Real-Time News Analysis (RTNA) for Improved Social Relationship Discovery

19-21 June 2007

Paper ID number: I-028

Janet F. O'May
janet.omay@us.army.mil
(410) 278-4998

Joan E. Forester
forester@us.army.mil
(410) 278-4977

Computational & Information Sciences Directorate
Army Research Laboratory
The U.S. Army's Corporate Laboratory


END
ZIONISM
World Crisis Web

Who is going to use the Real-Time News Analysis (RTNA) ?

U.S. ARMY RESEARCH LABORATORY

Army Soldier

Analyst

Scientist

Graphical User Interface

- **User directed input through the GUI**
  - Geographic region of interest
  - Characteristics of interest
  - Dates of interest
  - Key Words
  - News sources
  - Data target

- World
  - Americas
    - North America
      - United States
        - North East
          - Maryland
            - Harford County
              - Aberdeen
  - Asia
    - Middle East
      - Bahrain
      - Cyprus
      - Egypt
      - Iran
      - Iraq
        - Northern Region
        - Southern Region
      - Israel
      - Jordan
      - Kuwait
      - Lebanon
      - Oman
      - Qatar
      - Saudi Arabia
      - Syria
      - Turkey
      - UAE
      - Yemen

- Business
- **E**conomic
- Education
- Entertainment
- Health
- **I**nfrastructure
- **M**ilitary
- News
- **PMESII**
- **P**olitics
- Religion
- **S**ocial
- Sports
- Technology – Science
- User Defined

- All
- CBS
- CNN
- Fox
- Google
- MSNBC
- The ONION
- Wired
- User Defined

Find the news

- Google API
  - Actionable data - small amount but timely & meaningful
    - Soldier in the field
    - Commander
  - Reference data - a larger data set that is well filtered and preprocessed thus requiring more time
    - Analyst
    - ARL researcher
- Web crawler
  - All news stories
    - ARL researcher
      - Trend analyst
    - Relationship Discovery Service (RDS)

News Extraction

- **Extract/scrape/normalize news article**
  - Remove source-specific formatting
  - Remove other artifacts
- **Identify duplicate and near-duplicate news articles**
- **Save articles**
  - Text format
  - XML format

Multi-layered Knowledge Extraction

Text Mining – unstructured or semi-structured data sets

- Information extraction – (ThingFinder)
  - Identify key phrases & relationships within text
- Topic tracking
  - User directed
- Summarization - Used on lengthy documents
- Categorization
  - Word count
- Message Understanding System
  - Pattern-Matching
  - Syntax-Driven
  - Feature Selection
  - Semantics-Driven
- **Parsing**
- **Tagging**
- **Filtering**
- **Clustering**
- **Classifying**
- **Fusing**

Changing Intelligence Requirements

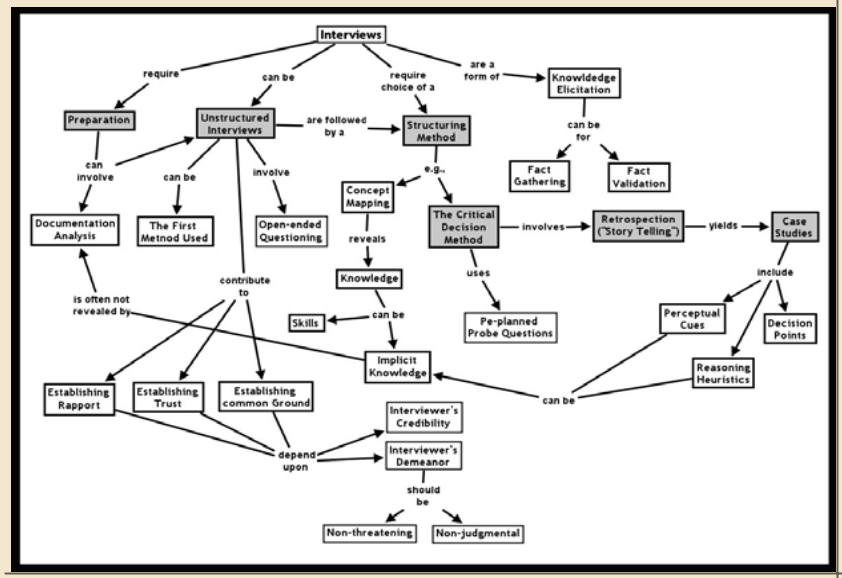
- INSCOM indicated interest in two sources of data:
  - Traditional data (SIGINT, MASINT, and HUMINT) obtained and processed in near real-time
  - Non-traditional data (financial, civil affairs, social context) obtained and processed
    - new methods of visualization
    - new methods to protect data
- Lt. Gen. Keith Alexander, Former Army G2 "said he is looking to industry and academia to help better organize and visually present information from multiple intelligence databases. " [1]
- Need to work with Intel Analysts to find out what we can do to make their job easier

[1] Article "Actionable Intelligence relies on every Soldier"
By Joe Burlas
Army News Service
http://www4.army.mil/ocpa/read.php?story_id_key=5847

Topics for the Intelligence Estimate

- Economics and psychology
  - The civilian population is passing through friendly lines in large numbers and taking refuge with friendly forces. They have little clothing, food, or medical supplies.
- Sociology
- Politics
- Science and technology
- Material
- Transportation
  - The civilian populace are using trucks, cars, oxcarts, and hand carts in their flight to friendly lines.
- Manpower
- Hydrography (electrical power)
- Population
- Religion

Concept Maps



Objective: Use cognitive maps to facilitate military tactical decision making

Description: Implement techniques to create and dynamically update military (tactical) situational concept maps

Approach: Apply mathematical graph theory to enable concept maps with weighted links usable as a guide for the deployment of mission assets

Impact

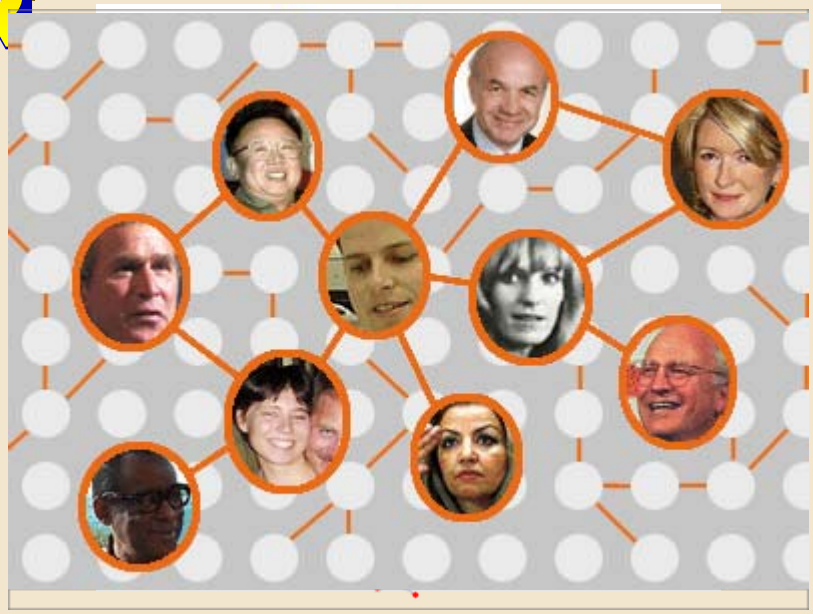
- Facilitates the determination of mission objectives
- Incorporates a cost-benefit profile useful in the development of actionable intelligence
- Provides improved cognitive understanding of a situation within a single diagram

Collaboration

- University of West Florida's Institute of Human and Machine Cognition (IHMC)
- U.S. Army Intelligence & Security Command (INSCOM)

U.S. ARMY RESEARCH LABORATORY

Social Relationship Discovery



- Based on Dynamic Network Analysis (DNA), a computational approach to modeling and simulating interactions among people, knowledge, resources, and tasks

- Evaluate currently available SNA software tools
- Research current algorithms
- Enhance existing algorithms for military intelligence applications
- Build interfaces to access software tools
- Improve visualization of SNA output
- Provide user directed queries

Collaboration

- Carnegie Mellon University's Center for Computational Analysis of Social and Organizational Systems (CASOS)

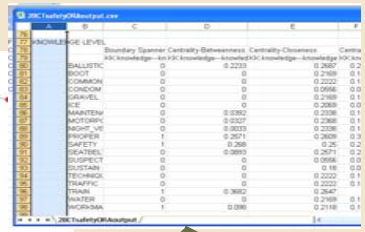
U.S. ARMY RESEARCH LABORATORY

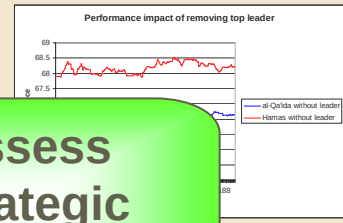
What Is Dynamic Network Analysis?

- A computational approach to modeling and simulating interactions among people, knowledge, resources, and tasks
- Dynamic Network Analysis (DNA) combines:
  - **Social network analysis**
  - **Link analysis**
  - **Multi-agent modeling**
- Applies to networks that are:
  - **Large**
  - **Multi-mode**
  - **Multi-link**
  - **Dynamic**
  - **Uncertain**
- Uses:
  - **Real world empirical data**
  - **Social, behavioral, organizational research findings**

U.S. ARMY RESEARCH LABORATORY

Texts





Build Network

Find Points of Influence

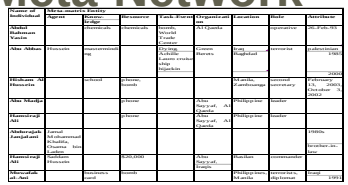
Assess Strategic Interventions

AutoMap
Automated extraction of network from texts
Visualizer

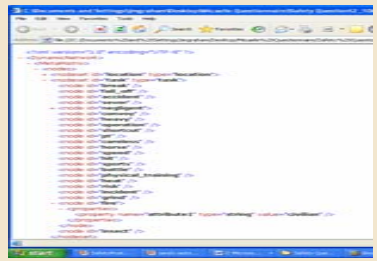
ORA
Statistical analysis of dynamic networks
Visualizer

DyNet
Simulation of dynamic networks
Visualizer

Meta-Network



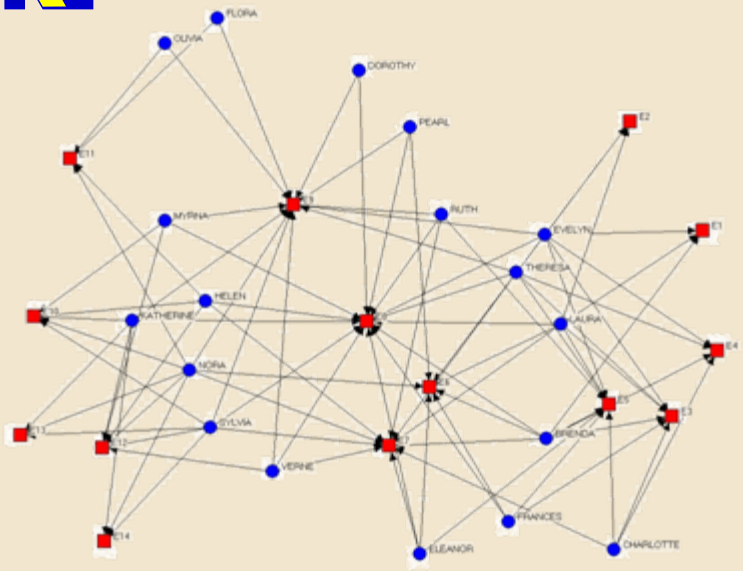
DyNetML



Unified Database(s)

CASOS

UCINET

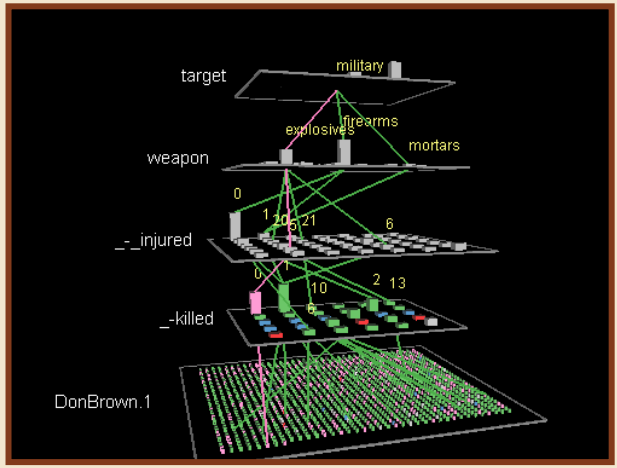


"UCINET is a comprehensive program for the analysis of social networks and other proximity data."

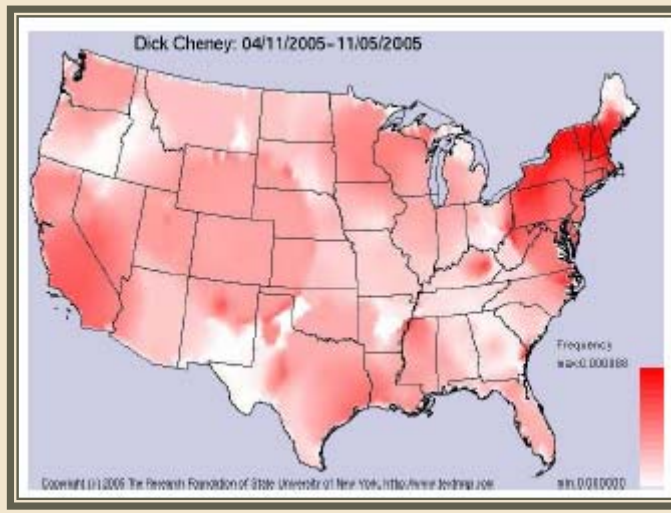
"UCINET contains network analytic routines, plus general statistical and multi-variate analysis tools such as multi-dimensional scaling, correspondence analysis, factor analysis, cluster analysis, multiple regression, etc."

- Information obtained from Analytic Technologies, Inc. website: http://www.analytictech.com/ucinet/ucinet_5_description.htm

Analyze & Visualize



- **Analyzing –** XML files or semi-structured data
  - Clustering
  - Data Mining
  - Fusing
  - Dynamic Network Analysis (DNA)
- **Visualizing**
  - Starlight – Pacific Northwest National Laboratory (PNNL)
  - Air Force Research Laboratory (AFRL) Dayton OH
    - Visualization of Social Network
  - Spatial Analysis of News Sources - Stony Brook University
  - Worldmapper – Sheffield University, UK



Future SNA Directions

- Develop software that will evaluate disparate data and select the “best” SNA software package to analyze the data through pre-defined heuristics
- Improve visualization techniques to provide the data to the Analyst quickly to improve analysis – new methodologies for data display

Wrap-up

U.S. ARMY RESEARCH LABORATORY

Visualize

Historical data

RAND CORPORATION

National Memorial Institute MIPT for the Prevention of Terrorism

UNIVERSITY OF MARYLAND

MINDSWAP

News extraction

Terror databases integration through an ontology

Dynamic PMESII Network Analysis

Multi-layered Knowledge Extraction

- Parse
- Tag
- Filter
- Mine
- Cluster
- Classify
- Fuse

Political, Military, Economic, Social, Infrastructure, Information (**PMESII**)

U.S. ARMY RESEARCH LABORATORY

Janet F. O'May

Operations Research Analyst

Tactical Collaboration & Data Fusion Branch

U.S. Army Research Laboratory
ATTN: AMSRD-ARL-CI-CT
Building 321, Office 17
Aberdeen Proving Ground, MD
21005-5067

Phone: **410.278.4998**
DSN: 298.4998
FAX 410.278.4988

janet.omay@us.army.mil

U.S. ARMY RESEARCH LABORATORY

Joan E. Forester

Operations Research Analyst

Tactical Collaboration & Data Fusion Branch

U.S. Army Research Laboratory
ATTN: AMSRD-ARL-CI-CT
Building 321, Office 15
Aberdeen Proving Ground, MD
21005-5067

Phone: **410.278.4977**
DSN: 298.4977
FAX 410.278.1996/4988

forester@us.army.mil