AutoMap Component Reference Manual

Michael W. Bigrigg
August 2009
CMU-ISR-09-124

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 work was supported in part by the Office of Naval Research under Contract No. N000140811223 to do the componentization of AutoMap, and Contract No. N000140811186 for work on thesauri, and Contract No. N000140811186 for extensions to handle beliefs and attributes; by the Army Research Institute under Grant No. W91WAW07C0063 to handle the part of speech analysis and anaphora resolution; and by the Air Force Office of Scientific Research under Contract No. FA9550-05-1-0388. 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 Office of Naval Research, the Air Force Office of Scientific Research, the Army Research Institute or the U.S.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
AUG 2009		**00-00-2009 to 00-00-2009**

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

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

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**73**	

Key Words: Semantic Network Analysis, Dynamic Network Analysis, Component Programming, Social Networks, AutoMap

Abstract

AutoMap is software for computer-assisted Network Text Analysis (NTA). The AutoMap components are a collection of programs that implement the features available in AutoMap.

Table of Contents

AutoMap Components

The AutoMap components are a collection of over 60 java executable programs for use in Network Text Mining applications. These components are used by interface programs such as AutoMap GUI, AutoMap Script, and SORASCS.

How To Use

The components require Java 1.6 or higher and are callable via a command line interface. The following example shows how to call the Lowercase routine.

C:\> java –cp am3.jar edu.cmu.casos.automap.Lowercase C:\in C:\out

The C:\in and the C:\out directories are example input and output directories and are given as command line parameters passed to the Lowercase routine.

The AutoMap components are not directly callable from a Java program except to the extent to which any main() may be called. No guarantee is given to components being called in this manner.

Classifications

Preprocessing Filter

Generation Routine

Supplemental Procedure

File Utility

CEMap Component

Internal Command

External Tool

Component Reference

This section gives an alphabetical list of all AutoMap components. This list may not include routines that are used by the AutoMap components

indirectly or routines that may be called by 3rd party libraries that AutoMap may make use of.

Mandatory arguments are listed in <angle brackets> with the pipe | symbol being used to identify accepted options. Optional arguments are in [square brackets]. Routines marked with an asterisk (*) denote components that heavily leverage a 3rd party library, other than libraries provided with Java.

AddAttribute

Classification: Supplemental Procedure

Usage:

java –cp am3.jar edu.cmu.casos.automap.AddAttribute <semanticnet_dir> <output_dir> <attrib.csv>

Parameters:

<semanticnet_dir>	input directory of DyNetML files
<output_dir>	output directory of DyNetML files
<attrib.csv>	

AddAttribute3Col

Classification: Supplemental Procedure

Usage:

java -cp am3.jar edu.cmu.casos.automap.AddAttribute3Col <semanticnet_dir> <output_dir> <attrib.csv>

Parameters:

<semanticnet_dir>

<output_dir>

<attrib.csv>

AddImages

Classification: Supplemental Procedure

Usage:

java -cp am3.jar edu.cmu.casos.automap.AddImages <network file> <outfile> [-k]

Parameters:

<network file>

<outfile>

-k flag preserves pre-existing images

AddMeta

Classification: Internal Command

Usage:

java -cp am3.jar edu.cmu.casos.automap.AddMeta <metadata_file> <input_dir> <output_dir> <column_type>

Parameters:

<metadata_file>

<input_dir>

<output_dir>

<column_type>

AddTimeFileName

???

Usage:

java -cp am3.jar edu.cmu.casos.automap.AddTimeFileName <input_dir>

Parameters:

<input_dir>

Anaphora

????

Usage:

java -cp am3.jar edu.cmu.casos.automap.Anaphora <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

AttributesToProps

Classification: Supplemental Procedure

Usage:

java -cp am3.jar edu.cmu.casos.automap.AttributesToProps <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

BeliefEnhancement

Classification: Supplemental Procedure

Usage:

java -cp am3.jar edu.cmu.casos.automap.BeliefEnhancement <input_dir> <output_dir> <belief_enhancement_file> <network type m|s>

Parameters:

<input_dir>

<output_dir>

<belief_enhancement_file>

<network type m|s>

BigramText

Classification: Internal Command

Usage:

java -cp am3.jar edu.cmu.casos.automap.BigramText <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

CRFSuggestion

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.CRFSuggestion <input_dir> <output_dir> <crf_dir>

Parameters:

<input_dir>

<output_dir>

<crf_dir>

CharsetToolkit

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.CharsetToolkit <input> <output_dir> [Directory|perFile]

Parameters:

<input>

<output_dir>

[Directory|perFile]

Note: input can either be a directory or a file

ClickIt

Classification: Supplemental Procedure

Usage:

java –cp am3.jar edu.cmu.casos.automap.ClickIt <network_file> <output_file> <location>

Parameters:

<network_file>

<output_file>

<location>

CombineThesauri

Classification: Supplemental Procedure

Usage:

java -cp am3.jar edu.cmu.casos.automap.CombineThesauri <output_file> [thes_file] [thes_file] ...

Parameters:

<output_file>

[thes_file]

ConceptList

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.ConceptList <input_dir> <output_dir> <gram_type>

Parameters:

<input_dir>

<output_dir>

<gram_type>

ConvertTextEncoding

Classification: Preprocessing Filter

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.ConvertTextEncoding <input> <output_dir> <encoding> [perDirectory|perFile]

Parameters:

<input>

<output_dir>

<encoding>

[perDirectory|perFile]

Note: input can either be a directory or a file

Dedupe

Classification: File Utility

Usage:

java -cp am3.jar edu.cmu.casos.automap.Dedupe <input_dir> <output_dir> <duplicate_dir> <log_dir>

Parameters:

<input_dir>

<output_dir>

<duplicate_dir>

<log_dir>

Delete

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.Delete <input_dir> <output_dir> <D|R> <delete_file>

Parameters:

<input_dir>

<output_dir>

<D|R>

<delete_file>

DyNetMlEditor

Classification: External Tool

Usage:

java -cp am3.jar edu.cmu.casos.automap.DyNetMlEditor <input_file> <output_file> <node_list_file>

Parameters:

<input_file>

<output_file>

<node_list_file>

ExtractNumerics

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.ExtractNumerics <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

FetchMail

Classification: CEMap Component

Usage:

java -cp am3.jar edu.cmu.casos.automap.FetchMail <protocol> <username> <password> <server> [folder] <output_body> <output_header>

Parameters:

<protocol>

<username>

<password>

<server>

[folder]

<output_body>

<output_header>

protocol may be any of imap, imaps, pop3, or pop3s

server string is of the form address[:port]

FilterDirectory

Classification: File Utility

Usage:

java -cp am3.jar edu.cmu.casos.automap.FilterDirectory <input_dir> <output_dir> <filter>

Parameters:

<input_dir>

<output_dir>

<filter>

FilterText

Classification: File Utility

Usage:

java -cp am3.jar edu.cmu.casos.automap.FilterText <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

General

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.General <input_dir> <output_dir> <thesaurus>

Parameters:

<input_dir>

<output_dir>

<thesaurus>

GenerateMetaNetwork

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.GenerateMetaNetwork
<metalist_dir> <semanticlist_dir> <properties_dir> <output_dir>

Parameters:

<metalist_dir>

<semanticlist_dir>

<properties_dir>

<output_dir>

GenerateSemanticNetwork

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.GenerateSemanticNetwork <concept_dir> <semanticlist_dir> <properties_dir> <output_dir>

Parameters:

<concept_dir>

<semanticlist_dir>

<properties_dir>

<output_dir>

KStemMain

***

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.KStemMain <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

KStemmer

***

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.KStemmer <input_dir> <output_dir> <stem_capitalization y|n>

Parameters:

<input_dir>

<output_dir>

<stem_capitalization y|n>

Keep

Classification: Internal Command

Usage:

java -cp am3.jar edu.cmu.casos.automap.Keep <keep_file> <input_dir> <output_dir> <adjacency>

Parameters:

<keep_file>

<input_dir>

<output_dir>

<adjacency>

KeepConcepts

Classification: Internal Command

Usage:

java -cp am3.jar edu.cmu.casos.automap.KeepConcepts <input_dir> <output_dir> <keep_file>

Parameters:

<input_dir>

<output_dir>

<keep_file>

KeyWordInContext

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.KeyWordInContext <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

Lowercase

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.Lowercase <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

LowercaseConcepts

Classification: Internal Command

Usage:

java -cp am3.jar edu.cmu.casos.automap.LowercaseConcepts <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

Merge

Classification: Supplemental Procedure

Usage:

java -cp am3.jar edu.cmu.casos.automap.Merge <input_dir1> <input_dir2> <output_dir>

Parameters:

<input_dir1>

<input_dir2>

<output_dir>

MergeDeleteLists

Classification: Supplemental Procedure

Usage:

java -cp am3.jar edu.cmu.casos.automap.MergeDeleteLists <output_file> [delete_list_file] [delete_list_file] ...

Parameters:

<output_file>

[delete_list_file]

MetaNet

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.MetaNet <concept_dir> <semantic_dir> <properties_input_dir> <output_dir>

Parameters:

<concept_dir>

<semantic_dir>

<properties_input_dir>

<output_dir>

MetaNetText

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.MetaNetText <input_dir> <output_dir> <metanet_thesauri>

Parameters:

<input_dir>

<output_dir>

<metanet_thesauri>

NamedEntity

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.NamedEntity <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

PStemmerMain

***

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.PStemmerMain <input_dir> <output_dir> <porter_language>

Parameters:

<input_dir>

<output_dir>

<porter_language>

PdfConverter

***

Classification: File Utility

Usage:

java -cp am3.jar edu.cmu.casos.automap.PdfConverter <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

PosTagger

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.PosTagger <input_dir> <output_dir> <POS_dir> <posType aggregate|ptb> <save_output_as csv|txt>

Parameters:

<input_dir>

<output_dir>

<POS_dir>

<posType aggregate|ptb>

<save_output_as csv|txt>

PositiveThesauri

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.PositiveThesauri <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

Properties

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.Properties <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

RemoveExtraWhite

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.RemoveExtraWhite [-t] [-n] <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

-t flag causes excess tabs to be removed

-n flag causes excess newlines to be removed

RemoveNumbers

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.RemoveNumbers <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

RemovePunct

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.RemovePunct <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

RemoveSym

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.RemoveSym <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

RemoveUserSym

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.RemoveUserSym <input_dir> <output_dir> <symbol_file>

Parameters:

<input_dir>

<output_dir>

<symbol_file>

SemanticList

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.SemanticList <input_dir> <output_dir> <window_size> <direction B|U> <window_reset>

Parameters:

<input_dir>

<output_dir>

<window_size>

<direction B|U>

<window_reset>

window_reset should be A, W=n, S=n, or P=n, n a positive integer

SemanticNet

Classification: Generation Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.SemanticNet <concept_dir> <semantic_dir> <properties_dir> <output_dir>

Parameters:

<concept_dir>

<semantic_dir>

<properties_dir>

<output_dir>

SpiderDriver

* * *

Classification: CEMap Component

Usage: java -cp am3.jar edu.cmu.casos.automap.SpiderDriver [options] <output_dir> <URL>

Parameters:

<output_dir>

<URL>

[options]

--documents

--max-size=n

--quota=size

--random-wait

--timeout=n

--tries=n

--user-agent=string

--wait=n

All numeric values must be nonnegative, a limit of 0 means no limit

SubmatrixSelection

Classification: Internal Command

Usage:

java -cp am3.jar edu.cmu.casos.automap.SubmatrixSelection <meta_list_dir> <types_file> <output_dir>

Parameters:

<meta_list_dir>

<types_file>

<output_dir>

ThesauriSort

Classification: Supplemental Procedure

Usage:

java -cp am3.jar edu.cmu.casos.automap.ThesauriSort <thesaurus_file> <output_file>

Parameters:

<thesaurus_file>

<output_file>

Threshold

Classification: Internal Command

Usage:

java -cp am3.jar edu.cmu.casos.automap.Threshold <input_dir> <output_dir> <threshold>

Parameters:

<input_dir>

<output_dir>

<threshold>

Union

Classification: Supplemental Procedure

Usage:

java -cp am3.jar edu.cmu.casos.automap.Union <input_dir> <output_dir> <union_file_name> <union_type>

Parameters:

<input_dir>

<output_dir>

<union_file_name>

<union_type>

UnionConcepts

Classification: Supplemental Procedure

Usage:

java -cp am3.jar edu.cmu.casos.automap.UnionConcepts <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

UnionKeyWordsInContext

Classification: Supplemental Procedure

Usage:

java -cp am3.jar edu.cmu.casos.automap.UnionKeyWordsInContext <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

Uppercase

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.Uppercase <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

UppercaseConcepts

Usage:

java -cp am3.jar edu.cmu.casos.automap.UppercaseConcepts <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

WhiteNumbers

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.WhiteNumbers <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

WhitePunct

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.WhitePunct <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

WhiteSym

Classification: Preprocessing Filter

Usage:

java -cp am3.jar edu.cmu.casos.automap.WhiteSym <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>

WordDocConverter

* * *

Classification: File Utility

Usage:

java -cp am3.jar edu.cmu.casos.automap.WordDocConverter <input_dir> <output_dir>

<input_dir> <output_dir>

WordList

Classification: Internal Routine

Usage:

java -cp am3.jar edu.cmu.casos.automap.WordList <input_dir> <output_dir>

Parameters:

<input_dir>

<output_dir>