AutoMap: Extracting usable information from unstructured texts

Michael W. Bigrigg  
bigrigg@cs.cmu.edu

Prof. Kathleen M. Carley  
kathleen.carley@cs.cmu.edu

AutoMap is a text mining tool that enables the extraction of network data from texts. AutoMap can extract four types of information:

- **content** (concepts, frequencies and meta-data such as sentence length)
- **semantic networks** (concepts and relationships)
- **meta-networks** (ontologically coded concepts and relationships – named entities and links)
- **sentiment and node attributes** (attributes of named entities)

Parts of speech tagging, machine learning techniques and proximity analysis enable computer-assisted Network Text Analysis (NTA). NTA encodes the links among words in a text and constructs a network of the linked words.

AutoMap is available as a stand-alone application as well as a module in *ORA & SORASCS.

AutoMap: Data Extraction

Unstructured Data: Texts

AutoMap: Data Extraction

CSV & DyNetML Output

Post Processing
- Event inference
- Belief inference
- Node attribute data fusion
- Lat-Lon data fusion

ORA: Analysis

Specialized Text analytics:
- Theme identification – LSA
- Hot Topics
- Semantic Network
- Parts of Speech
- Communicators
- Communicative Power
And – all social and dynamic network tools

Unstructured Data: Texts

AutoMap: Data Extraction

CSV & DyNetML Output

Post Processing
- Event inference
- Belief inference
- Node attribute data fusion
- Lat-Lon data fusion

ORA: Analysis

Specialized Text analytics:
- Theme identification – LSA
- Hot Topics
- Semantic Network
- Parts of Speech
- Communicators
- Communicative Power
And – all social and dynamic network tools

This work is part of the Dynamics Networks project at the center for Computational Analysis of Social and Organizational Systems (CASOS) of the School of Computer Science (SCS) at Carnegie Mellon University (CMU). Support was provided, in part, by the Office of Naval Research (ONR HSCB N000140811223, MMV - N00014-06-1-0104, MURI N000140811186) , the Army Research Institute (W911W07C0063) and CASOS. The views and conclusions contained in this document are those of the author sand should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research, the Army Research Institute or the U.S. government.

www.casos.cs.cmu.edu