



*ISI 2010 Workshop on Current
Issues in Predictive Approaches to
Intelligence and Security Analytics
26 May, 2010*

Text analysis and entity extraction in
asymmetric threat response and prediction

Agenda

- ATRAP Overview
- Performing Entity Extraction
- Text Analysis in ATRAP
- Future Work



The Asymmetric Threat Response and Analysis Program (ATRAP)

Our Sponsor

- Supported by the U.S. Army Intelligence Battle Lab, Fort Huachuca, AZ
- Task Order Numbers:
 - 9T7ZDAIS705 (ATRAP 1)
 - 9T8ZDAIS803 (ATRAP 2)
 - 9Q9SDAIS903 (ATRAP 3)

Our Project Team

- University of Arizona
 - Dr. Jerzy Rozenblit, PI
 - Brian Ten Eyck, PM
 - Dr. Erwin Chan
 - Dr. Ferenc Szidarowski
 - Dr. Jonathan Sprinkle
- Ephemian, Inc.
 - System development
 - Database architecture
 - GUI Framework

ATRAP: the Intelligence Analyst's Cognitive Amplifier

➤ Battle Lab Objectives:

- Search and sort over large data sets
- Automate mundane data fusion tasks
- Visualize links, patterns, and predicted courses of action
- Augment commanders' situational awareness and decision-making capabilities

➤ Persistent Design Philosophy

- Analytic transparency
- Maximal user flexibility
- Prepare for integration into Operational Environment



Core ATRAP Functionality

- Search documents and entities
- Text Highlighter Tool
 - Review structured and unstructured text
 - Facilitates data entry, link assignment
- 2D Relationship Viewer
- 3D ThoughtSpace™
 - Geotemporal visualization environment
 - Full pan, zoom, rotate
 - “Play back” function over user-specified AOI, DTG
- Template Builder
 - Model threats, TTP's
 - Match to ingested data → predict courses of action



ThoughtSpace

The screenshot displays the ThoughtSpace software interface. The main window shows a map of the Caribbean region, including Florida, the Gulf of Mexico, and the Bahamas. A red diamond icon representing an interdicted vessel is highlighted. A pop-up window provides details for this interdicted vessel:

Interdiction
Maritime Interdiction
A1

Attributes

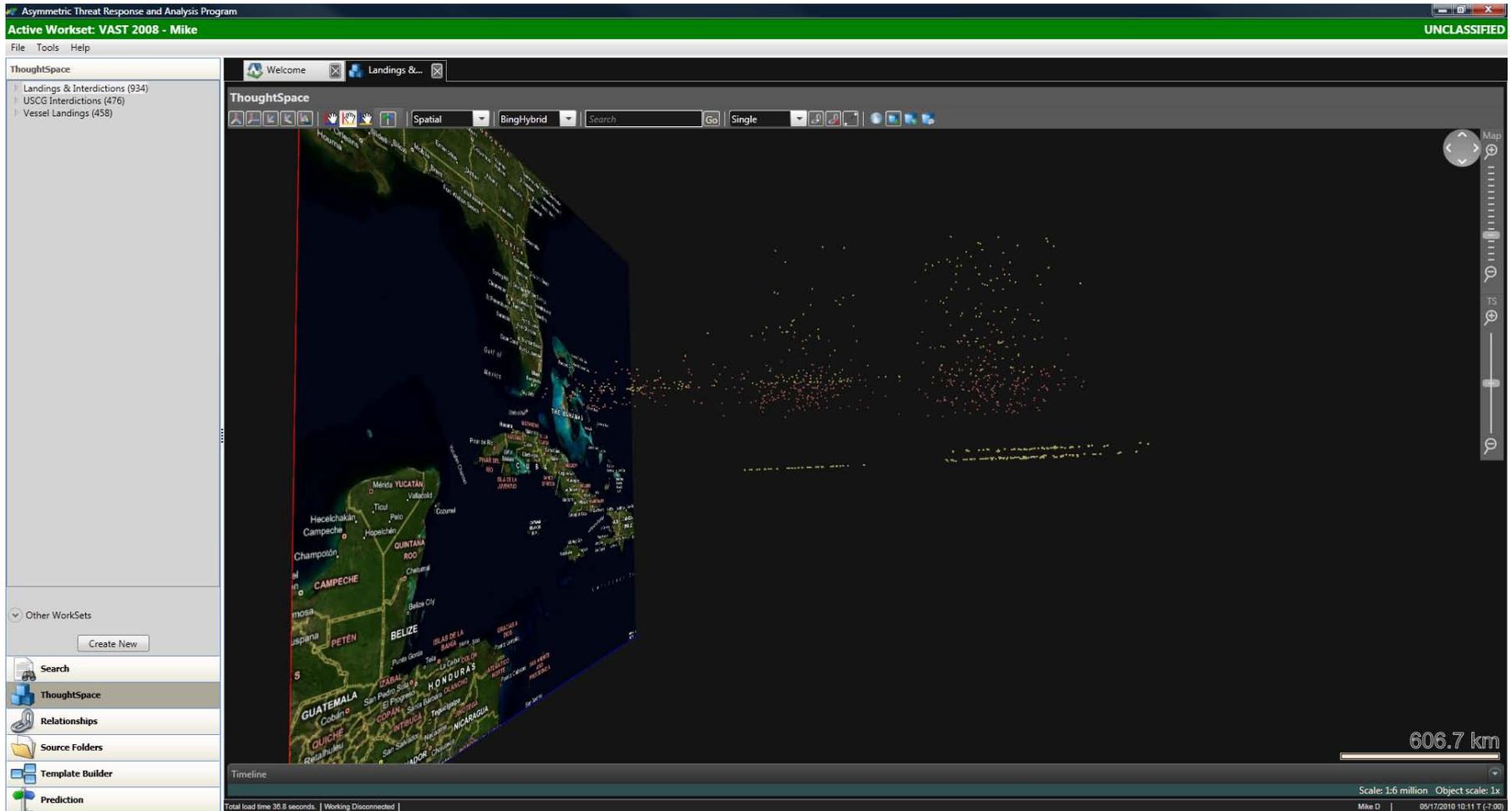
Passengers	18
Fatalities	0

Known Locations

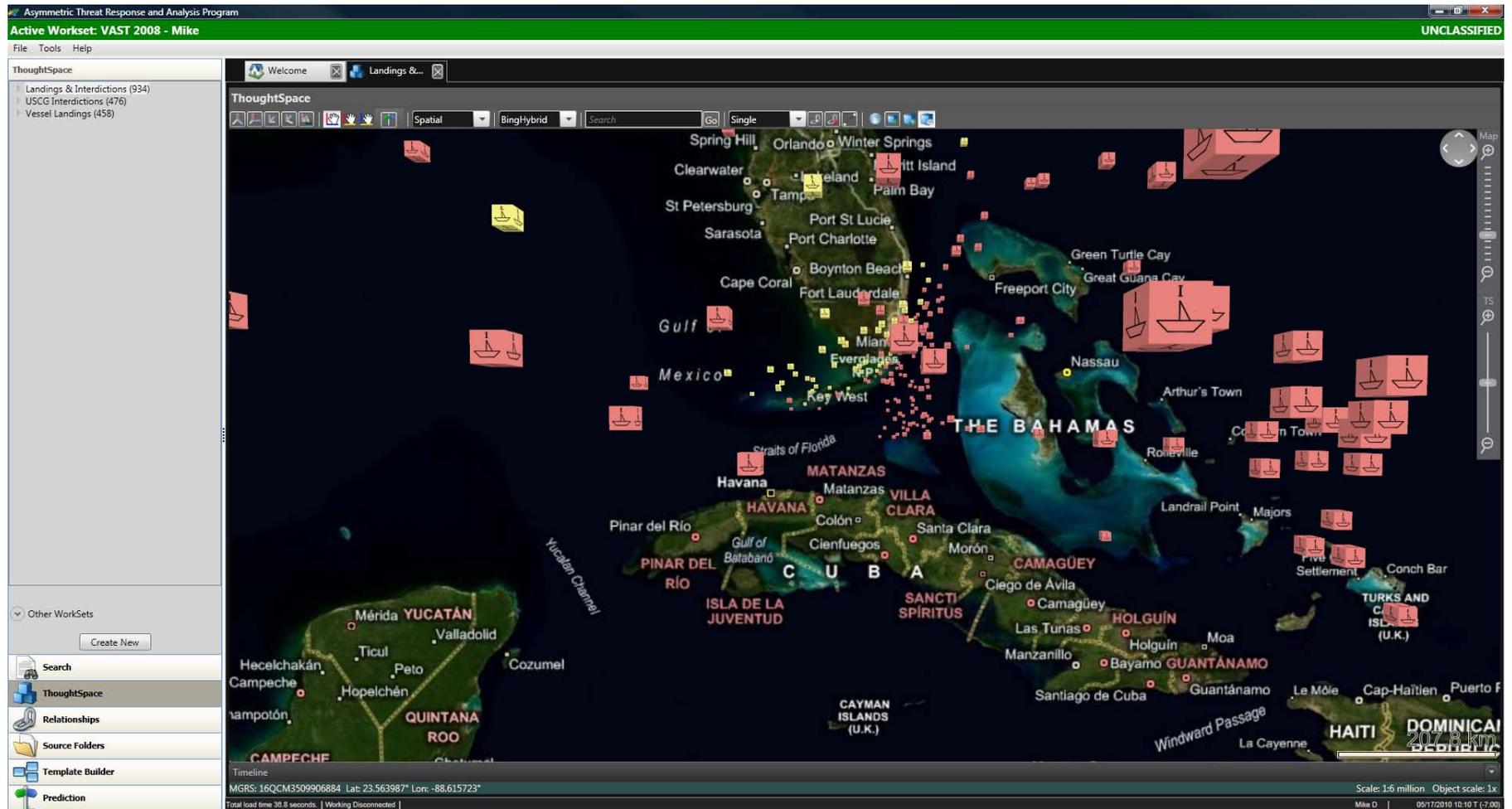
Time	Location	Last Modified
03/12/2005 17:00 T (-7:00)	17RNG427986128Z	05/11/2010 09:01 T (-7:00)

The interface also includes a timeline at the bottom showing a histogram of interdictions over time, with a scale of 1:6 million and an object scale of 1x. The status bar at the bottom indicates the total load time is 38.8 seconds and the system is working disconnected.

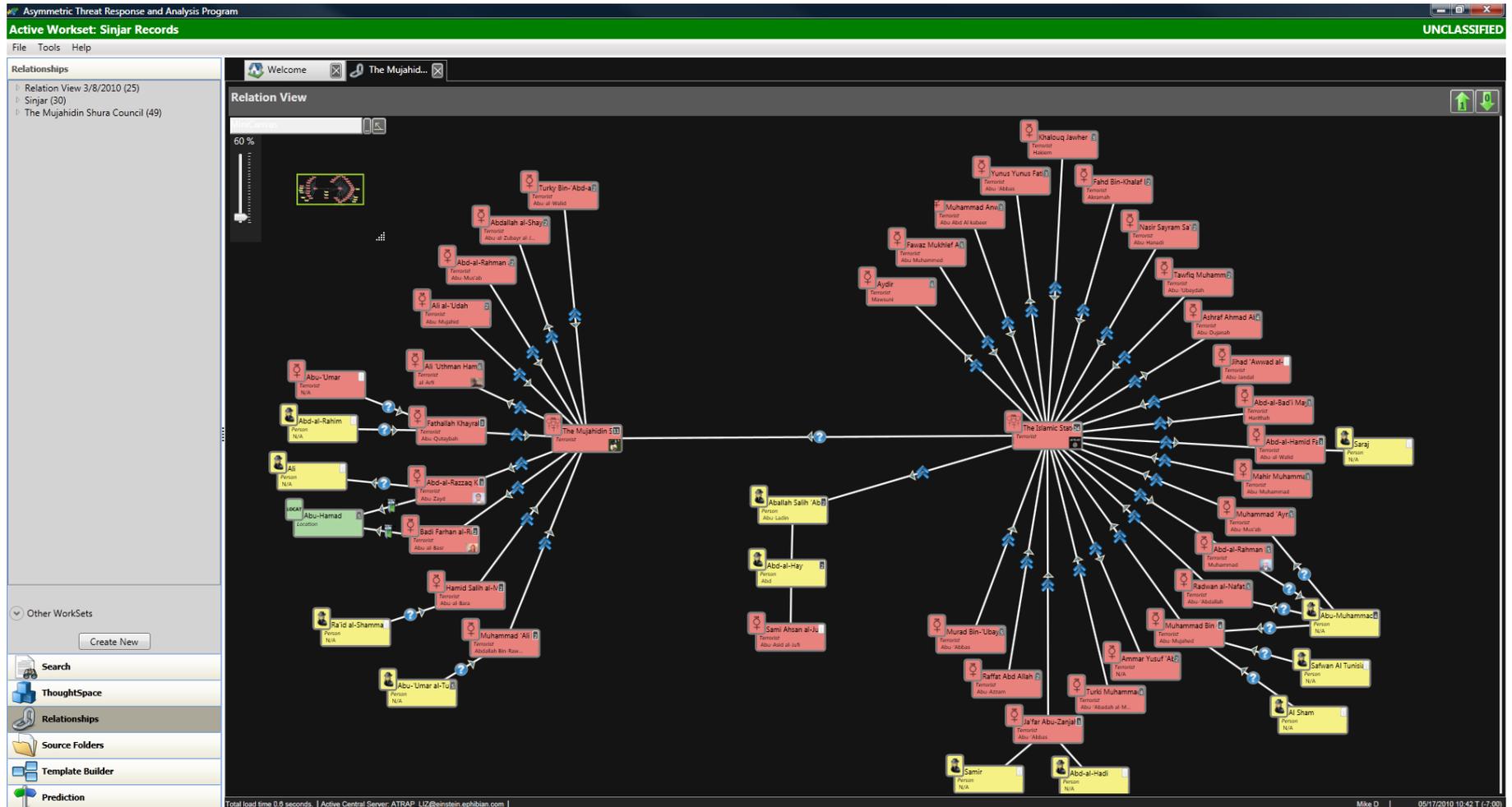
ThoughtSpace



ThoughtSpace



2D Relationship Viewer



Entity Recognition

- Common problem in Natural Language Processing
- Extract strings from text that refer to entities
- Entity types for ATRAP:
 - Person: *Mohammad al-Fahid*
 - Location: *Aleppo, Syria; US Embassy*
 - Organization: *Democratic Front for the Liberation*
 - Equipment: *Long Range Artillery, Prilutsky M1914*
 - Event: event keywords and predicate-argument relations

ATRAP Data Sets

- Sample corpora provided by U.S. army
 - HUMINT and TACREP documents
 - Scenarios used to train intelligence analysts
 - Time-stamped documents
 - 250,000 words total

- Corpus annotation for entities in progress:
 - 3,948 Persons
 - 5,298 Locations
 - 5,742 Organizations

Entity extraction: Person, Location, Organization

- First step: document classification
 - Is the document written in “standard” English?
 - Is the document written in “military” English?
 - USMTF – U.S. Military Text Formatted messages

- Subsequent step dependent upon type of text
 - Atypical capitalization patterns in military texts are problematic for statistical systems trained on standard written English,
 - But also make it easy to detect entities with simple pattern matching techniques

Capitalization in military documents

- Capitalization used to place emphasis on words
 - *Will Displaced Persons or Refugees impede operations in the vicinity of designated objectives?*
 - *Decontamination priority is: Maneuver, Routes, Fire Support, and Logistic assets.*

- Capitalization also indicative of names and locations
 - Alexander ABADJIAN
 - Rakhshan NEZAMI
 - TEHRAN, Ariana
 - MOSCOW, Donovia

Entity extraction: Person, Location, Organization

- “Standard” English: machine learning
 - Sequence classifier
 - Automatic Content Extraction (ACE) corpus
 - Conference on Natural Language Learning (CoNLL) shared task
- “Military” English:
 - (Large annotated corpora not available)
 - Pattern matching through regular expressions
 - Take advantage of peculiarities of writing style
- Additional procedures, e.g.:
 - Identify foreign names
 - Word lists to identify/filter entities (name list, common English vocabulary, names of building types, etc.)

Entity extraction: Person

- English: pattern match through regular expressions
 - (Title) (Firstname) (Middle Initial) Last Name
 - List of English names
- Foreign names
 - Classifier looks for non-English character trigrams in capitalized strings
 - *Bohdana* → “boh”, “ohd”, “hda” are atypical in English
 - Arabic name patterns:
Marwan Al-Shehhi, Ahmed Alhaznawi, Ibn Al-Shaykh al-Libi
- Identify lists of names
 - *Aksana Titov, Bohdana Abramstov, Cyzarine Nazarov, Fayina Byko, Galina Tikhomirov, Irina Fomin, and Kisa Leonov.*

Entity extraction: Locations

- Explicitly indicated in text :
 - *Country of Origin: Minaria*
 - *LOC: Kemalia*
 - *POB Fizuli, Arzak*
- Noun phrases ending with location words
 - *building, fort, apartment, river, embassy, etc.*
 - May begin with locational modifiers:
 - Indicated by suffix of word
 - **Donovian**, **Chinese**, **Minarian**, **Pakistani**
- Addresses
- Lists of geographic locations

Entity extraction: Organizations

- Explicitly indicated in text :
 - *Organization: Formasi Qan Usyanci (FQU)*
 - *Affiliation: Sokolov Corporation*

- Utilize lists of organization types:
 - *committee, organization, cartel, mafia, etc.*
 - May begin with locational modifiers
 - e.g. *Donovian Mafia*

Evaluation: Person

- Compare to existing named entity recognition software
 - Sparse Network of Windows classifier, trained on news text from CoNLL-2003 shared task
- Apply to ATRAP data for Person evaluation
 - (Location, Organization annotation in progress)
- Results for pattern matching only:

TABLE I. PERFORMANCE OF ATRAP AND SNOW PERSON ENTITY RECOGNIZERS ON ATRAP DATA

	Precision	Recall	F-measure
ATRAP	0.843	0.663	0.743
SNOW	0.346	0.489	0.405

Evaluation: Location, Organization

- No large, fully annotated corpus available on military/terrorism domain
 - Test on MUC-3 data: Latin American terrorism
 - Caveat: MUC answer keys indicate a *subset* of the entities in documents (recall scores less meaningful)
- Precision scores for pattern-matching only:

		ATRAP	SNoW
Precision	PER	.893	.988
	ORG	.878	.922
	LOC	.911	1.000

Entity extraction: events (work in progress)

➤ “The Syrian Army allegedly killed over 100 Syrian and Lebanese smugglers.”

➤ Want to extract a logical proposition:

killed(Syrian Army, 100 Syrian and Lebanese
smugglers)

Entity extraction: events (work in progress)

- Step one:
 - Automatically construct domain-specific lexicons based on keywords and comparison to large databases of documents
 - Discover words relevant to domain: e.g. *bombed, killed, detonate, attack, destroy*
- Step two:
 - Develop a classifier for discovering arguments of predicates
 - Utilize ACE-2005 annotated corpus
 - Sentence: *The Syrian Army allegedly killed over 100 Syrian and Lebanese smugglers.*
 - Event: *killed*
 - Subject: *The Syrian Army*
 - Target: *over 100 Syrian and Lebanese smugglers.*

Highlighter Tool: Unprocessed Document

Asymmetric Threat Response and Analysis Program
Active Workset: Sinjar Records
UNCLASSIFIED

File Tools Help

Sources

ar
'Abd-al-'Aziz Muhammad al-Jahni.doc (820 KB)
'Abd-al-Rahman al-Fayturi Hamad al-Zaytuni.doc
Ahmad Bin-Muhammad Muhammad al-Jahni.doc
Ahmad Bin-Na'ili Diraghtish.doc (4 MB)
'Abd-al-Bad'i Maylud Muhammad.doc (34 KB)
Ahmad Bin-Shubat Bin-'Abd-al-Rahman al-Zufa
'Abd-al-Nasir Haws Biliqasim al-Saghir.doc (815
al-Raway'i Salim Sa'id al-Sa'iri.doc (815 KB)
'Abd-al-Hakim 'Umar Tahir.doc (1224 KB)
'Ali Ahmad 'Ali Hammud.doc (816 KB)
'Awad Fayhan 'Awad al-'Utaybi.doc (815 KB)
'Abdallah Bin-Salim Bin-Muhammad al-Dusari.doc
Adam Shabraq Hamid.doc (2 MB)
'Ali al-'Udah.doc (62 KB)
'Ali Bin-Muhammad al-Madayyan.doc (815 KB)
'Ali Bin-Hasan Bin-'Ali al-Dhiyi.doc (815 KB)
Ahmad 'Ali Musa.doc (5 MB)
'Abd-al-Rahman Bin-'Ali Bin-Muhammad al-'Asi
'Amir Qasim Muflih al-Jardat.doc (5 MB)
'Abdallah Awlad al-Tumi.doc (819 KB)
'Abdullah Sami Ahmad Muhammad.doc (5 MB)
Aydir.doc (34 KB)
Ahmad 'Ali Sa'id.doc (814 KB)
'Abd-al-Rahim Bin-Hamdun.doc (816 KB)
Awass 'Ali 'Abd-al-Karim.doc (36 KB)
Abd-al-Hamid Fahim 'Abdallah al-Turabli.doc
Ahmad Farhat Ahmad.doc (8 MB)
'Abd-al-Kamil Bin-Ahamad Musawid.doc (5 MB)
'Aballah Salih 'Abd-al-Karim Ibrahim.doc (814 K
'Abd-al-Rahman Bin-'Ali al-Rajhi.doc (37 KB)
'Abd-al-Rahman Wa'il 'Ammash al-Harbi.doc (5
Ali 'Uthman Hamad al-'Arfi.doc (59 KB)
'Abd-al-'Aziz Bin-'Atiq Salih al-'Atiq.doc (7 MB)
'Ali 'Abdallah 'Ali al-Ghamidi.doc (6 MB)
'Aballah Salih 'Abd-al-Karim Ibrahim.doc (814 K
'Abd-al-'Aziz 'Abd-al-Rabb Salih al-Yafi'.doc (3
exxonmobil-profit-jumps-oil-price.htm (10 KB)
Amin Faysal Ahmad al-'Urayfi.doc (815 KB)
Amin 'Ali Haxan al-Ahmadri.doc (2 MB)

Other WorkSets

Search
ThoughtSpace
Relationships
Source Folders
Template Builder
Prediction

Welcome 'Abd-al-'Azi...

Data Extraction Tool

Add Mode Highlight to Add Add Type Entity ThoughtSpace Not Selected (null) Quick Add Off

'Abd-al-'Aziz Bin-'Atiq Salih al-'Atiq.doc

The Mujahidin Shura Council]

Personal Information Form

Name:	'Abd-al-'Aziz Bin-'Atiq Salih al-'Atiq
Alias:	Abu-Suhayb
Address:	Riyadh, Saudi Arabia
Phone Numbers:	0096612750665 home 00966555246937 father
Date of Birth:	16 JUL 1983
Coordinator:	
Occupation in Homeland:	
Duty:	Journalist
Items Entrusted:	A watch + \$200 + 4000 Liras + Passport
Date of Entry:	10 NOV 2006
Route of Entry:	Abu-Hamad
Assigned to:	
Last Will and Testament:	
1-	
2-	
3-	

Document Media

Document Information

Entity

Relationship Picker

Document Media

Total load time 12.1 seconds | Active Central Server: ATRAP_LIZ@einstein.ephbian.com | Mike D | 05/17/2010 10:19 T (7:00)

Highlighter Tool: Shared Mark-ups

The screenshot displays the ATRAP Highlighter Tool interface. The main window shows a document titled 'Abd-al-'Aziz Bin-'Atiq Salih al-'Atiq.doc'. The document content includes a photograph of a man, a flag icon, and the text 'The Mujahidin Shura Council]'. Below the text is a 'Personal Information Form' table with the following data:

Name:	Abd-al-'Aziz Bin-'Atiq Salih al-'Atiq
Alias:	Abu-Suhayb
Address:	Riyadh, Saudi Arabia
Phone Numbers:	0096612750665 home 00966555246937 father
Date of Birth:	16 JUL 1983
Coordinator:	
Occupation in Homeland:	
Duty:	Journalist
Items Entrusted:	A watch + \$200 + 4000 Liras + Passport
Date of Entry:	10 NOV 2006
Route of Entry:	Abu-Hamad
Assigned to:	
Last Will and Testament:	
1-	
2-	
3-	

The right sidebar shows 'Document Information' with a list of entities: Mike D, Abu-Hamad, Riyadh, Saudi Arabia, Bryan M, Abd-al-'Aziz Bin-'Atiq Salih al-'Atiq, and The Mujahidin Shura Council. The status bar at the bottom indicates 'Total load time: 13 seconds' and 'Active Central Server: ATRAP_L12@beinstein.ephilbian.com'.

Highlighter Tool: Entity Dossier

The screenshot displays the ATRAP (Asymmetric Threat Response and Analysis Program) interface. The main window is titled 'Active Workset: Sinjar Records' and shows a list of document sources on the left. The central pane displays a document titled 'Abd-al-Rahman Ahmad Nasir Khatir.doc' with a photo of a man and a highlighted text snippet: '[The Mujahidin Shura Council]'. Below the document is a 'Personal Information Form' with the following data:

Name:	Abd-al-Rahman Ahmad Nasir Khatir
Alias:	Abu-Yasir
Address:	Az Zarqa', Jordan
Phone Numbers:	House 0096253994028
Date of Birth:	18 FEB 1975
Coordinator:	Bakr
Occupation in Homeland:	
Duty:	Martyr
Items Entrusted:	32 Jordanians [TC: Jordanian Dinars] + 200 SL + passport
Date of Entry:	02 NOV 2006
Route of Entry:	Abu-Hamad
Assigned to:	
Last Will and Testament:	
1-	
2-	
3-	
4-	
5-	

The right-hand pane shows a detailed 'Entity Dossier' for 'Abd-al-Rahman Ahmad Nasir Khatir'. It includes fields for Name, Affiliation (Unknown), Entity Type (Person), Reliability (A 1), and various attributes such as DOB (02/18/1975 07:00), Description (32 Jordanians, Martyr), Alias (Abu-Yasir), Address (Az Zarqa', Jordan), and Phone number (House 009625399402). It also lists Aliases (Abu-Yasir, 05/17/2010 10:26 T (-7:00)), Media (a sunset image), and Known Locations (Tabular, Map).

At the bottom of the interface, there is a status bar showing 'Total load time 2.6 seconds', 'Active Central Server: ATRAP_LIZ@enstein.ephilban.com', and the user 'Mike D' on '05/17/2010 10:30 T (-7:00)'.

Highlighter Tool: Creating Relationships

The screenshot displays the ATRAP (Asymmetric Threat Response and Analysis Program) interface. The main window is titled "Active Workset: Sinjar Records" and shows a "Data Extraction Tool" with a list of documents on the left and a central view of a document titled "Abd-al-Rahman Ahmad Nasir Khatir.doc". A photograph of a man is visible in the document view. A yellow highlight is placed over the text "The Mujahidin Shura Council" in the document. A "Personal Information Form" is displayed below the document, with fields for Name, Alias, Address, Phone Numbers, Date of Birth, Coordinator, Occupation in Homeland, Duty, Items Entrusted, Date of Entry, Route of Entry, and Assigned to. A "Create Relationships" dialog box is open, showing the process of relating "The Mujahidin Shura Council" to "Abd-al-Rahman Ahmad Nasir Khatir". The dialog box includes a "Type" dropdown set to "Associates, Generally", a "Description" field, and radio buttons for "Confirmed" and "Suspected" relationships. The "Start Date" and "End Date" are both set to "05/17/2010 10:27". A "Relationship Picker" window is also visible on the right, showing a list of entities and their relationships. The interface includes a menu bar (File, Tools, Help) and a status bar at the bottom.

Active Workset: Sinjar Records

File Tools Help

Sources

Sinjar

- 'Abd-al-'Aziz Muhammad al-Jahni.doc (8 MB)
- 'Abd-al-Rahman al-Fayturi Hamad al-Za' Ahmad Bin-Muhammad Muhammad al-Ahmad Bin-Na'il Diraghtah.doc (4 MB)
- 'Abd-al-Bad'i Maylud Muhammad.doc (3 MB)
- Ahmad Bin-Shubat Bin-'Abd-al-Rahman 'Abd-al-Nasir Haws Bilqasim al-Saghir.doc
- al-Raway'i Salim Sa'id al-Sa'iri.doc (815 KB)
- 'Abd-al-Hakim 'Umar Tahir.doc (1224 KB)
- 'Ali Ahmad 'Ali Hammud.doc (816 KB)
- 'Awad Fayhan 'Awad al-'Utaybi.doc (815 KB)
- 'Abdallah Bin-Salim Bin-Muhammad al-F Adam Shabraq Hamid.doc (2 MB)
- 'Ali al-'Udah.doc (62 KB)
- 'Ali Bin-Muhammad al-Madayyan.doc (8 MB)
- 'Ali Bin-Hasan Bin-'Ali al-Dhiyri.doc (815 KB)
- Ahmad 'Ali Musa.doc (6 MB)
- 'Abd-al-Rahman Bin-'Ali Bin-Muhammad 'Amir Qasim Muflih al-Jardat.doc (6 MB)
- 'Abdallah Awlad al-Tumi.doc (819 KB)
- 'Abdullah Sami Ahmad Muhammad.doc Aydir.doc (34 KB)
- Ahmad 'Ali Sa'id.doc (814 KB)
- 'Abd-al-Rahim Bin-Hamdun.doc (816 KB)
- Awas 'Ali 'Abd-al-Karim.doc (36 KB)
- Abd-al-Hamid Fahim 'Abdallah al-Turabi Ahmad Farhat Ahmad.doc (8 MB)
- 'Abd-al-Kamil Bin-Ahamad Musawi.doc
- 'Aballah Salih 'Abd-al-Karim Ibrahim.doc
- 'Abd-al-Rahman Bin-'Ali al Rajihi.doc (37 KB)
- 'Abd-al-Rahman Wa'il 'Amash al-Harbi Ali 'Uthman Hamad al-'Arfi.doc (59 KB)
- 'Abd-al-'Aziz Bin-'Atiq Salih al-'Atiq.doc
- 'Ali 'Abdallah 'Ali al-Ghamidi.doc (6 MB)
- 'Aballah Salih 'Abd-al-Karim Ibrahim.doc
- 'Abd-al-'Aziz 'Abd-al-Rabb Salih al-Yafi', exxonmobil-profit-jumps-oil-price.htm
- Amin Faysal Ahmad al-'Urayfi.doc (815 KB)
- Amin 'Ali Hasan al-Ahmedi.doc (7 MB)

Other WorkSets

Search

ThoughtSpace

Relationships

Source Folders

Template Builder

Prediction

Welcome

'Abd-al-'Azi...

'Abd-al-Rah...

Aydir.doc

Data Extraction Tool

Add Mode: Highlight to Add

Add Type: Entity Attribute

ThoughtSpace: Not Selected (null)

Quick Add: Off

'Abd-al-Rahman Ahmad Nasir Khatir.doc

Relationship Picker

-- Relate --

Abd-al-Rahman Ahmad Nasir Khatir

The Mujahidin Shura Council

Ashraf Ahmad Abu-Bakr al-Hasan

Abu-Hamad

Abd-al-Rahman Ahmad Nasir Khatir

The Mujahidin Shura Council

Ashraf Ahmad Abu-Bakr al-Hasan

Abu-Hamad

Personal Information Form

Name: Abd-al-Rahman Ahmad Nasir Khatir

Alias: Abu-Yasir

Address: Az Zarqa', Jordan

Phone Numbers: House 0096253994028

Date of Birth: 18 FEB 1975

Coordinator: Bakr

Occupation in Homeland:

Duty: Martyr

Items Entrusted: 32 Jordanians [TC: Jordanian Dinars] +

Date of Entry: 02 NOV 2006

Route of Entry: Abu-Hamad

Assigned to:

Last Will and Testament:

1-

2-

3-

4-

5-

Create Relationships

Relate The Mujahidin Shura Council to Abd-al-Rahman Ahmad Nasir Khatir

Swap

Type: Associates, Generally

Description:

Relationship Is: Confirmed Suspected

Start Date: 05/17/2010 10:27

End Date: 05/17/2010 10:27

Ok Cancel

Delete This Relationship

Document Information

Entity

Relationship Picker

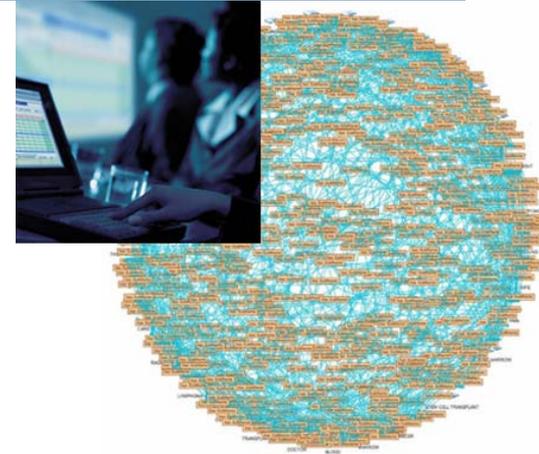
Document Media

Total load time 2.6 seconds. | Active Central Server: ATRAP_LIZ@einstein.ephilban.com

Mike D 05/17/2010 10:29 T (-7:00)

Ongoing Research and Development

- Enhanced NLP Techniques
- Enhanced Behavioral Models
- Game-Theoretic Decision Modeling
- Data Provenance
- Multi-Level Data Security



ASYMMETRIC THREAT RESPONSE AND ANALYSIS PROGRAM

Dr. Jerzy Rozenblit
Principal Investigator
jr@ece.arizona.edu

Brian Ten Eyck
Project Manager
bteneyck@ece.arizona.edu

Dr. Erwin Chan
Computational Linguistics
echan3@u.arizona.edu