List and removed in pre-processing.

- 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

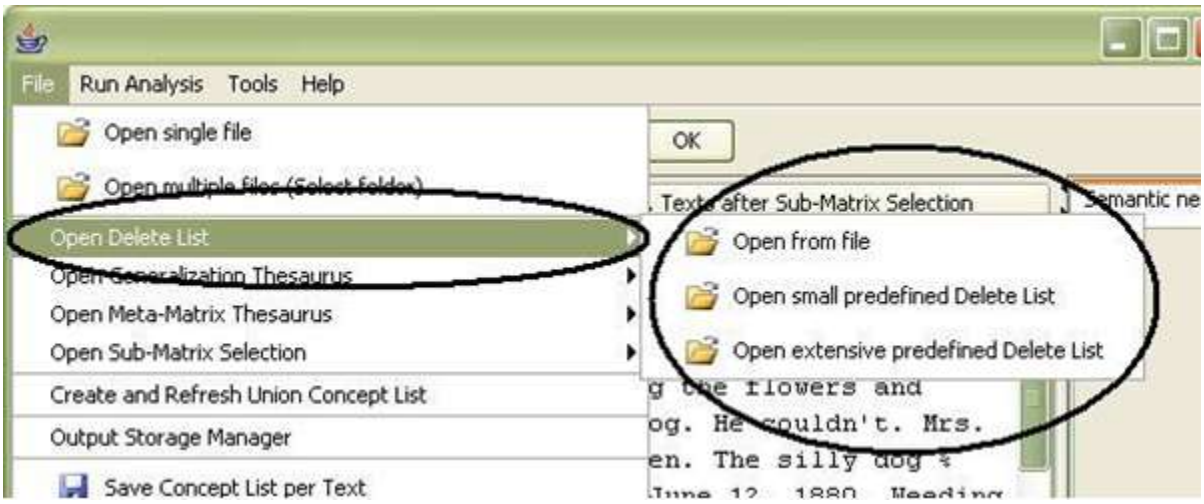
The Delete List is NOT case sensitive.
You can use the predefined Delete Lists that AutoMap offers, or create your own Delete List.
All Delete Lists can be edited.

138.3 Open a Delete List

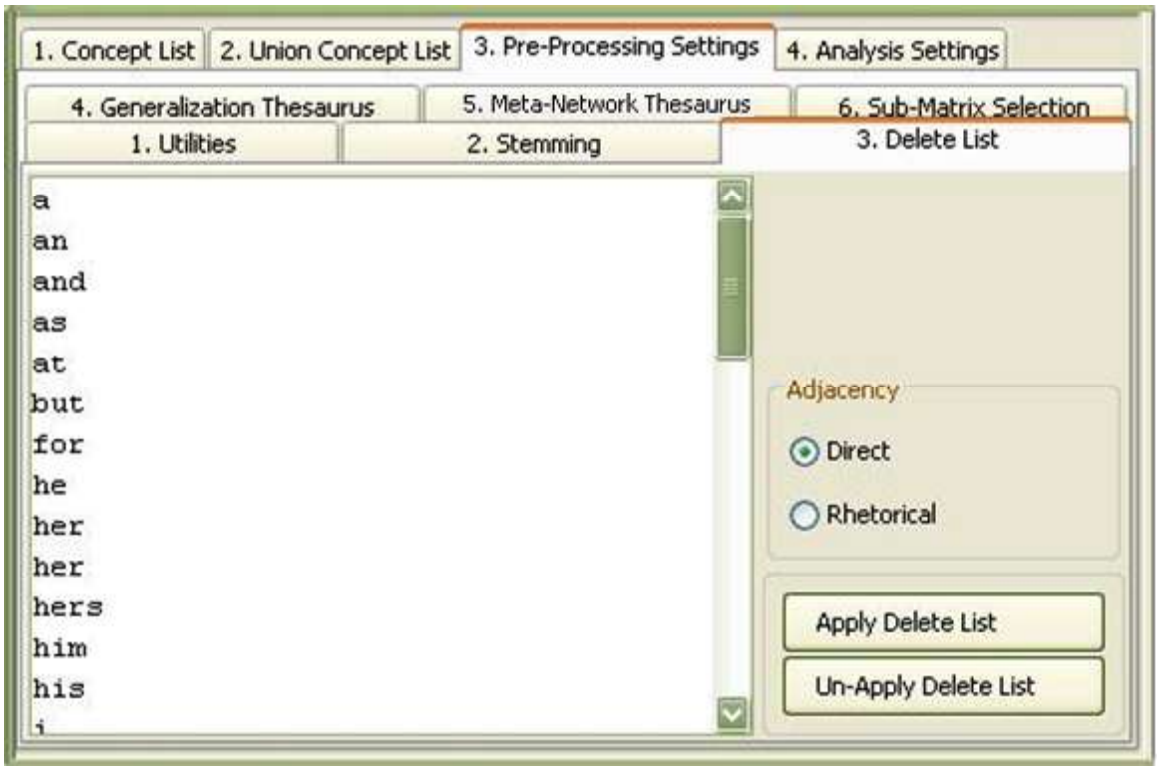
From the main menu select the *File => Open Delete List*. Choose one of the following options:

- Open from file: A file chooser appears. Select a previously created delete list
- 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 the *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 3. Delete List.*



The Delete List can be edited from within *Text Processing Panel => Tab 3. Delete List*.
The types of Delete Lists available in AutoMap are:
Small predefined Delete List

a, an, and, some, many, this, that, these, those, the, all, one, every

Extensive predefined Delete List

a, an, and, as, at, but, for, he, her, hers, him, his, i, it, its, me, mine, my, nor, of, or, our, she, so, that, the, their, theirs, them, they, to, us, we, who, whoever, whom, whomever, will, would, you, your, yours, yourself

138.3.1 NOTE : If you wish to apply a Delete List and a Thesaurus, it is recommend to apply the Delete List first, followed by the Thesaurus.

When applying a delete list, AutoMap:

- Searches the text file(s) for concepts specified in the Delete List
- Deletes matching concepts from the texts
- Displays the resulting texts in *Text Processing Panel => Tab 4. Texts After Deletion*

Steps to apply a Delete List

1. Choose an adjacency option in Text Processing Panel => Tab 3. Pre-Processing => SubTab 3. Delete List. There are types of Adjacency:
    - Direct Adjacency (default) : removes unspecified content which results in , 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.
2. Click the Apply Delete List button to remove concepts form all currently loaded text
3. The newly pre-processed texts can be viewed in *Text Display Panel => Tab 4. Texts After Deletion*

138.4 Multiple Delete Lists

Multiple delete lists can be applied to the same text by loading, and applying the first list then loading, and applying a subsequent Delete List. Any number of Delete Lists can be applied in this manner.

138.5 Un-apply a Delete List

To un-apply a Delete List that was just applied to the data, go to the *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 3. Delete List* and press the [Un-Apply Delete List] button. The Delete List in Panel 2 will be cleared.

138.6 Save an applied Delete List

To specify a location for your Delete List file, choose *File => Output Storage Manager*. Click the Save file as… button to the right of the default file pathway. In the file chooser window, navigate to the directory to save the file. Press the [Save Settings] button at the bottom of the Output Storage Manager then close the Output Storage Manager window.

From the Automap main menu select *File => Save Delete List*. The Delete List will be saved to the directory specified in the Output Storage Manager.

138.7 Save text(s) after Delete List

From the main menu select *File => Output Storage Manager*. Click the Save file as… button to the right of the default file pathway for Texts After Deletion. In the file chooser window, navigate to the directory to save the file. Press the [Save Settings] button at the bottom of the Output Storage Manager then close the Output Storage Manager window.

From the Automap main menu select *File => Save Text(s) after Delete List Applied*. The Delete List will be saved to the directory specified in the Output Storage Manager.

138.8 Modify a Delete List

In *Text Processing Panel => Tab 3. Pre-Processing Settings => 3. Delete List* A loaded Delete List can be:

- Add concepts : Create an empty line and type in a new concept
- Modify concepts : Go to the desired line and edit a concept
- Drop concepts : Highlight the concept and press the <Delete> key

138.9 Create a delete list from scratch in AutoMap

There are two ways to create a new delete list:

Within AutoMap

In *Text Processing Panel => Tab 3. Pre-Processing Settings => 3. Delete List* click inside the empty field. Type on concept per line and press [Enter] when done to start a new concept. Do not leave and Empty Lines. Deletes Lists are not case-sensitive. When finished press the [Apply Delete List] button.

Outside of AutoMap

Open a text editor capable of producing standard text files. Type on concept per line and press [Enter] when done to start a new concept. Do not leave and Empty Lines. Deletes Lists are not case-sensitive. Save the file as a text file. From the Automap main menu select *File => Open Delete List => Open from file*.

Any of these Delete Lists can be saved for future use.

138.9.1 NOTE : Delete Lists can be opened in Excel from the Open menu item.

138.10 Modify a delete list in Microsoft Excel

The same rules for editing Delete Lists in a text Editor apply to editing them in Excel.

139 Lesson 3

139.1 Using a Generalization Thesaurus

139.2 Load text document(s) into AutoMap

The Generalization Thesaurus can be used on one or more files simultaneously.

From the main menu select either *File ⇒ Open single file* or *File => Open multiple files*. Navigate to the directory and chose the file(s) to load. The text will be displayed in *Text Display Panel => Tab 1. Original Texts*. If multiple files were loaded use the arrow keys on the menubar to view the other texts.

139.3 Pre-Processing

After text(s) are loaded any pro-processing can be applied such as Delete List, Named-Entity Recognition, Symbol Removal, and Stemming.

139.3.1 NOTE : If you wish to apply both a Delete List and a Generalization Thesaurus, be sure to use first the Delete List before the Generalization Thesaurus.

139.4 Open a Generalization Thesaurus

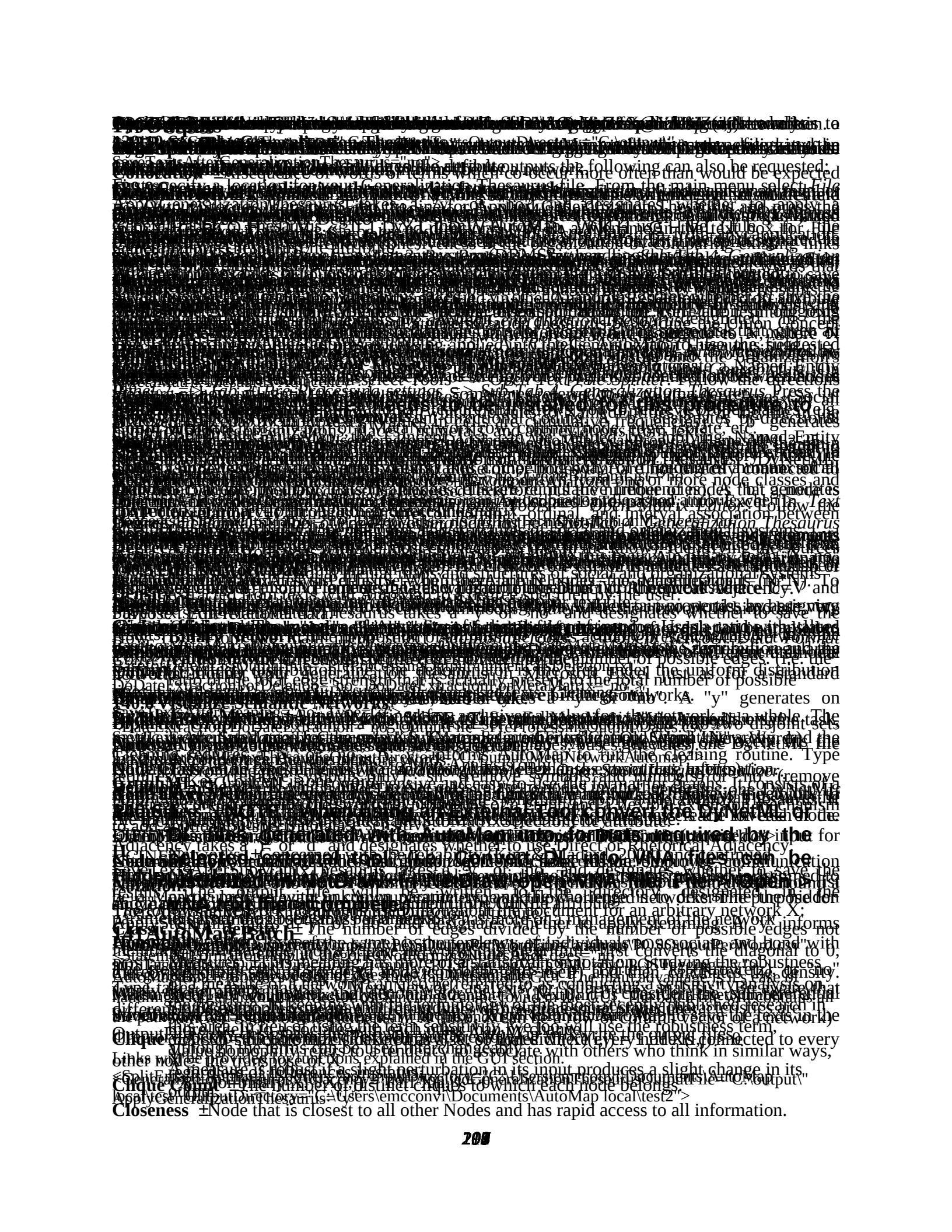
From the Menu bar select *File => Open Generalization Thesaurus*. The thesaurus will be displayed in *Text Processing Panel => Tab 3. Preprocessing Settings => SubTab 4. Generalization Thesaurus*.

[illegible]

[illegible] outputs the following can also be requested:

[illegible] Thesaurus file. From the main menu select *File* [illegible]

[illegible] Thesaurus. Press the [illegible] generates [illegible] Entity [illegible] or more node classes and [illegible] *Text* [illegible] ordinal, and interval association between [illegible] *Generalization Thesaurus* [illegible] Thesaurus takes a "y" or "no" and designates whether to save the [illegible] of possible edges. I.e. the [illegible] the total number of possible [illegible] A "y" generates on [illegible] two disjoint sets [illegible] Type [illegible] (remove [illegible] Adjacency takes a "y" or "no" and designates whether to use Direct or Rhetorical Adjacency. [illegible] Thesaurus takes a "y" or "no" and designates whether to save the [illegible] for an arbitrary network X: [illegible] The number of edges divided by the number of possible edges not [illegible] Thesaurus. Press the [illegible] converts the diagonal to 0, [illegible] Studying the robustness [illegible] density." [illegible] use the robustness term, [illegible] connected to every [illegible] with others who think in similar ways, [illegible] explained in the GUI section. [illegible] in its input produces a slight change in its [illegible] distinct cliques to which each node belongs. [illegible] Apply Generalization Thesaurus = "Users\emcconvi\Documents\AutoMap local\test2" >



Closeness ±Node that is closest to all other Nodes and has rapid access to all information.

4. Be sure to separate the words by columns.
5. The Thesaurus is NOT case sensitive.

139.9 Create a generalization thesaurus list from scratch

A Generalization Thesaurus can be created from scratch in AutoMap.

Got to *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 4. Generalization Thesaurus*. In the text field there are two columns titled concept and generalization. To edit items inside this text field, 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.

From the main menu select *File => Create and Refresh and Refresh Union Concept List*. On *Panel 2 => Tab 3. Pre-Processing settings => Sub-Tab 4. Generalization Thesaurus* Press the [Load Union Concept List] button

Once loaded into AutoMap, the Concept List can be refined by applying Named-Entity Recognition and Deletion Lists prior to application of the Generalization Thesaurus.

139.10 Save the Generalization Thesaurus

First specify a location for your Generalization Thesaurus file. From the main menu select *File => Output Storage Manager*. Find the line for Applied Generalization Thesaurus and press the Save file as …. Navigate to the desired directory, type in a filename in the text box for File name: and press [Open].

After returning to the Output Storage Manager window press the [Save Settings] button to save the pathways. Then close the window.

From the main Menu select *File => Save applied Generalization Thesaurus*.

139.11 Open the Generalization Thesaurus in Microsoft Excel

139.11.1 NOTE : Excel needs to be installed for this procedure to work.

Open up Excel. From the main menu select *File => Open*. Navigate to the directory where the Generalization Thesaurus was saved and open it.

Alternatively, the Generalization Thesaurus can be copied and pasted into Excel. In *Text Processing Panel => Tab 3. Pre-Processing Settings => Sub-Tab 4. Generalization Thesaurus* place the cursor in the text field. Press [Control] and [A] simultaneously to select all the text. Then copy the text with [Control] and [C]. In Excel place the cursor in a blank Excel spreadsheet and press [Control] and [V] to paste the generalization thesaurus into the spreadsheet.

139.12 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.

140 Outputs

140.1 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 : From the main menu select *Tools => Open Text File Splitter*. Follow the directions specified in the user interface.

140.2 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 : From the main menu select *Additional Tools => Open CompareMap*.

140.3 Merge DyNetML Files

Purpose : Merge multiple DyNetML files into 1 DyNetML file.
Example:
This might be needed for example when DyNetML files that was generated per text during Sub-Matrix Text Analysis need to be consolidated into 1 DyNetML file that represent the entire text set.
Output : One DyNetML file
How To : From the main menu select *Additional Tools => Open DyNetML File Merger*. Follow the directions specified in the user interface.

140.4 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 : From the main menu select *Additional Tools => Open Matrix Editor*. Follow the directions specified in the user interface.

140.5 Edit Network Data

Purpose : Edit relational data.
How To : From the main menu select *Additional Tools => Open Network Data Format Converter*. Follow the directions specified in the user interface.

140.6 Visualize Semantic Networks

Purpose : Visualize mental models and social structure.
How To : From the main menu select *Additional Tools => Open SocialInsight Visualizer*.

140.6.1 NOTE : When using an External Tool convert the DyNetML or DL files generated with AutoMap into formats required by the selected external tool. (i.e. Convert DL into VNA files can be visualized in NetDraw. In NetDraw open VNA file: File > Open > VNA text file > complete.)

141 AutoMap Batch

AutoMap Batch provides a way to use AutoMap without the GUI. The main advantage is it's ease of handling a larger amount of files. Most functions can be found in the GUI. The batch file will contains all the pathways and toggles settings necessary for the end user.

Links will be provided for functions explained in the GUI section.

141.1 Things to Remember

141.1.1 NOTE : Any pathway you denote in the XML file MUST exist or it will not be created. Create any directories you need before starting.

141.1.2 SUGGESTION : Create a parent directory into which all of input and output files can be placed. This will make it easier to find any file.

141.2 Running AutoMap Batch

141.2.1 Open up the file AutoMap_XML.xml

Modify the AutoMap_XML.xml which resides in the home directory where AutoMap is installed. This it typically in the Program Files in Windows or the Home Folder in Linux or the Application Folder (OS X).

Open Command Prompt (Windows) or Terminal (Linux).

Change directory to AutoMap home directory

Run AutoMap_script.bat or AutoMap_script.sh (Linux).

142 The AutoMap_script.XML.xml file.

Following is a description of each line of the XML file along with the values possible. Where available there is a link back to the GUI section describing the function.

142.1 Input

<InputText inputDirectory="C:\Users*{name_of_user}*\Documents\AutoMap local\Prop">

This will contain the full pathway to the text files to use. All text files to be analyzed must reside in this folder.

142.2 Utilities

<TextConceptList CreateTextWiseConceptList="no" OutputDirectory="Output">

CreateTextWiseConceptList takes a "y" or "no" and designates whether or not to create an individual concept list for each loaded text.

If "y" is used then the pathway for these concept lists needs designated in OutputDirectory. If "no" is used then AutoMap ignores the pathway.

<UnionConceptList ContextListFile="PreProcessingOutput\UnionConceptsInContext.csv" CreateUnionConceptContextList="no" CreateUnionConceptList="no" OutputFile="C:/test1/ProcessedData/UnionConcept/ucl.csv">

ContextListFile is the pathway to the file containing a Context List.

CreateUnionConceptList takes a "y" or "no" and designates whether or not to create an individual concept list in context.

CreateUnionConceptContextList takes a "y" or "no" and designates whether or not to create a concept list in context.

If "y" in CreateUnionConceptContextList then the pathway for this lists needs designated in OutputFile. If "no" is used then AutoMap ignores the pathway.

<NamedEntitiesExtraction BaseforGeneralizationThesaurus="no" CreateNamedEntitiesList="no" OutputFile="PreProcessingOutput\NamedEntities.csv">

CreateNamedEntitiesList takes a "y" or "no" and designates whether to create a Named Entity file.

A "y" will create a Named Entities List (which defaults to NamedEntities.csv) in the file in the OutputFile attribute.

<N-GramExtraction BaseforGeneralizationThesaurus="no" CreateBiGramList="no" OutputFile="PreProcessingOutput\bigram.csv">

CreateBiGramList takes a "y" or "no" and designates whether to create bigram file.

A "y" creates a BiGram list from all loaded text. The pathway to the file is specified in OutputFile attribute.

<NumericalsExtraction
BaseForMetaMatrixOutputFile="PreProcessingOutput\numericalsMetaMatrix.csv"
BaseforDeleteListOutputFile="PreProcessingOutput\numericalsDeletion.txt"
DoDateExtractionForDeletion="no" DoDateExtractionForMetaMatrix="no">

<DateExtraction DoDateExtraction="no" OutputFile="PreProcessingOutput\Dates.csv">

DoDateExtraction takes a "y" or "no" and designates whether to extract date information.

A "y" creates a file in the directory specified in the OutputFile attribute.

<TFIDFComputation OutputFile="PreProcessingOutput\TFIDF.csv" TFIDFComputation="no">

TFIDFComputation takes a "y" or "no" and designates whether the TFIDF file is created.

A "y" creates a file in the directory specified in the Outfile attribute.

<POSTagging POSOutputPerWord="y" POSOutputPerWordFile="c:\input\POSOutputPerWord.csv" POSTaggedTextsOutputDirectory="c:\input\" POSTaggingWithXML="y" POSTagging="y">

POSTagging takes a "y" or "no" and designates whether the Parts of Speech file will be written.
A "y" creates a file in the directory specified in the POSTaggingTextsOutputDirectory attribute containing the POS Tags for the texts.

<AnaphoraResolution AnaphoraResOutputDirectory="c:\input\" DoAnaphoraRes="y"
UseResolutionAsIs="y">

DoAnaphoraRes takes a "y" or "no" and designates whether the Anaphora will be created.
A "y" creates a file in the directory specified in the AnaphoraResOutputDirector attribute containing the POS Tags for the texts.
The UseResolutionAsIs requires a "y" or "no". A "y" tells AutoMap to use the suggested resolutions without editing, a "no" allows for the manual editing of resolutions.

142.3 PreProcessing

<Deletion Adjacency="r" AppliedDeleteListOutputFile="PreProcessingOutput" ApplyDeleteList="no"
DeleteListFile="" SaveTextsAfterDelition="no"
TextsAfterDeletionOutputDirectory="PreProcessedTexts">

ApplyDeleteList takes a "y" or "no" and designates whether to apply a delete list.
A "y" will apply the delete list specified in the DeleteListFile attribute. A pathway and filename need to be specified.
Deletion Adjacency requires a "r" (rhetorical) or "d" (direct).
SaveTextsAfterDelition takes a "y" or "no". A "y" will save the texts after deletion is applied to the directory specified in the TextsAfterDeletionOutputDirectory attribute.

<Cleaning CleanedTextsOutputDirectory="PreProcessedTexts" Cleaning="y"
SaveTextsAfterCleaning="no" Type="s">

Cleaning requires a "y" or "no". A "y" informs AutoMap to run the cleaning routine. Type requires a "s" (remove symbols only), "sn" (remove symbols and numbers) or "no" (remove nothing). The SaveTextsAfterCleaning requires a "y" or "no". A "y" tells AutoMap to save the file in the directory specified by the CleanedTextsOutputDirectory attribute.

<Stemming ApplyStemming="y" KStemCapitalization="no" SaveTextsAfterStemming="no"
StemmedTextsOutputDirectory="PreProcessedTexts" Type="k">

ApplyStemming takes a "y" or "no" and designates whether to apply stemming. A "y" informs AutoMap to apply a stemmer.
The stemmer selected is specified in Type which takes a "p" (Porter), "k" (Krovretz), or "no" (apply no stemmer).
SaveTextsAfterStemming requires a "y" or "no". A "y" informs AutoMap to save the texts in the directory specified by StemmedTextsOutputDirectory.

<GeneralizationThesaurus Adjacency="no" AppliedGeneralizationThesaurusOutputFile="C:\output\"
ApplyGeneralizationThesaurus="y"

GeneralizationThesaurusFile="C:\Users\emcconvi\Documents\AutoMap local\Prop\CombinedThesari.csv" GeneralizedTextsOutputDirectory="C:\output\" SaveTextsAfterGeneralizationThesaurus="no">

ApplyGeneralizationThesaurus takes a "y" or "no" and designates whether to apply a Generalization Thesaurus. If "y" then AutoMap will use the file in the GeneralizationThesaurusFile attribute.

Adjacency takes a "d" or "r" and designates whether to use Direct or Rhetorical adjacency.

SaveTextsAfterGeneralizationThesaurus takes a "y" or "no" and designates whether to save the results. The output file will be written to the directory designated in the GeneralizedTextsOutputDirectory attribute.

<MetaMatrixThesaurus Adjacency="r" AppliedMetaMatrixThesaurusOutputFile="C:\output\metathes.csv" ApplyMetaMatrixThesaurus="y" MMOraFile="" MapToOraTypes="y" MetaMatrixThesaurusFile="C:\Users\emcconvi\Documents\AutoMap local\Prop\CT_HamasMM_2-25-08.csv" SaveTextsAfterMetaMatrixThesaurus="no" TextsAfterMetaMatrixThesaurusOutputDirectory="C:\output\" categories="knowledge,agent,resource,task,event,organization,location,role,action,attribute,when">

ApplyMetaMatrixThesaurus takes a "y" or "no" and designates whether to apply a MetaMatrix Thesaurus. If "y" then AutoMap will use the file in the MetaMatrixThesaurusFile attribute.

Adjacency takes a "r" or "d" and designates whether to use Direct or Rhetorical Adjacency.

SaveTextsAfterMetaMatrixThesaurus takes a "y" or "no" and designates whether to save the results. The output file will be written to the directory designated in the TextsAfterMetaMatrixThesaurusOutputDirectory attribute.

<SubMatrixSelection Adjacency="r" ApplySubMatrix="y" FullMetaMatrixFile="C:\input\SubMatrix\Submat.txt" SaveFullMetaMatrix="y" SaveTextsAfterSubMatrixThesaurus="no" SubMatrixSelectionFile="C:\input\Submat.txt" TextsAfterSubMatrixThesaurusOutputDirectory="C:\output\MetaNetworkAutomap">

ApplySubMatrix takes a "y" or "no" and designates whether to apply a MetaMatrix Thesaurus. If "y" then AutoMap will use the file in the SubMatrixSelectionFile attribute.

Adjacency takes a "r" or "d" and designates whether to use Direct or Rhetorical Adjacency.

SaveTextsAfterSubMatrixThesaurus takes a "y" or "no" and designates whether to save the results. The output file will be written to the directory designated in the TextsAfterSubMatrixThesaurusOutputDirectory attribute.

<StatementFormation ApplyCategoriesNotFormingNodesAndEdges="no" CategoriesNotFormingNodesAndEdgesFile="" Directionality="u" MeasurementTypForSubMatrixAnalysis="nta" Strenght="y" TextUnit="s" TextUnitResetNumber="2" WindowSize="5" ResetAfterGivenWordNumber="y" WordResetNumber="46">

ApplyCategoriesNotFormingNodesAndEdges takes a "y" or "no" and designates whether to apply a MetaMatrix Thesaurus. If "y" then AutoMap will use the file in the CategoriesNotFormingNodesAndEdgesFile attribute.

TextUnit takes a "t" (for coding across the entire text), "p" (for resetting window after set number of paragraphs), "s" (for coding links in a given number of sentences), or "c" (for links a given number of clauses). WindowSize takes a number.

TextUnitResetNumber takes a number and is used if contains a p, s, or c. Use this attribute to set the number of paragraphs, sentences, of clauses before the window resets.

Strength takes a "y" or "no". A "y" designates coding cumulative frequency of statements. A "no" designates a binary link coding.

Directionality takes a "u" or "b". A "u" designates uni-directional coding. A "b" designates bi-directional coding.

MeasurementTypForSubMatrixAnalysis takes a "nta" (network Text Analysis) or "sna" (Social Network Analysis(. A "u" designates uni-directional coding. A "b" designates bi-directional coding.

<Output ConceptByConceptInStatementMatrix="no" ConceptsAnalyzed="no"
ConceptsAnalyzedByTexts="no" ConceptsInStatementsAndIsolatesByTexts="no"
ConceptsInStatementsMatrix="no" DlPerMaps="no" DlPerTermDistributionMatrices="no"
DoPOSExtraction="y" DynetmlForSemanticNetworkAnalysisPerText="y"
DynetmlForSemanticNetworkAnalysisPerTextSet="no" DynetmlForSubMatrixAnalysisPerText="y"
DynetmlForSubMatrixAnalysisPerTextSet="no" EdgeAttributeExtract="no" MapFileOutput="no"
StatFileOutput="no" Statements="no" StatementsByTexts="no">

For each text analyzed a map file and state file are generated. For each analysis a Stats_overall.csv file is generated. The outputs are written to the directory designated in OutputDirectory attribute. In addition to the default outputs the following can also be requested:

ConceptsAnalyzed takes a "y" or "no". A "y" will generate statistics on all concepts analyzed.

ConceptsInStatementsMatrix takes a "y" or "no". A "y" generates statistics on all concepts linked into statements.

Statements takes a "y" or "no". A "y" generates statistics on all identified statements.

EdgeAttributeExtract takes a "y" or "no". A "y" extracts edge attributes if two nodes share the same attribute and attaches the link as an edge attribute.

ConceptByConceptInStatementMatrix takes a "c", "b", "no". A "c" generates a matrix of concepts by concepts. A "b" generates the same matrix only in binary form. A "no" generates no matrix

ConceptByConceptInStatementMatrix takes a "c", "b", "no". A "c" generates a matrix of all analyzed concepts by input texts (Values in cells are cumulative frequencies). A "b" generates this matrix in binary form. A "no" generates no matrix

ConceptsInStatementsAndIsolatesByTexts takes a "c", "b", "no". A "c" generates a matrix of all analyzed concepts by input texts (Values in cells are cumulative frequencies). A "b" generates this matrix in binary form. A "no" generates no matrix

StatementsByTexts takes a "c", "b", "no". A "c" generates a matrix of statements by statements (Values in cells are cumulative frequencies). A "b" generates this matrix in binary form. A "no" generates no matrix

MapFileOutput takes a "y" or "no". A "y" stores one map per text file.

StatFileOutput takes a "y" or "no". A "y" Stores one stat file per text.

DynetmlForSemanticNetworkAnalysisPerText takes a "y" or "no". A "y" generates one DyNetML file per text.

DynetmlForSemanticNetworkAnalysisPerTextSet takes a "y" or "no". A "y" generates on DyNetML file that represents all texts. NOTE : This output applies to Map Analysis only!

DynetmlForSubMatrixAnalysisPerText takes a "y" or "no". A "y" generates one DyNetML file per text. NOTE : This output applies to Map Analysis only!

DynetmlForSubMatrixAnalysisPerTextSet takes a "y" or "no". A "y" generates one DyNetML file per text. NOTE : This output applies to Map Analysis only!

DlPerMaps takes a "y" or "no". A "y" generates output maps in DL format (required as input for UCINET).

DlPerTermDistributionMatrices takes a "y" or "no". A "y" generates output term distribution list and matrices in DK format. (required as input for UCINET)

<Analysis OutputDirectory="C:\input\dynetml" Type="sub matrix analysis">

Type takes a "map analysis", "Meta Network analysis" or "sub matrix analysis". Be aware that different types of analysis require different types of pre-processing procedures.

OutputDirectory designates the pathway where AutoMap will write the output files.

<SplitEmails RunEmailSplitter="no" InputDirectory="C:\Users\emcconvi\Documents\AutoMap local\test" OutputDirectory="C:\Users\emcconvi\Documents\AutoMap local\test2">

RunEmailSplitter takes a "y" or "no". A "y" will run the email splitter.

InputDirectory indicates the directory in which the emails are stored.

OutputDirectory indicates the directory where outputs will be written.

<BeliefEngine InputDirectory="C:\output" MergeListFile="C:\input\stuff\mergelistkk.txt" OutputDirectory="C:\output\" RunBeliefEngine="no">

RunBeliefEngine takes a "y" or "no". A "y" runs the belief post processor.

InputDirectory designates the directory for the XML files. If this attribute is empty the Belief engine will run on the files in Analysis OutputDirectory

MergeListFile designates the pathway to the merge file list. NOTE : This file should be a text file.

OutputDirectory designates the pathway for the XML output files. If this is empty the files will be stored in the Analysis OutputDirectory.

<PropertyReplacement InputDirectory="C:/input/dynetml" ListToConvert="_specific,_general" OutputDirectories="C:/input/dynetml/1,C:/input/dynetml/2" RunPropertyReplacement="y">

RunPropertyReplacement takes a "y" or "no". A "y" runs the Property Replacement.

InputDirectory designates the pathway for the XML output files. If this is empty the files will be default to Analysis OutputDirectory.

OutputDirectories designates the pathway for the output files. If there are multiple phases to convert then enter multiple separated with a comma These will be run sequentially such that the last folder will contain files with every phase listed converted to an attribute.

ListToConvert lists the phrases to convert to attributes. More than one attribute can be listed and is separated with a comma, and the final xml files will have all phrases converted.

143Basic Terms

A

Adjacency Network – A Network that is a square actor-by-actor (i=j) network where the presence of pair wise edges are recorded as elements. The main diagonal, or self-tie of an adjacency network is often ignored in network analysis.

Aggregation – Combining statistics from different nodes to higher nodes.

Algorithm – A finite list of well-defined instructions for accomplishing some task that, given an initial state, will terminate in a defined end-state.

Attribute – Indicates the presence, absence, or strength of a particular connection between nodes in a Network.

B

Betweenness – Degree an individual lies between other individuals in the network; the extent to which a node is directly connected only to those other nodes that are not directly connected to each other; an intermediary; liaisons; bridges. Therefore, it's the number of nodes that a node is connected to indirectly through their direct links.

Betweenness Centrality – High in betweenness but not degree centrality. This node connects disconnected groups, like a Go-between.

>

Bigrams – Bigrams are groups of two written letters, two syllables, or two words, and are very commonly used as the basis for simple statistical analysis of text.

Bimodal Network – A network most commonly arising as a mixture of two different unimodal networks*.

Binarize – Divides your data into two sets; zero or one.

Bipartite Graph – Also called a bigraph. It's a set of nodes decomposed into two disjoint sets such that no two nodes within the same set are adjacent.

C

Centrality – The nearness of a node to all other nodes in a network. It displays the ability to access information through edges connecting other nodes. The closeness is the inverse of the sum of the shortest distances between each node and every other node in the network.

Centralization – Indicates the distribution of connections in the employee communication network as the degree to which communication and/or information flow is centralized around a single agent or small group.

Classic SNA density – The number of edges divided by the number of possible edges not including self-reference. For a square network, this algorithm* first converts the diagonal to 0, thereby ignoring self-reference (a node connecting to itself) and then calculates the density. When there are N nodes, the denominator is (N*(N-1)). To consider the self-referential information use general density.

Clique – A sub-structure that is defined as a set of nodes where every node is connected to every other node.

Clique Count – The number of distinct cliques to which each node belongs.

Closeness – Node that is closest to all other Nodes and has rapid access to all information.

Clustering coefficient – Used to determine whether or not a graph is a small-world network.

Cognitive Demand – Measures the total amount of effort expended by each agent to do its tasks.

Collocation – A sequence of words or terms which co-occur more often than would be expected by chance.

Column Degree – see Out Degree*.

Complexity – Complexity reflects cohesiveness in the organization by comparing existing links to all possible links in all four networks (employee, task, knowledge and resource).

Concor Grouping – Concor recursively splits partitions and the user selects n splits. (n splits -> 2^n groups). At each split it divides the nodes based on maximum correlation in outgoing connections. Helps find groups with similar roles in networks, even if dispersed.

Congruence – The match between a particular organizational design and the organization's ability to carry out a task.

Construct – A reduced form of Construct is found in ORA in the Near Term Impact Report.

Count – The total of any part of a MetaNetwork row, column, node, edge, isolate, etc.

CSV – File structure meaning Comma Separated Value. Common output structure used in database programs for formatting data.

D

Degree – The total number of edges to other nodes in the network.

Degree Centrality – Node with the most connections. (i.e. In the know). Identifying the sources for intel helps in reducing information flow.

Density –

- **Binary Network:** The proportion of all possible edges actually present in the Network.
- **Value Network:** The sum of the edges divided by the number of possible edges. (i.e. the ratio of the total edge strength that is actually present to the total number of possible edges).

Dyad – Two nodes and the connection between them.

Dyadic Analysis – Statistical analysis where the data is in the form of ordered pairs or dyads. The dyads in such an analysis may or may not be for a network.

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).

DyNetML – DynetML is an xml based interchange language for relational data including nodes, ties, and the attributes of nodes and ties. DyNetML is a universal data interchange format to enable exchange of rich social network data and improve compatibility of analysis and visualization tools.

E

Eigenvector Centrality – Node most connected to other highly connected nodes. Assists in identifying those who can mobilize others

F

FOG – (F)uzzy (O)verlapping (G)roups. Gives a better understanding of individuals spanning groups. Fuzzy groups are a more natural and compelling way of thinking of human social groups.

G

General density – The number of edges divided by the number of possible edges including self-reference. For a square network, this algorithm* includes self-reference (a node connecting to itself) when it calculates the density. When there are N nodes, the denominator is (N*N). To ignore self-referential information use classic SNA* density.

Gini coefficient – The measure of inequality of a distribution of income. Uses a ratio with values between 0 and 1: the numerator is the area between the Lorenz curve of the distribution and the uniform (perfect) distribution line; the denominator is the area under the uniform distribution line.

Graph Level Metric – A metric defined for, and gives a value for, the network as a whole. The metric is calculated once for the network. Examples are Centralization, Graph Hierarchy, and the maximum or average Betweenness.

GraphML – GraphML is a comprehensive and easy-to-use file format for graphs. It consists of a language core to describe the structural properties of a graph and a flexible extension mechanism to add application-specific data.

H

Hidden Markov Model – A statistical model in which the system being modeled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters from the observable parameters.

Homophily – (i.e., love of the same) is the tendency of individuals to associate and bond with similar others.

- Status homophily means that individuals with similar social status characteristics are more likely to associate with each other than by chance.
- Value homophily refers to a tendency to associate with others who think in similar ways, regardless of differences in status.

I

In-Degree – The sum of the connections leading to a node from other nodes. Sometimes referred to row degree.

Influence network – A network of hypotheses regarding task performance, event happening and related efforts.

Isolate – any node which has no connections to any other node

L

Lattice Network – A graph in which the edges are placed at the integer coordinate points of the n-dimensional Euclidean space and each node connects to nodes which are exactly one unit away from it.

Link – A specific relation between two nodes.

Link Analysis – A scientific area focused on the study of patterns emerging from dyadic observations. The relationships are typically a form of co-presence between two nodes. Also multiple dyads that may or may not form a network.

M

Main Diagonal – in a square network this is the conjunction of the rows and cells for the same node.

Math Terms

These mathematical terms and symbols are used: Let S be any set:

- card(S) = |S| = the cardinality of S (the cardinality of the node-sets is represented as |A|, |K|, |R|, |T|)
- $\Re$ denotes a real number
- Z denotes an integer

Network Algebra – The part of algebra that deals with the theory of matrices.

Measure – A measure is a function that maps one or more networks to R^n. Measures are often scalar (n=1) or vector valued with n=|V| or n=|U|.

MetaNetwork – A statistical graph of correlating factors of personnel, knowledge, resources and tasks. These measures are based on work in social networks, operations research, organization theory, knowledge management, and task management.

Monte Carlo – A random optimization of your organization

Morpheme – A morpheme is the smallest meaningful unit in the grammar of a language.

Multi-node – More than one type of node (people, events, locations, etc.).

Multi-plex – Network where the links are from two or more relation classes.

Multimode Network – Where the nodes are in two or more node classes.

N

Named-Node Recognition An Automap feature that allows you to retrieve proper names (e.g. names of people, organizations, places), numerals, and abbreviations from texts.

Neighbors – Nodes that share an immediate edge to the node selected.

Network – Set of links among nodes. Nodes may be drawn from one or more node classes and links may be of one or more relation classes.

A network N is a triple consisting of two sets of nodes, called U and V, and a set of edges E is an element of UxV. Thus, we write N=(U,V,E). An element e = (i,j) in E indicates a relationship or tie between nodes i is an element of U and j is an element of V. A network where U=V and therefore E is an element of VxV is called unimodal*; otherwise the network is bimodal*. We write G=(V,E) for unimodal* networks. For our purposes, unimodal networks will not contain self loops, which means that (i,i) is NOT an element of E for i is an element of V.

Newman Grouping – Finds unusually dense clusters, even in large networks.

Nodes – General things within a node class (e.g. a set of actors such as employees).

Node – A specific node (e.g., Joe, Martha, Bob; or, airplanes, buses, bicycles).

Node Class – The type of items we care about (knowledge, tasks, resources, agents).

Node Level Metric – is one that is defined for, and gives a value for, each node in a network. If there are x nodes in a network, then the metric is calculated x times, once each for each node. Examples are Degree Centrality*, Betweenness*, and Cognitive Demand*.

Node Set – A collection of nodes that group together for some reason.

Notation:

The following network notation is used throughout the document for an arbitrary network X:

- X(i,j) = the entry in the ith row and j^{th} column of X
- X(i,:) = i^{th} row vector of X
- X(:,j) = j^{th} column vector of X
- sum(X) = sum of the elements in X (also, X can be a row or column vector of a network)
- dich(X) = dichotomize (make binary) X, so that dich(X)(i,j) = 1 iff X(i,j) > 0
- X' = the transpose of X
- ~X = for binary X, ~X(i,j) = 1 iff X(i,j) = 0

- X@X = element-wise multiplication of two matrices (e.g. Z=X@Y => Z(i,j) = X(i,j)*Y(i,j))

O

ODBC – (O)pen (D)ata (B)ase (C)onnectivity is an access method developed by the SQL Access group in 1992 whose goal was to make it possible to access any data from any application, regardless of which database management system (DBMS) is handling the data.

Ontology – "The Specifics of a Concept". The group of nodes, resources, knowledge, and tasks that exist in the same domain and are connected to one another. It's a simplified way of viewing the information.

Organization – A collection of networks.

Out-Degree – The sum of the connections leading out from a node to other nodes. This is a measure of how influential the node may be. Sometimes referred to as column degree.

P

Pendant – Any node which is only connected by one edge. They appear to dangle off the main group.

Q

QAP Correlation – Calculates measures of nominal, ordinal, and interval association between the relations in two matrices, and uses quadratic assignment procedures to develop standard errors to test for the significance of association.

R

Random Graph – One tries to prove the existence of graphs with certain properties by assigning random edges to various nodes. The existence of a property on a random graph can be translated to the existence of the property on almost all graphs using the famous Szemerédi regularity lemma*.

Reciprocity – The percentage of nodes in a graph that are bi-directional.

Redundancy – Number of nodes that access to the same resources, are assigned the same task, or know the same knowledge. Redundancy occurs only when more than one agent fits the condition.

Relation – The way in which nodes in one class relate to nodes in another class.

Robustness – Two different definitions:

- Networks – Concerned with the reliability (Kim & Madard, 2004) and continued functioning of a network following an intervention. The robustness of a network is particularly relevant in communication-type and flow-oriented networks. The purpose for understanding robustness of a network has more of a management of the network connotation.
- Measures – This meaning has more of a statistical connotation. Studying the robustness of a measure of a network can also be referred to as conducting a sensitivity analysis on the measure. In keeping with the terminology of the most-recently published research in this area, in lieu of using the term sensitivity, we too will use the robustness term, although the terms can be used interchangeably.
  A measure is robust if a slight perturbation in its input produces a slight change in its output.

Row Degree – see In Degree*.

S

Scale-Free Network – Some nodes act as highly connected hubs (high degree), although most nodes are of low degree. Scale-free networks' structure and dynamics are independent of the system's size N, the number of nodes the system has. A network that is scale-free will have the same properties no matter what the number of its nodes is.

Self-Loop – A node with a connection to itself.

Simmelian Ties – Two nodes are Simmelian Tied to one another if they are reciprocally and strongly tied to each other and strongly tied to at least one third party in common.

Simulated Annealing – A method of finding optimal values numerically. It's a search method as opposed to a gradient based algorithm. It chooses a new point, and (for optimization) all uphill points are accepted while some downhill points are accepted depending on a probabilistic criteria.

The term Simulated Annealing draws its inspiration from metallurgy, where atoms within a metal are heated thereby dislodging them from a metal's internal structure transforming the metal into another atomic state. In this way, your organization is heated changing its components in the attempt to arrive at an optimized state.

Small-World Network – Small-World Networks will have sub-networks that are characterized by the presence of connections between almost any two nodes within them.

Social Network Analysis – 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).

Sphere of Influence – One node's direct relationship with one of it's neighbors as a function of specified path length.

Stemming – Stemming detects inflections and derivations of concepts in order to convert each concept into the related morpheme.

.stl file format – This file format is native to the stereo lithography CAD software created by 3D Systems. STL files describe only the surface geometry of a three dimensional object without any representation of color, texture or other common CAD model attributes and can us both ASCII and binary representations

Szemerédi's Regularity Lemma – A fundamental structural result in external graph theory due to Szemerédi (1978). The regularity lemma essentially says that every graph can be well-approximated by the union of a constant number of random-like bipartite graphs, called regular pairs.

T

Thesauri – Associates concepts with more abstract concepts

- Generalization Thesaurus – Typically 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.
- Meta-Network Thesaurus – Associates text-level concepts with meta-network categories.

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

Tie – see Edge

Topology – The study of the arrangement or mapping of the elements (links, nodes, etc.) of a network, especially the physical (real) and logical (virtual) interconnections between nodes.

Total Degree Centrality – The normalized sum of a node's row and column degrees.

Trails – Analyzes the trails that a node class makes through another node class; for example, how vessels pass through ports.

U

Unimodal networks – These are also called square networks because their adjacency network* is square; the diagonal is zero diagonal because there are no self-loops*.