NetMapper User Guide

V1 5/2020

Eric Malloy and Kathleen M. Carley

Netanomics

NetMapper is a language technology tool that can be used for text mining to support network analysis. NetMapper can operate in multiple modes and on over 40 languages for diverse forms of texts. NetMapper is interoperable with ORA.

A high level overview of the modes of operation is shown in table 1. Note that, not all forms of analysis are available for all forms of text.

Table 1. High Level Features					
Feature	Text blocks	Text micro-blocks			
Extracting semantic network	yes	no			
Extracting meta-network	yes	no			
Sentiment for key words based	yes	yes			
on context					
CUES	yes	yes			

A text block is a document or set of texts that have been turned in to a single document. Examples include: news articles, journal abstracts, the content from a set of tweets by the same user all in the same word file.

A text micro-block is a file, JSON or CSV, where each entry or row refers to a single text and all other entries are attributes of that text. Examples include: a json file containing multiple tweets or a csv file such that each row refers to a different journal article. In these files there is a variable that contains text, e.g., the content of the tweet or a column in a csv containing journal titles.

Each text, each text – whether a block or micro-block - is processed separately and results are exported. The user can choose what types of results to export. Any results from NetMapper can be imported in to ORA. Based on the language of the text, NetMapper will read the text either from left to right top to bottom or right to left, or right to left bottom to top. NetMapper will treat the items in the texts as "concepts" such that a concept can be one or more words. For example, "monkey" is a concept as is "man-in-the-moon." NetMapper identifies concepts and, for text blocks, the links between them. Links are based on relative proximity in the text.

NetMapper extracts and classifies teams based on tokenization and thesauri. NetMapper can be rapidly customized to support extraction of just terms of interest, rather than extracting all terms in the document.

NetMapper is lexicon based and employs an extensive set of thesauri, translation files and delete lists in over 40 languages. In addition, it also supports the use of user-generated domain thesauri and delete lists. Hence, users who are working in a specialized area or on a specialized topic can fine tune what concepts are extracted using the domain files.

Output: Formats

The output from NetMapper is in TSV and in the xml format read by ORA.

XML: When you use NetMapper to analyze text blocks you can choose to extract a semantic network and/or a meta-network. In the semantic network, all concepts are treated as the same ontological category, knowledge. The result is a single network showing concept to concept linkages for that text block. In the meta-network, all concepts are cross classified into an ontological category such as agent or task. The results is a meta-network that can have many subnetworks in it.

TSV: When you use NetMapper you can provide a set of keywords, for which the sentiment is identified. These keywords and the associated sentiment per text is then exported in a TSV file. A second TSV file that NetMapper can generate is associated with CUES. This file contains information per text on the frequency with which each of the CUES appears in the text.

Input: Types of Texts

Currently, to use NetMapper the user must have a set of texts. These may be text blocks, or text microblocks.

Text blocks are larger documents containing multiple sentences organized into one or more paragraphs. These texts should be pre-processed into a .txt file. Generally, images should be removed. Examples of types of texts include:

- News documents
- Journal articles
- Blog posts
- Word files containing all the tweets by a single user

NetMapper can accept data in multiple formats:

- US-ASCII
- UTF-8
- UTF-16
- UTF-32

Text micro-blocks are smaller documents, such as tweets or citations to journals. These texts should be combined into a single json or csv (or tsv) file that includes per entry/row the information on a particular micro-block. Examples of types of texts of this sort include:

- tweets
- reddit comments
- youtube comments

• bibliographic citations

NetMapper can accept data in multiple formats:

- json
- csv
- tsv

Concept

A concept is a word or phrase that serves as a single ideological idea. Examples are president and John F. Kennedy. When concepts are not categorized into an ontological category they are treated as being of type knowledge. Alternatively, they can be categorized into a set of ontological categories. These categories are based on the ORA ontology.

Ontological Categories

If meta-networks are generated, NetMapper uses a pre-defined ontology and automatically classifies concepts into this ontology. The ontological categories and types are described below:

- Agent specific, generic
 - o Individual actors
 - Specific unique often with first and last name Jamie O'Connor
 - Generic non-unique and often a role haberdasher
- Organization specific, generic
 - Groups, corporations, populations
 - Specific unique IBM
 - Generic a type Non-government organization
- Location specific, generic
 - Places things can be at
 - \circ Specific unique with lat and lon or place on map United States of America
 - Generic may be at multiple locations hill
- Event specific, generic
 - Major happenings that impact groups
 - Specific occur once World War I
 - Generic multiple occurrences Tornado
- Knowledge
 - Branches of knowledge
 - o Topics of interest
- Resource
 - Things that are not purely mental disease, food, wire
- Task
 - Activities eat

- Belief
 - o "isms" Catholicism
 - Sentiment positive, negative
 - Belief statements right to bear arms

Note that the "Knowledge" ontological category is used for semantic networks.

Types of Networks Supported

Two types of networks can be extracted: semantic networks and meta-networks. The semantic networks are concept to concept networks. From an ORA perspective, all concepts are treated as being of type knowledge. The meta-networks are concept to concept where each concept is also linked to its ontological category.

Operation

Within NetMapper, the user has a variety of options. The user can choose to remove words, to classify concepts into common terms, and can construct and use their own domain thesauri.

- Types of items in texts that can be removed
 - Stop words
 - Punctuation
 - o Numbers
 - 27 languages are supported
- Many concepts can be classified into common terms
 - o Thesauri based
 - Special thesauri are included for disease, sports, numeric expressions, in over 40 languages
- Domain thesauri
 - These are user designed
 - They can contain specialized words
 - They may be in a special language

NetMapper has a large number of thesauri that it will routinely use unless the user specifies not to use them. These universal thesauri support the activities suggested above. These thesauri include ones for specific agents, organizations, events, and locations. There are also universal thesauri for the generic agents, organizations, events, locations, knowledge, tasks, resources, and beliefs.

Thesauri

The purpose of a thesauri is two fold: First, it specifies how a concept should be referred to. Thus it provides information about what to translated ConceptFrom (the item in the raw text) to ConceptTo (the concept that will be visible in the output file). This provides the user with a way of reducing complexity and so the number of concepts in the networks by: 1) by converting a set of synonyms to a common word; 2) overcoming common typos; and 3) clustering words into topical areas based on user

choice. This also provides the user with a way of adding attributes such as the default valence for sentiment calculation.

Types of Thesauri

- Universal Thesaurus
 - NetMapper has a set of universal thesauri and translation files.
 - These thesauri are used by default; however, you can choose not to use them.
- Domain Thesaurus
 - A domain thesaurus is a user supplied file.
 - You may add multiple domain thesauri using the add button. If you select any domain thesauri in the list and click the remove button, it will be deleted from the list.

Thesauri Format

The format of a thesauri file is a tab separated file with a set of columns specifying relevant information. Row 1 must contain the header for that column using the name specified below with exact spelling and case. All and only the following fields can be included.

Thesauri row 1 headers:

- 1. conceptFrom
 - This is a required field
- 2. concetpTo
 - This is a required field
- 3. metaOntology
 - This is a required field
 - Only allowed values are:
 - Agent
 - Organization
 - Location
 - Event
 - Knowledge
 - Resource
 - Task
 - belief
- 4. nodetype
 - This is a required field
 - Only allowed fields are generic or specific for agent, organization, location and event
 - For other ontological classes this =must be blank so blank for knowledge, resource,

task or belief

- 5. Category 1 < Optional>
- 6. Category 2 < Optional>
- 7. Category 3 < Optional>
- 8. Country < Optional>
- 9. First Name < Optional>
- 10. Last Name <Optional>

- 11. Gender < Optional>
- 12. Suffix <Optional>
- 13. Language <Optional>
- 14. Acronym <Optional>
- 15. Valence <Optional>
- 16. Evaluation <Optional>
- 17. Potency <Optional>
- 18. Activity <Optional>
- 19. Affect Mean < Optional>
- 20. Military Role < Optional>
- 21. Political Role < Optional>
- 22. Religious Role <Optional>
- 23. Abusive <Optional>
- 24. Exclusive <Optional>
- 25. PowerAnger <Optional>
- 26. PowerEncourage <Optional>
- 27. PowerFear < Optional>
- 28. PowerForbidden < Optional>
- 29. PowerGreed <Optional>
- 30. PowerLust <Optional>
- 31. PowerSafety <Optional>
- 32. Absolutist <Optional>
- 33. Equivocal <Optional>
- 34. Connective <Optional>
- 35. NamedEntity <Optional>
- 36. Pronoun_Level <Optional>
- 37. Adverb < Optional>
- 38. OtherUsage <Optional>
- 39. Inclusive <Optional>

Not all the columns need values for every entry. What is required is conceptFrom (this is what you want search for), conceptTo (this is what you want it replaced with when its found), metaOntology the ontology it has to be either agent, event, organization, location, or knowledge, and nodeType which is either "specific" or "generic".

Each line after the header row contains information on a concept.

The file must be saved as UTF-8 (without BOM). To do this, do the following. In excel save the file as unicode. This creates a tab separated file that is UTF-16. Then using another tool like Notepad++, VIM, Emacs, etc .. re-save as utf8 without BOM.

NetMapper has a set of pre-defined thesauri in a large number of domains. The user can choose to use these or not. By default they are all applied. In addition the user can choose to create and use a domain thesauri.

In a domain thesauri there must be at least four columns. These are conceptFrom, conceptTo, Ontology, and nodetype.

Delete List

A delete list defines a set of concepts that should be deleted and not included in the resultant coded network. NetMapper has a set of pre-specified delete lists. These are the universal delete lists. By default, all universal delete lists will be applied. The user, however, can choose not to apply any or all of these delete list and/or can add a customized domain delete list.

Types of Delete Lists

- Universal Delete List
 - NetMapper has a set of universal delete lists available.
 - These delete lists are used by default; however, you can choose not to use them.
- Domain Delete List
 - These are user supplied files. The files contain terms that will be deleted from the text and will therefore not show up in the generated networks.

Delete List Format

The format of a delete list is a csv file with only one column. Each concept to be deleted is in its own line. The files have only a single concept per line. A concept may contain more than a single word. Most all Unicode characters are accepted with the exception of tabs.

Available Universal Delete Lists

The set of universal delete lists contain concepts for:

- Time
- Measurement
- Symbols
- Stop words
- Numbers
- Regular expressions

Application of Thesauri and Delete Lists

Thesauri and delete lists are applied in the following order:

- 1. Domain Thesauri
- 2. Universal Thesauri
- 3. Domain Delete List
- 4. Universal Delete Lists

Link Generation

NetMapper creates networks through link generation. After the thesauri and delete lists are applied then NetMapper extracts the networks by identifying links among concepts. A link is placed between two concepts just in case they are within the window of operation. The window of operation is defined by either or both number of concepts and syntactic structure (e.g., number of clauses, sentences, or paragraphs). There exist default choices which have been found empirically to lead to the best results for the type of texts being examined. However, the user can choose to change the defaults by specifying:

- Window size based on number of words
- Window size based on number of syntactic units
- Whether or not deleted terms are "counted" in defining the size of the window

Finally, NetMapper auto detects language and moves correctly either from left to right up to down, or the reverse.

Outputs

- NetMapper generates the following outputs. DyNetML files for import to ORA
 - Meta Network (With or Without Unknowns)
 - Semantic Network (With or Without Unknowns)
- Original Text with modifications
 - Just UT Concepts in the Text
 - UT and DT Concepts
 - With or Without Deleted Concepts
- CSV files containing sentiment scores

Description of NetMapper Operations

The following pages describe the set of screens in NetMapper and the different parameters that the user can set, the input files of relevance, and the output files that are possible to generate.

Using the options described below you will be able to load in files to process, choose which delete lists or thesauri to use, select specialized options, choose where links are placed, and choose the context for sentiment.

Start: Data Entry

When you launch Netmapper you will see tabs for four separate pages: Files, Advanced Settings, Delete Lists and Thesauri. Each of these has a distinct function. You will also see in the upper right the option Help. If you click on Help you will get a pop-up section that contains limited on-line help. In the bottom right you will see the word next. If you click on next you will go to a new page that lets you run NetMapper. Unless you want to alter the defaults you only need to add files on the Files Tab and then click Next.

Files Tab

The purpose of the files page is to tell NetMapper what text files you want to process, what format they are in, and where you want the data to be stored. It's all about file management.

T L	C*1	· · · · · · · ·		1			NI.INA	C E' 4
IND	the nade	IC THA 1	rirct tan t	nat w/III •	COO WINDN	VOULISUNCE	Netivianner	
1 IIC	IIICS DUGC	13 1110 1	m st tab t			you launch		JUC I ISUIC I.
						1		0

File Help Image: Price Pr	🕪 NetMapper
Files Advanced Settings Delete Lists Thesauri Domain Thesaurus Add Remove Domain Delete List Add Remove Concepts Of Interest case sensitive Concepts Of Interest List Add Input Files Add Raw Text File Remove File Import Tweets Input Files Add Raw Text File Remove File Import CSV File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	File Help
Domain Thesaurus Add Remove Domain Delete List Add Remove Concepts Of Interest case sensitive Concepts Of Interest case sensitive Concepts Of Interest List Add Remove Input Files Add Raw Text File Remove File Import Tweets Import CSV Input F File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	Files Advanced Settings Delete Lists Thesauri
Domain Delete List Add Remove Concepts Of Interest case sensitive Concepts Of Intersest List Add Remove Input Files Add Raw Text File Remove File Import Tweets Import CSV Input F File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	Domain Thesaurus Add Remove
Domain Delete List Add Remove Concepts Of Interest case sensitive Concepts Of Interset List Add Remove Input Files Add Raw Text File Remove File Import Tweets Import CSV Input F File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	
Domain Delete List Add Remove Concepts Of Interest case sensitive Concepts Of Interset List Add Remove Input Files Add Raw Text File Remove File Import Tweets Import CSV Input Files File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	
Concepts Of Interest case sensitive Concepts Of Intersest List Add Remove Input Files Add Raw Text File Remove File Import Tweets Import CSV Input F File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	Domain Delete List Add Remove
Concepts Of Interest case sensitive Concepts Of Intersest List Add Remove Input Files Add Raw Text File Remove File Import Tweets Import CSV Input F File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	
Concepts Of Interest case sensitive Concepts Of Intersest List Add Remove Input Files Add Raw Text File Remove File Import Tweets Import CSV Input F File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	
Concepts Of Intersest List Add Remove	Concepts Of Interest case sensitive
Input Files Add Raw Text File Remove File Import Tweets Import CSV Input F	Concepts Of Intersest List Add Remove
Input Files Add Raw Text File Remove File Import Tweets Import CSV Input F	
Input Files Add Raw Text File Remove File Import Tweets Import CSV Input F	
File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	Input Files Add Raw Text File Remove File Import Tweets Import CSV Input F
File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	
File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	
File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	
File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	
File Name Akan Amharic Arabic Armenian Assamese Awadhi Azerbaij	
And Annone Arole Annend Assonese Awduli Azerbaij	File Name Akan Ambaric Arabic Armenian Assamese Awadhi Azerbaiti
	The Name Avail Annanc Arabic Annenian Assamese Awauni Azerbaij
Back Next >	< Back Next >

Figure 1. Files Page

- Domain Thesaurus
 - Select a domain thesaurus if the user has one prepared for the input data.
- Domain Delete List
 - Select a domain delete list if one has been prepared for the input data.
- Input Files
 - Lists all the files to be processed.
 - Use the "Add Raw Text File" button to bring up a file dialog box. Multiple files or a directory can be selected. The selected files will be processed as text files.

- The remove button is used to remove individual files from the list. This is done by selecting a file or files and clicking the remove button.
- \circ The "Import Tweets" button is used to import twitter data in JSON.
- If you choose to import Tweets then you get the screen shown in Figure 2.

	Profile: Custom Profile	•	
Select a file:	er_CASOS_SI_case_study. 1	1527706970077.anonymized.json	Browse
JSON Field		Field Type	
entities.media	a].sizes.thumb.h	Tweet ID	•
extended_en	tities.media[].sizes.small.h	Text	•
user.statuses	s_count	Text	•
entities.media	a[].sizes.large.h	Text	-
extended_en	tities.media[].sizes.thumb.re	Select	-
retweet_cour	nt	Select	-
user.is_trans	lator	Select	-
reply_count		Select	-
entities.media	a[].sizes.thumb.resize	Select	-
entities.urls[]	.indices	Select	-
extended_en	tities.media[].sizes.small.w	Select	-
user.entities.	description.urls[].url	Select	-
user.follower	s_count	Select	-
lang		Select	-
user.name		Select	-
filter_level		Select	-
user.entities.	description.urls[].expanded	Select	-
entities.media	a[].expanded_url	Select	•
user.profile_t	background_tile	Select	-
user.translat	or_type	Select	• •
Ok	Load Profile Save	Profile Clear All Canc	el

Figure 2. Twitter import screen

- The Profile drop down is used to select a JSON profile that describes what field is to be used as the tweet id and fields are to be processed as text.
- Only one field can be selected for tweet id.
- Multiple fields can be selected as text fields.
- The "Default JSON Twitter" profile is a NetMapper supplied profile that cannot be changed. It uses "id" as Tweet ID and "text" as the Text field to process.
- You may select the "Custom Profile" option and create your own profile.

- Use the browse button to select a JSON file or directory containing JSON files.
 When that is done, the files are parsed and the table below is populated with all the fields found in all the JSON files.
- Field type "Text" or "Tweet ID". Only one field maybe the "Tweet ID".
- Once a valid profile (there must be at minimum one field selected as "Text") is either loaded or created and saved. The "Ok" button will be enabled, press that to go back to the Files Page.
- "Load Profile" allows the user to load a previously saved profile.
- "Save Profile" allows the user to save a custom profile, all custom profiles must be saved before they can be used. When the user clicks the "Save Profile" button, they will be prompted to enter a name for their profile. That name is the name that will show in the Profiles drop down box at the top of the dialog.
- "Clear All allows the user to clear all selections"
- "Cancel" will return the user to the Files Page without adding any files to be processed.
- If you choose to import CSV then you get the screen shown in Figure 3.

🔅 Import Chracter Separated Files (C	:SV)							
Profile: Custom Profile 🗸								
Character separator ,								
Select a file: C:\Users\jonilith\Documents	SampleData_Netmapper.csv Browse							
CSV Column	Field Type							
screen_name	Select 🗸 🔺							
id_str	Unique ID 🗸							
name	Select 👻							
statuses_count	Select 👻							
time_zone	Select 👻							
description	Select 👻							
reply_to_user_id	Select 👻							
listed_count	Select 👻							
reply_to_status_id	Select 👻							
contributors_enabled	Select 👻 E							
verified	Select 👻							
favourites_count	Select 👻							
friends_count	Select 👻							
is_translator	Select 👻							
lat	Select 👻							
retweet_status_id	Select 👻							
status_source	Text 👻							
status_retweet_count Select 🗸								
status_id Select 👻								
has_default_profile	Select 👻 👻							
Ok Load Profile	Save Profile Clear All Cancel							

Figure 3. CSV import screen

- The Profile drop down is used to select a CSV profile that describes what field is to be used as the tweet id and fields are to be processed as text.
- Only one field can be selected for tweet id.
- Multiple fields can be selected as text fields.
- You may select the "Custom Profile" option and create your own profile.
- Use the browse button to select a CSV file or directory containing CSV files.
 When that is done, the files are parsed and the table below is populated with all the fields found in all the CSV files.
- Field type "Text" or "unique id". Only one field maybe the "unique id".
- Once a valid profile (there must be at minimum one field selected as "Text") is either loaded or created and saved. The "Ok" button will be enabled, press that to go back to the Files Page.
- "Load Profile" allows the user to load a previously saved profile.

- "Save Profile" allows the user to save a custom profile, all custom profiles must be saved before they can be used. When the user clicks the "Save Profile" button, they will be prompted to enter a name for their profile. That name is the name that will show in the Profiles drop down box at the top of the dialog.
- "Clear All allows the user to clear all selections"
- "Cancel" will return the user to the Files Page without adding any files to be processed.

• Language Selection

- Allows the user to select additional (to English) languages to use for processing each input file. Note that English is the default and is always chosen.
- Column Popup Menu (Figure 4)
 - The column popup menu has two items, "Select For All" and "Deselect for All".
 - "Select for All" Will select that language for all the files.
 - "Deselect for All" Will deselect that language for all the files.
- Row Popup Menu (Figure 5)
 - The Row Popup menu has three options "Select All Languages", "Deselect All Languages" and "Clone Entry to All Others"
 - "Select All Languages" will select all the languages for the row that was right clicked on.
 - "Deselect All Languages" will unselect all the languages (except for English) for the row that was right clicked on.
 - "Clone Entry to All Others" will select the same languages for all the files that are select for the current row.

🛓 Text to Network Wizard			<u> </u>				L		x
File Help									
Files Advanced Settings De	elete Lis	ts Thesa	uri						
Demain Theory you		Domou							
Domain Thesaurus	kaa	Remov	e						
INO FILE SELECTED									
Domain Delete List	dd	Remov	e						
No File Selected									
Input Files Add	Rem	ove	Input	Files: 12					
C:\darfur\2001_06_14_child_s	oldier.t	xt							•
C:\darfur\2001_09_06_oil.txt									
C:\darfur\2004_01_23_child_s	oldiers.	txt							=
C:\darfur\2004_05_20_famine	.txt								
C:\darfur\2004_07_20_rains.t	xt								
C:\darfur\2004_11_18_famine	.txt								
C:\darfur\2004_12_15_pollo.t	xt								
C:/bartur/2004_12_20_0il.txt									-
File Name	Akan	Amharic	Arabic	Armenian	Assamese	Awadhi	Azerbaijani	Balochi	Re
2001 06 14 child soldier tyt					- SSGIIICSC				
2001_09_06_oil_txt		Select	t for All						- Âl
2004 01 23 child soldiers				H					
2004 05 20 famine.txt		Desel	ect for A	AII					
2004 07 20 rains.txt									=
2004_11_18_famine.txt									
2004_12_15_polio.txt									
2004_12_20_oil.txt									
2004_12_22_Annan.txt									
2004 12 23 polio.txt									•
							< Back	Ne	xt >

Figure 4. Column Pop-Up Menu

🛃 Text to Network Wizard				6		X
File Help						
Files Advanced Settings Delete Lists Thesauri						
Domain Thesaurus Add Remove						
No File Selected						
Domain Delete List Add Remove						
No File Selected						
Input Files Add Remove Input File	s: 12					
C:\darfur\2001_06_14_child_soldier.txt						
C:\darfur\2001_09_06_oil.txt						
C:\darfur\2004_01_23_child_soldiers.txt						-
C:\darfur\2004_05_20_famine.txt						=
C:\darfur\2004_07_20_rains.txt						
C:\darfur\2004_11_18_famine.txt						
C:\darfur\2004_12_15_polio.txt						
C:\darfur\2004_12_20_oil.txt						~
File Name Select All Languages	enian	Assamese	Awadhi	Azerbaijani	Balochi	Be
2001_06_14_child_						_ ^
2001_09_06_oil.txt Deselect All Languages						
2004_01_23_child_ Clone Entry to All Others						
2004_05_20_famine.cxc						-==
2004_07_20_rains.txt						_
2004_11_18_famine.txt						
2004_12_15_polio.txt						
2004_12_20_oil.txt						
2004_12_22_Annan.txt						
2004 12 23 polio.txt						
	_			< Back	Ne	xt >

Figure 5. Row Pop-Up Menu

Advanced Settings Tab

The advanced settings tab providers the user with a series of options that allow for more fine-grained control of how networks are extracted from the raw texts (see Figure 6). This is a completely optional tab – i.e., you can just accept the defaults and never look at this tab.

NetMapper						
File Help						
Files Advanced Settings Delete Lists Thesauri						
A concept is one or more words that together form an ideational unit - e.g., man, or cour A thesaurus defines what all words or phrases should be translated into which specific co This includes both translating plurals, or tenses to a common form such as tries to try and translating less used synonyms to a common form such as provides backing to support; a indentWidth);other languages into English	ntry of origin. nceptn I trying to try; nd translating fr	om				
Include concepts in outputs. This will affect all tsv and dynetML files generated with the	exception of the	Measures				
Universal Thesaurus						
Concepts that have been repaced by the Universal Thesaurus Entries will be in the output	Its					
Concepts that have been repaced by the Domain Thesaurus Entries will be in the outputs The domain thesaurus contains the translations and conversions created by the user. Creating and using a domain thesaurus provides the user with the ability to make use of t terms for the project being done, and to classify those concepts by type.	terms of art and	specialization				
Unknown Terms						
Concepts that were not identified in the any thesauri or the COI will be in the ouptuts. An unknown term is a single term (denoted by a set of contiguous letters separated from others by spaces or end of sentence or clause symbols). Even when both the domain and universal thesaurus are used there are likely to be terms in the documents processed that do not occur in a thesaurus. These are unknown terms. When unknow terms are selected they will in the meta-network be classified into the type "other." Selecting unknown terms to be used the first time you run NetMapper is a good way of identifying those terms that you might want to include in a domain thesaurus.						
Only Concepts Of Interest						
Only the concepts in the COI and Domain Thesaurus will be in the outputs						
Ignore deleted words in search frame. Each deleted word will not be counted toward	d the frame limit					
Use binary link weights which will be a 1 or 0. Otherwise linke weights are the numbe	er of times that s	ame link o				
Treat British and American spellings the same (grey vs gray)						
Automatically expand contractions. This will turn "we're" to "we are" and so on						
✓ Use common typos list (English)						
V Automatically Detect Language of Input Files						
	< Back	Next >				

Figure 6. Advanced Settings Tab

• Advanced Settings Tab

- Include concepts in Meta-Network
 - All Concepts Everything
 - Only From Universal Only include terms from the universal thesaurus
 - Only From Domain Only include terms from the domain thesaurus
- Include concepts in Semantic Network
 - All Concepts Everything
 - Only From Universal Only include terms from the universal thesaurus
 - Only From Domain Only include terms from the domain thesaurus
- Ignore deleted words in search frame
 - Means that NetMapper will not count terms that have been deleted when counting words within the search window to create links
- Use binary link weights
 - All edges will have a weight of 1
- o Treat British and American Spellings that same
 - British spellings will be treated as their American equivalent as opposed to being processed as a different word.
- Use Common Typos list
 - NetMapper has the ability to fix some basic typos that often occur. This selection tells NetMapper to make those assumptions.
- o Automatically Detect Language of Input Files
 - If selected NetMapper will attempt to determine the language of each individual input file. This may cause other translation thesauri to be loaded to process any given input file.

Delete Settings Tab

The delete settings tab is used to specify which delete lists you want to use (See Figure 7). By default all existing universal delete lists, in all the user selected languages, are used.

🖆 Text to Network Wizard		
File Help		
Files Advanced Settings Delete Lists Thesauri		
☑ Universal Delete List		
☑ Delete symbols		
Delete numbers in all selected languages		
Delete measurement phrases in all languages		
Delete time phrases in all selected languages		
	< Back	Next >

Figure 7. Delete List Tab

- Delete Lists
 - o Universal Delete List
 - By selecting this option all words in the universal delete list will be deleted from the text and ignored for a good deal of NetMappers processing.
 - Delete symbols
 - Many non alpha numeric symbols that are also not punctuation will be deleted
 - Delete numbers in all select languages
 - Numbers will be removed, that means 1,2,3 will be deleted as well as one, two, and three.
 - Delete measurement phrases in all languages

- Selecting this options will remove terms like inches, centimeters and so on
- Delete time phrases in all selected languages
 - Similar to delete measurement this option will delete terms like hour, minute and second.

Thesauri Tab

The thesauri tab is used to specify which of the universal thesauri are used (see Figure 8). By default all are selected.

🖆 Text to Network Wizard		
File Help		
Files Advanced Settings Delete Lists Thesauri		
☑ Use universal thesaurus (English)		
Select Translation ConceptTo Granularity: ConceptTo 🗸		
л	< Back	Next >

Figure 8. Thesauri Tab

• Thesauri Tab

1

• Use universal thesaurus (English)

- This will mean that a large number of predefined thesaurus entries will be used in processing text. This also includes all the translation thesauri. The general scope of the thesauri to be included by selecting this option covers many well known agents, locations, events and knowledge.
- o Select Translation ConceptTo Granularity
 - There are several levels of granularity that a term can be translated too. In order of most specific to most general they are ConceptTo, Category 1, Category 2.

Next: Running NetMapper

Once you click next you will see a new page focusing on the choices you need to make to build the networks (Figure 9). You will be asked what type of networks you want to build, and how you want to set the windows. Once you have entered all the relevant data press Next in the lower right hand side.

NetMapper		. 🗆	x
File Help			
Network Type A meta-network is a network in which the concepts have been classified into types (e.g., agent, organization, location). In this case you can easily choose just a type of no e.g., to just look at the agents and their connection to each other. A link indicates that the within a certain distance of each other. Meta Network A semantic network is a network in which each node is a concept. The links in this network represent whether the two concepts occurred within a certain dista Semantic Network List of filtered concepts found in each text (filtered in adavanced settings such as COI or D Concept List List of concepts, frequency and sentiment type information for each concept NetMapper TSV Statistics about each text Usage Measures	ode, two concepts o ance of each o omain only	occurred	e test.
Search Window Type Sentence			Ŧ
Search Window Width			
Window Width = Entire Document			2 👻
Sentiment Window Width			3
	< Back	Ne	xt >

Figure 9. Network Selection

- Network Generation Page
 - Network Type
 - Generate Meta Networks (not available for JSON or CSV).
 - Generate Semantic Networks (not available for JSON or CSV)
 - Generate a list of valid concepts for each text. This will be filtered based on your selections for concepts filtering in the advanced tab. Default is all concepts.

- Generate the NetMapper tsv file (rnmf.tsv is the extension) which is a list of all concepts in each text with statistics for each such as frequency, sentiment, and uncertainty.
- Generate Usage Measures, this is a tsv with many statistics about each text including
 - concept count
 - reading difficulty
 - named entity
 - abusive
 - exclusive
 - poweranger
 - powerencourage
 - powerfear
 - powerforbidden
 - powergreed
 - powerlust
 - powersafety
 - absolutist
 - equivocal
 - connective
 - postive
 - negative
 - 1st person
 - 2nd person
 - 3rd person
 - pronoun#
 - numbers
 - expletive
 - all caps
- Automatically Generated Reports:
 - URL Format
 - TSV file of origins (i.e. tweet id) to URLs found in the document.
 - Date Format
 - TSV file of origins (i.e. tweet id) to dates found in the document.
 - Hashtag Format
 - TSV file of origins (i.e. tweet id) to hashtags found in the document.
 - Zip Code Format
 - TSV file of origins (i.e. tweet id) to zip codes found in the document.
 - Twitter Handle Format
 - TSV file of origins (i.e. tweet id) to Twitter handles found in the document.

- Phone Number Format
 - TSV file of origins (i.e. tweet id) to phone numbers found in the document.
- Emoticon Format
 - TSV file of origins (i.e. tweet id) to emoticons and emojis found in the document.
- Report Naming
 - Input file name + report name + format
 - Ie. If your input file is myinput.txt then the usage measures format file will be named
 - myinput.txt.usage_measures.tsv
- Search Window Type
 - Word will determine the search width by number of words in the window.
 - Sentence will determine the search width as number of sentences in the window.
- o Search Window Width
 - The number of words or sentences that should be used in the window when determining links between terms in the network.
- Sentiment Window Width
 - The window width to be used for the sentiment network.

Next: Output

Your next task is to tell NetMapper where to put the output files (see Figure 10). You provide a root directory and NetMapper will generate the requested output, label each file uniquely, and put it into this output directory.

RetMapper 1.0.0.14	of the billion beautiful setting to a	
File Help		
Enter a directory in which to save the out	put:	
C: \Users \ionilith \Documents \nmoutput		Browse
Files	Output Files Root	
C:\eula.1040.txt	eula. 1040. txt	
C:\eula.1033.txt	eula. 1033. txt	
C:\eula.2052.txt	eula.2052.txt	
C:\eula.1036.txt	eula. 1036.txt	
C:\eula.1031.txt	eula. 1031.txt	
C:\eula.1028.txt	eula. 1028.txt	
C:\eula.1041.txt	eula. 1041.txt	
C:\eula.1042.txt	eula. 1042. txt	
C:\eula.3082.txt	eula.3082.txt	
		ack

Figure 10. Output Directory Selection

- Output Page
 - \circ Output Direction is the location to put all the output files
- Files list
 - In this list you can choose to change the root naming convention of a given input file. This means that if you have a file named symbols.txt, all ouput files will be named symbols.<whatever extensions are necessary>. However if another file in the list has

the same root name, the results from one will overwrite the other. By manually changing the root, that can be prevented.

Run

1

When you are finished, you simply press run in the bottom right of the Output page. Notice, until you press run you can go back and forth between the different pages and tabs.