Open Source Intelligence

C. H. Best

Joint Research Centre

European Commission
Acknowledgements

• Europe Media Monitor team

• http://press.jrc.it
Programme

1. Open Source Intelligence
   • Some Definitions
   • Not just text - satellite examples

2. Diversion: The Web and the Dark Web
   • Web dynamics
   • Terrorists use of the Web

3. Web Mining and Intelligence
   • News Monitoring
   • News Analysis / Visualisation

4. OSINT Tools
Open-Source Intelligence

- Term Originates from Security Services
- 9/11 Shock to the system of old style intelligence
- “Failure to connect the dots.” Agency secrecy, mistrust, elitism. 9/11 Commission Report
- Terrorists skilled use of internet was an eye opener.
- “To fight a network like Al Qaeda you need to behave like a network”, John Arquillia
- Connected world: Business intelligence
Growth of Open Source Intelligence

• Pre 9/11 intelligence services were closed and relied on Humint, Sigint and classified information
• Realisation that open source could have foreseen attacks -> reassessment in use of OS, & in sharing intel between agencies.
• Now - Active monitoring and tracking of terrorist activities on the Internet – most national police and security services
• Importance of on-line news for governments and international organisations. 24 hour CNN news effect.
• EUROSINT Forum 2006, ISI conferences 2004
• NATO OS handbook, US Army OS manual, NATO OS Training
OSINT - definitions

- Open Source Information (OSINF) is data which is available publicly – not necessarily free.
- Open Source Intelligence (OSINT) is proprietary intelligence recursively derived from OSINF.
- OSINF Collection is monitoring, selecting, retrieving, tagging, cataloging, visualising & disseminating data.
- OSINT is the result of expert analysis of OSINF.
Intelligence Sources

- HUMINT
- SIGINT
- OSINT

Classified

Public
Who uses OSINT

- Security Services, Law Enforcement etc.
- Governmental Organisations
  - EU, NATO, AU Situation Centre
  - IAEA – Nuclear Safeguards
  - UN DPK, WHO, NGOs
- Large Companies
  - Oil Companies
  - Multinationals
- Business Intelligence
What is intelligence anyway?

- An Intelligence product is nearly always a report.
  - Derived from collected information
  - Not just text
  - Maps, images, data
  - both from classified and open source information

- A report derived only from Open Sources may be shared with others
  - Inter-agency collaboration
Typical Agency, Company, UNHQ, Council SitCen etc.

- Open Source Products
  - News Flash alerting – news monitoring
  - Daily Current Awareness Briefings – news monitoring
  - Situation updates – news and field reports
    - Darfur, Iraq weekly summary
  - Specific Studies – Search Engines, Subscription databases, news archives
    - Iraq WMD,
    - 9/11 Commission Report
    - Radicalisation
Who is involved?

Tool Builder

OSINF Collector/researcher

Analyst

Minister/General/Commissioner/CEO

We are here

Classified Information
Information Overload

Data
≈ 10,000 messages per day
(IMINT, SIGINT, HUMINT, Intel Reports)

Data Search
Analyst creates a profile to sort messages

Extraction
Analyst reads selected messages to extract facts

Interpretation
Analyst builds a link analysis model

Conclusion
Analyst forms an opinion and authors a report

Report

The system sorts messages by key words -- limited precision.
The analyst can study ≈ 200 messages per day.
The analyst manually builds (at best) partial, static link models.
The final output is text: a message, web page or document.
OSINT Cycle

Workflow data management tools
OSINT Processes

Collect → Process → Analyse → Visualise → Collaborate

- Collect
  - Multilingual Inf. Retrieval
  - Search
  - Crawler
  - News feeds
  - Machine Trans.

- Process
  - Inf extraction
  - Entities
  - Places
  - classify

- Analyse
  - Link analysis
  - Relationships
  - Geolinking
  - Trends
  - Statistics

- Visualise
  - Networks
  - Relations
  - Time graphs
  - Maps

- Collaborate
  - Intel wiki
  - IM
  - Case DB
  - Publish
Open Source Information Sources

• The Internet - 80%
• Subscription services
  – Lexis Nexis, News Agencies
  – Factiva, Dialog, BBC Monitoring, Janes
• Commercial Satellites
  – Space Imaging
• Geospatial Data Providers.
Features of OSINF

- Information Overload – must be selective
- Language – Machine translation, multi-lingual tools
- Sources – important to identify, rank reliability. Information comes first – source second. Opposite to Humint.
- Impact. Speed - rumor, falsities, propaganda, events fly across the world creating mayhem– eg. Danish cartoons
- Sharing information. This is essential to link together small signals and discover inter-connections. Goes against secrecy.
- Copyright and legality minefield
OSINF is NOT JUST TEXT - image acquisition & analysis And Data Fusion

Example: High Resolution Commercial Satellite

- IKONOS - 1 m resolution
- Quickbird - 60cm resolution.
- Sample Applications:
  - Intelligence - Iran Nuclear Programme
  - Damage assessment – Lebanon
  - Image analysis – refugee camps
Image intelligence
Monitoring Iran Nuclear Programme

- Combination of open source data and high resolution satellite images
- Data mining by EMM and conventional search engines
- Expert knowledge of the nuclear fuel cycle
- Independent assessment for EU policy makers
- IAEA perform similar studies
Traditional industries of Masuleh fading away

TEHRAN, Oct. 10 (MNA) -- New anthropological studies of Masuleh will be carried out soon, Gilan Cultural Heritage and Tourism Department anthropologist Saeid Hosseini announced on Monday.

Art news in brief

Classical singer’s new album coming out soon

TEHRAN, Oct. 10 (MNA) -- Iranian classical singer Alireza Eftekari is set to release his new album entitled “Guide of Balkh”, which is centered around verses by Molana Rumi. Master Jalal Zolfonun plays setar on the album. Eftekari also worked on a music video produced by musician Fereydun Shabazian on the theme of Sacred Defense which was recently broadcast on Iranian TV.

East Azerbaijan town may hold apple festival

TEHRAN, Oct. 10 (MNA) -- Experts believe that the garden city of Zenuz could attract hundreds of tourists if an apple festival was organized in the city.
گزارش تصویری/ بازدید اعضای کمیسیون انرژی مجلس از کارخانه فراوری اورانیوم (۱)

اعضاء کمیسیون انرژی مجلس شورای اسلامی عصر امروز از کارخانه فراوری اورانیوم اصفهان یکی از کرک‌های این نیروگاه نورث‌کرکس به‌دست آوردند.
Damage Assessment with Satellite Imagery

JRC Beirut

JRC & EUSC: South Lebanon

802 km² or 89% of the total area of S. Lebanon
- 326 residential buildings affected in southern suburbs of which 60% completely destroyed, mostly (88%) high buildings with more than 5 floors
- Main area affected: Haret Horaik: 15% of its buildings destroyed or damaged
- Population impacted: up to 30,000 people
Example 1: Destroyed residential buildings (Beirut)
Beirut - Damage to infrastructure

- Damaged highway sections and overpasses (6)
- Damaged industrial buildings (8)
- All airport runways damaged
Web Mining & Intelligence

First Diversion : Understanding Web Trends
Evolving communications

Phone: one to one

Broadcast: One to many

Web 1.0: many to many

Web 1.5: User participation

1920

1960

1993

1996
Web 2.0 – convergence of 3 things

Easy user publishing

Easy syndication - RSS

Interactivity – Ajax/DHTML/Special Client

2002

But really ……
So-called Web 2 enabled by widespread deployment of broadband

Billions of Internet users

Regional Share Digital Information

Total On-line Digital Content $1.6 \times 10^{20}$ bytes

2010 – 70% of content will be from users

Source: IDC
However - still only 10 users per web server!

Information jungle

2007-100 million web servers

Indexed by search engines

Hidden Web

So

web is evolving from passive surfing to active publishing

Source - Netcraft
Explosion in User publishing

Social Network Sites
- YouTube: 120 million clips daily viewed
- MySpace > 116 million users in 3Y
- Photo sharing, Flickr, webshots

Wikipedia Phenomenon
- 1.4 Million articles english
- 112 languages, 300K authors
- Collaborative accreditation works!

Blogs

50 Million Weblogs Tracked (July, 2006)
Doubling in size about every 6 months for the past 3 years.

The Blogosphere is over 100 times bigger than it was 3 years ago, when Technorati was incorporated.

June 2006 - Posts by Language
- English 32%
- Spanish 11%
- Japanese 31%
- Russian 5%
- German 5%
- Dutch 1%
- Other 5%
Article Count for the Ten Largest Wikipedias

Source: http://stats.wikimedia.org/EN/TablesArticlesTotal.htm
Cultural Variations
Future Social Change

Today’s “geeks” → Tomorrow=’s customer, user, voter, elderly….
Second Diversion: Dark Side of the Web

- Similar growth - easy publishing
- Racist and extremist propaganda
- Al Qaeda could not function without the Internet

- Acknowledge Jael Shahar
NEW YORK, April 18 — Al-Qaida’s apparent use of computers and inexpensive editing software to put together videotapes that emerged this week came as no surprise to U.S. intelligence officials, who told NBC News that laptops and high-speed Internet access had become terrorists’ primary tools for communicating over the past several
New Challenges....

• New technologies have changed the way terrorist organizations operate. The Global Jihad relies on global communications for:
  – Command and Control
  – Communications
  – Propaganda & PsyOps
  – Recruitment
  – Fundraising
  – Datamining
New Responses …

• At the same time, terrorist use of new technologies presents new opportunities for intercepting and disrupting terrorist activities.
The Internet as a Tool for Command & Control

• The same advantages the Internet brings to the general public and to business are useful to international terrorist groups:
  – speed
  – security
  – global linkage
  – anonymity
The Internet as a Tool for Command & Control

Secure Communications
• Public encryption is available to anyone, allowing secure communication.

Terrorist groups can thus be geographically dispersed and non-hierarchical
• Networks have proven capable of defeating much more powerful hierarchical institutions.
The Internet as a Tool for Command & Control

“Leaderless Resistance”
- The Internet supports the spread of a unifying ideology without any visible leadership.

Non-state actors
- Terrorist organizations can flourish independently of state sponsors, thus removing the restrictions imposed by the vulnerability to retaliation.
Secure Communication

- **Cryptography**
  - Changes the contents of a file or message so that it is unreadable by everyone except the intended recipient.
  - Encrypted messages are not hidden; their comings and goings can be detected and monitored.
  - Once the means of encryption have been revealed, the code breaker must uncover the key to decrypt the message.

- **Steganography** ("covered writing")
  - The practice of hiding messages within pictures, text, and objects.
  - Dates back to ancient Greek times.
  - As applied to modern communications systems, it has become a very effective form of sending imperceptible messages.
  - Unlike encrypted messages sent via email, steganography does not provide a trail to the sender or the recipient.
Simpler Internet Tricks

- Two terrorists on opposite sides of the globe agree to open 30 anonymous web-based e-mail accounts with 30 different passwords.
- On the first of the month the first account is used, on the second of the month the second account is used and so on, until each account is used once.
- Send emails on public computers, for example in libraries or cyber cafés, making them even more difficult to trace.

2. 2 persons access same email account. First saves a message as a draft email. Recipient then reads and deletes the draft.
- The e-mail has never been sent, and cannot be tracked.
More Tricks

• The language in the e-mails can be “coded”
• In preparing for the Sept. 11 attacks, Mohamed Atta and Ramzi bin al-Shibh pretended to be students as they exchanged e-mails, talking about "architecture" (the World Trade Center), "arts" (the Pentagon), "law" (the Capitol) and "politics" (the White House).
Use of Internet Cafes

- "The Internet is a very effective tool, and we use it to obtain and disseminate information quickly. Even people in Afghanistan can go across the border and send e-mail." - Yasir al-Sirri

- Publishing a Bin Laden video direct to worldwide TV is easy.
Information Terrorism

An attack, or threat of attack, on electronics, communications apparatus or computers to disrupt civilian/military infrastructure in order to achieve political aims.
Information Terrorism

An attack, or threat of attack on, electronics, communications apparatus or computers to disrupt civilian/military infrastructure in order to achieve political aims.

**Electronic Warfare**
- Disruption of communications

**Hacker Warfare**
- Damage to vital hardware
- Damage to control systems

**Psychological Warfare**
- Damage to information systems

**Netwar**
- Propaganda/Hactivism
- Communications, command & control

**Type of Attack**
- Jamming
- HERF guns
- EMP bombs
- Chipping
- Backdoors
- Trojan Horse
- Worms
- Viruses
- Direct broadcast
- Internet
- Encoded messages
From Forum to Jihad
The Internet as a Tool of Psychological Warfare

How terrorists manipulate public opinion in Western democracies
The Internet as a Tool for Psychological Warfare

- **Psychological and propaganda warfare**
  - To disrupt or damage what a target population knows—or thinks it knows—about itself and the world around it.

- **Democracies** are particularly vulnerable to information warfare via the media.

- **The Internet circumvents censorship**
  - True censorship is virtually impossible.

- **The Internet provides free speech without accountability**
  - It can be impossible to trace the origin of content.
The Internet as a Tool for Psychological Warfare

- **The Internet supports targeted Psychological warfare** – Messages can be tailored to suit the audience.

- **Information on the Internet is “weightless.”** The website of a terrorist group can appear just as valid as that of a serious news agency.
Al-Muhajiroun
The Voice, The Eyes & The Ears of the Muslims

Focus on Jihad

Latest News & Information

US Helicopter Destroyed in Attack on Khost
Mujahideen fired missiles on Khost Airport and killed three Afghan mercenaries supporting the US and destroying...

Zaeef Asks for Support in our Prayers
"...I appeal to all of my friends and relatives to pray for the steadfastness of all prisoners in American jail," Zaeef wrote in his...

Pakistan on the Brink of Shariah

New on Al-Muhajiroun

Play Part 3 | Download Part 3

Main Event

Rally for Islam
This year sees the return of the annual "Rally for Islam 6," Challenging disbelief. Set to take place in Trafalgar Square, London WC2.

Search Site

Is the Detention in Camp X-Ray allowed in their own Law?
Following the abhorrent military leg of the "campaign against terror" the Crusades have demonstrated...

Mailing List

E-mail
E-mail
Counter-measures

- We can use active and passive attacks to disrupt terrorists' electronic networks.
- Active attacks include using computer viruses to infect enemy computers. Like a Phone Tap.
- Passive attacks monitor e-mails and transferred data, and watch traffic patterns.
Implications

• Every national police, security service has a Web surveillance monitoring team
• Good Jobs for Hackers in counter-terrorism!
• Passive monitoring and data mining tools are much in demand.

End of Diversion!
1. Web Mining and Intelligence
   - RealTime News Monitoring

   “Data Stream Mining”
Live Web Site Monitoring

• **Real Time News Monitoring & Aggregation**
  – EMM, Google News, Newstini, Daylife, Newsisfree etc.

• **Blog Monitoring**
  – Opinion mining, technorati

• **Active Site Crawlers**
  – Download full site, check automatically for changes

• **Page Watch**
  – Tell me when this web page changes
  – Copernic, OWL check,
News Monitoring - The Problem

• Check all my reliable news sources for articles continuously in my languages
• Give me a feed of any articles for my subject
• Alert me when something happens in my area
• Alert me when if a massive news story happens.
• Let me search for anything in particular over time

Information Retrieval Problem
Search engine Problem

Fast Topic Detection
Breaking news detection
Moreover

Silobreaker

Google
Newsisfree
Newstln

Daylife

EMM

BIG Commercial Aggregators

LexisNexis
Factiva

Agencies: AFP, AP, Reuters etc.

Various RSS readers
Mobile Solutions
Live News Monitoring and Aggregation

Attention span is 2-3 hours

Alerts must be FAST < 30 mins
News Fluxes have some special characteristics

Weekly pattern
Weekly Hourly Averages (٢٠٠٢/٨٠/٠٣ - ٢٠٠٢/٩٠/٠٣)

hourly updates
frequency
scoop

Different Time Zones for En, Fr, Es – not Se

Multiple news sites check
- Update frequencies
- Language dependencies
News Site Monitoring

- News sites update rapidly
- Multilingual Worldwide coverage
- Duplication/Syndication from a news agency
- Multiple similar stories measure “importance” of the story
- News search engines:
  - Indexing news must be in real time
  - Page ranking though is easy!
  - Time ordered results latest first
  - Source ranking
Update Frequencies

- 24 hour news eg. CNN, BBC every 10 mins
- Major Newspaper- every 30 mins
- Regional news sites 4 times per day
- Magazines, weeklies – once per day

Need a Directory of sites
- Source, location, language, ranking, type, update frequency
Headline check for each site – detect new articles.

- Headline Page HTML is unstructured
- RSS is the simplest structured format for news snippets.
- Must convert badly formed HTML to RSS avoiding adverts etc.
- HTML -> XHTML
- XHTML → RSS via XSLT
- Process is called “Scraping”.
- Cache of RSS – append new items, trigger article retrieval
- XSLT needs continuous maintenance for each site. If the site changes then the scraping breaks!
RSS-Really Simple Syndication

• XML format for summarising web site content.
• Ideal for News Sites – but can be used for data updates
• My Netscape invented RSS 0.9
• Web Blogs (Userland) went 0.91, 0.93 → 2.0
• BUT - All are based on simple XML constructs

<channel> BBC news</channel>
   <item>
     <title>A new Headline</title>
     <link>http://news.bbc.co.uk/news?id=1234</link>
     <description>A news story published on the BBC web site</description>
     <pubDate>12-03-04:10-05</pubDate>
   </item>
   ......
RSS Feeds

• The Good News
  – Many sites are now providing RSS feeds for their main news sections.
  – No XSLT maintenance problem.
  – Easy to build a news aggregations system
    • Netvibes, iGoogle, Newsisfree, Bloglines etc.

• The Bad News
  – Often the content is not that of the web page
  – Content delayed by upto 24 hours!
Operational Issues

• 24/7 operations essential
• Speed & reliability to have peaceful weekends
• attention time for news is measured in hours
• Successful systems must be the fastest and match users interests
• We test our response against google news
  – 2004 we were up to 1 hour faster
  – 2007 parity
EMM 3.0 (late 2006)

- Scraper
- Grabber
- Ingestor
- BNS
- Indexer
- Lang Tech
- Web Server
  - UI
  - Mail
  - SMS
  - WAP

- RSS
  - All items from source
  - Latest items from source
  - RSS’ (+content)

- Alert
  - RSS +content
  - RSS +content +tags
  - RSS +content +more tags

- Fanout

- In-line processes

- Need access to internet

Legacy (item) interface

EMM 3.0 (late 2006)
Features

• Some hard lessons learned
• Clear data flow model based on feed RSS view of data
• Clear protocol between dispatcher/scheduler
  – Optimized for local queues
• Clear distribution of process logic (scraper/grabber/alert etc.)
• Better performance
  – More data in post, better distributed processing model
Software detects all new articles. Then it follows all links to full article.
NATO ASI MMDSS 10-21 Sep 2007

Software extracts article text

Detects any Of 10000 Keywords In 0.1 secs

Sorts articles into 600 categories (Alerts)

US presents new Darfur UN draft

The US has put out a new version of its UN draft resolution on the Darfur crisis, containing a modified threat of sanctions against Sudan.

It says the UN Security Council will consider such measures if violence continues in the region.

The earlier draft contained a direct threat of sanctions if Sudan failed to disarm pro-government militias.

Earlier, fresh attempts to find a solution to the Darfur crisis made no progress at talks in Nigeria.

The US ambassador to the UN, John Danforth,
On the fly Topic (ALERT) Detection

- Objective: Filter real time multilingual news
- Topic Detection and Tracking (DARPA 1997-2002)
- TDT = Breaking news detection
- Pre-defined Alerts (Topics) – keyword patterns
  - Finite State pattern matching … speed
  - Weighted sums
  - Boolean combinations

more like fishing for news
EMM Alert Definitions

Tunable subject Alert definitions detect articles & Send SMS/email alerts

Boolean Combinations

Proximity constraints
Breaking News Detection – first generation

- Headlines appear with capitalised Nouns.
- Identify new Topics appearing above the background

Signal level depends on \( \frac{dn}{dt} \)

Time to live
Real Time Breaking News Detection  
First Version

\[ S = \frac{n_t}{\bar{n}_t} \beta \frac{n_F}{N_F} \]

- \( n_t \) = number of occurrences of topic in last hour
- \( \bar{n}_t \) = rolling time average of occurrences of monitored topic per hour
- \( \beta \) = multiplying coefficient – tuned per language en=6
- \( n_F \) = number of feeds (sources) from which the topics are derived
- \( N_F \) = feed scaling coefficient (current value 3)

Database monitors 0.5 Million capitalised Keywords over 3 week period
Rolling time averages of each topic are calculated every 10 minutes.
Tuned Stopwords, all languages, clustering of overlap topics

Tests show reaction times to major events (capture Sadam) of < 30 minutes
Thresholds tuned per language
Breaking News Detection in single Alerts/Topics

• Keep hourly expectation rates on flux of articles
• Statistical measure of fluxes of predefined news
Topic Alert Detection

Calculation:

Normalized fraction of articles for a Topic T each day.

\[ A_i = \frac{T_i}{N_i} \]

Average of the normalized value for previous N-1 days ("expected value")

\[ \bar{A} = \frac{\sum_{j=0}^{N-2} A_j}{N - 1} \]

Standard deviation of the sample. Assume a normal distribution.

\[ \sigma_s = \sqrt{\frac{\sum_{j=0}^{N-2} (A_j - \bar{A})^2}{N - 1}} \]

Current cumulative level high if > 0.99

\[ \phi = \Phi_{\sqrt{\sigma_s^2} + \bar{A} (N-1)} \]
TDT Version 2: Top Stories “Real Time Clustering”

- Improve clustering of news reports
- Must work real time and react to sudden events
- Improve performance in low coverage languages
- Improved minute by minute story tracking
- SMS + email alerts including “News Updates”
- Algorithm optimised over 4 months of manual checks.
RTC Algorithm in 18 languages

4 hour rolling corpus en,fr,de  extend period for other languages until >200 docs
repeat every 10 mins.

120 clusters

400*400
distance matrix

cosine distance

120 clusters

Infinite input response filter

~400 vectors

~ 6000 words
400 docs

Remove 100
Top V

First
200 words

Remove E>0.7
twice in
2 docs

filter unique
out words

Entropy
calculation
Real Time Story Tracking

• Story is a linked set of clusters with overlapping IDs

• Stories can merge and die
40000 English Articles

Entropy–Word Frequency Distribution

1 - S

Word occurrence

10^0 10^1 10^2 10^3 10^4 10^5 10^6
EMM uses 10 min clustering in 18 languages

- Language-independent ➔ multilingual
- Interesting cross-lingual comparisons
RTC

10 min clusters

New BN stories
Demo
Europe Media Monitor Services

EMM
European Media Monitor
News Tracking - Timelines

EMM in active use by EC
- 4000 email alerts/day
- 50000 active web users
- 500 SMS/day to VIPs
- 10000 keywords Real-Time
- 40000 articles/day
- 30 languages
- 600 topic alerts Real-time

RNS
Rapid News Service

Monitors all World Countries
Derives statistical indicators and time trends

Live news
Newsletters
Push SMS alerts
Press Reviews

All Europe’s news
Breaking News
40,000,000 articles
Processed since May 2002

EMM Services
Europe Media Monitor Services
An example specialised news monitor

Top News
updated
10 minutes
From African Sources
Clinton, Obama clash over US diplomacy

Sudan(5) Clinton supporters characterized it as a gaffe that underscored the freshman senator’s lack of foreign-policy savvy while Obama’s team claimed his response displayed judgment and a reputation of President Bush’s diplomacy. In a memo from Obama campaign spokesmen Bill Burton, the campaign contends...

Sudan: President Declares 2007 Year of Peace

atlasica 17:45 Tue, 24 Jul 2007

Sudanese President Omar al-Bashir has declared 2007 as the Year of Peace in Darfur.

This article also triggered: Tanzania, Libya, CurrentSituations, Africa,Unification, Chad.

Sudan: Meeting On Darfur Conflict Planned for Early August, UN And AU Envoy Say

atlasica 17:45 Tue, 24 Jul 2007

Khartoum(3) Sudan(6)

Pushing ahead with efforts to find a political solution to the conflict that has engulfed the Darfur region of Sudan, envoy from the United Nations and the African Union today invited leaders of movements that have not signed the main peace accord to “pre-negotiation” talks scheduled for early August in Arusha, Tanzania.

This article also triggered: Tanzania, CurrentSituations, Libya, Africa,Unification.

KENYA: Pastoralists lose 100,000 animals to livestock disease

elchristo 17:22 Tue, 24 Jul 2007

Sudan(5)

NAIROBI, 24 July 2007 (IRIN) - Herders in Turkana district in northwestern Kenya had lost an estimated 100,000 goats and sheep in the past year to pastos...
End Product
Can be emailed to lists of users or send to Mobile phones
Daily News products
Power of Syndication

• RSS allows content syndication
• Real time news and Information extraction for all
• Syndicate to 3d party sites
• Syndicate to Web2 portals
  – Netvibes
  – iGoogle
  – Bloglines
  – Mobile Phones
**ABU SAYYAF GROUP (ASG)**

**Mother Tongue Name:** 
جناح أبو صاموس

**Aliases:** Bearer of the Sword, al-Harakat al-Islamiyah

**Base of Operation:** Philippines

**Founding Philosophy:** The Abu Sayyaf Group (ASG), or Abu Sayyaf, is a radical Islamic terrorist group active in the Southern Philippines and Malaysia. Its stated goal is the creation of an independent Islamic state encompassing parts of Southern Thailand, the island of Borneo, the Sulu Archipelago, and Mindanao, areas where Moro Muslims, a minority ethnic group in the Philippines, make up the majority of the local population. The ASG is known to target Filipino and Western Christians in the Southern Philippines, though the group's influence is thought to have expanded to the regional level recently.

The ASG was founded in 1991 by radical Moro National Liberation Front (MNLF) members who objected to the MNLF's negotiations with the Philippine government. Due to the ASG's predisposition toward violent tactics, which include high-profile bombings, armed attacks, assassinations, and beheadings, it is seen to be more radical than its mother group. Abu Sayyaf, which literally means "father of the sword" in Arabic, sees itself as the rightful inheritor of the legacy of armed Moro resistance in the region and the torchbearer in the struggle for the establishment of an Islamic state in Southeast Asia. Ironically, the group rose into prominence in the mid-1990s after two large-scale plots (one to destroy 12 U.S. commercial aircraft simultaneously in mid-flight and one to assassinate Pope John Paul II) were foiled. It is thought to have two to five hundred core members.
News Visualisation
1. Linked category networks

- Articles belong to different categories and can contribute to a story
- Automatically detect those categories most linked together and to particular stories.
- Start with one category or one story to view all the related news.
  - Iran Nuclear
UN nuclear chief calls joint plan on Iran an important step

UN nuclear chief hails deal with Iran at IAEA

U.S. experts to inspect N. Korean nuclear facilities - paper

Oil falls $1 ahead of OPEC meet

SA oil empowerment deal

Rebels blow up Mexico oil pipeline

Mexico’s state oil company reports explosions at pipeline

Opec to debate token oil output rise

12,000 evacuated in Mexican pipeline blasts

SA oil empowerment deal

Rebels blow up Mexico oil pipeline

Mexico’s state oil company reports explosions at pipeline

Opec to debate token oil output rise

12,000 evacuated in Mexican pipeline blasts
Sep 13th Ecology

- Roadside diesel pollution poses heart danger - study
- Conservation union finds 16,300 species threatened
- Child mortality at record low; Unicef predicts further drop
- UK study clears mobile phones of short-term risks
- Ebola said depleting gorilla populations
- Official: Tumour link to mobiles
- Health experts rush to tackle Congo Ebola outbreak
- Contraceptive pill may cut cancer risk: study
- Japan PM Abe quits after year of scandal, crisis
- Foot-and-mouth disease confirmed at British farm
- Test to show if UK foot and mouth strain the same
Old Visualisation Idea: State of the World

- Human Events evolve in time and space.
- Automatic news monitoring software (EMM) tracks and records events reported around the world.
- EMM Classifies 24/7 news by Topic (subject).
- Software is “impartial” given wide coverage of opinion.
- Can software provide a “measure” of the “State” for all countries at a given time?
- Can software provide early warning of crises - “breaking news”? 
- Can software then record all this for later analysis?
State of the World

Physical

• Thermodynamic systems have a state $S(T,V,P...)$. Hide molecular complexity.
• Measured by Temp, Press, Entropy…
• Weather is rather more complex – chaotic
• Weather Maps

Social

• Social systems have a “state” (war, peace, growth, famine etc.) Hide complex issues.
• Measured by news reporting - indicators
• Real life is also complex even chaotic
• News Maps
News Maps derived from Indicators

Weather
Temp, rainfall

NEWS
Conflict, Famine

select THEME Conflict
Change detection News Maps

Countries more in the news
Compared to 5 Day average.

Countries more mentioned in Conflict articles than expected
Natural Disaster Indicator 16 Dec 2004

Terrorist Attack
Thematic Indicator
Information synthesis for EC Foreign Affairs Department

Provide desk officers with concise clear overviews of current situation in any country of the world. Integrate news, Maps & structural data

News hotspots for themes
Long Term News Analysis

• News Explorer database
  – Clusters
  – Entities
  – Places
  – Relations
  – Events
  – ……
Data mining news archives - Visualisation & Maps

Many dead as quake hits Indonesia
At least 15 people have died and hundreds have been injured after a strong earthquake shook Indonesia’s ancient royal capital of Yogyakarta, Reuters reports. AFP reported quoting police and witnesses that several people were killed and hundreds injured when the quake rocked the densely populated....

Date: Sat, May 27, 2006
Articles in Cluster: 107
Countries mentioned in the Cluster: Indonesia
Alerts mentioned in the Cluster: NaturalDisasters
read cluster

Search for survivors as Indonesia quake toll tops 3,000
Search for survivors as Indonesia quake toll tops 3,000 BANTUL, Indonesia (Reuters) - May 28, 2006 04:35 GMT Rescue workers dug

Date: Sun, May 28, 2006
Articles in Cluster: 64
Countries mentioned in the Cluster: Indonesia
Alerts mentioned in the Cluster:
Long timelines – simply largest story each day.
Social Networks - Relation Extraction

- Two entities are Related or “linked” through a phrase
- Machine Learning of relations.
  - “contacts” (met, phoned, discussed with, emailed etc.)
  - “support” (backed, applauded, welcomed, concurred etc.)
  - “criticise” (slamed, rejected, criticised, accused, etc.)
  - “family” (wife, son, daughter, lover, mistress etc)
- Mine 3 years of news reports
- Log dates and related topics
Learning “Contact Relation”

- “contact” means (meeting, phoning, emailing, etc.)
- First, using an initial set of patterns like “X met Y” we detect phrases like Tony Blair met George Bush
- Next, we search in the same news cluster other phrases which paraphrase the above relation, e.g. the talk between Tony Blair and George Bush
- Next, we add “the talk between X and Y” to the set of patterns and repeat the learning
- In such a way we found 105 contact relation patterns
- Automate to extract A “contacts” B from selected news reports
Contact Network for Iraq Insurgency
Iran Nuclear Standoff Structure
Social Network During Lebanon Conflict
EMM: Mining Social Networks

Key: Contact, family, support, criticise, generic
3. OSINT Web Mining Tools
Open Source Data Collection Tools

• Analysts need tools to cope with information overload

• Objectives
  – Automate Information Retrieval
  – Automate information extraction
  – Case management
  – Data Archival
Goals of OSInt Tools

- Information Retrieval from WEB or other.
- Extract “Facts” from unstructured Text

Structured

<table>
<thead>
<tr>
<th>Event Table</th>
<th>ID</th>
<th>Title</th>
<th>Start Date</th>
<th>End Date</th>
<th>Event Type</th>
<th>Killed</th>
<th>Location</th>
<th>Country</th>
<th>Target</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Attack</td>
<td>8/29/2010</td>
<td></td>
<td>Terrorist</td>
<td>0</td>
<td>Egypt</td>
<td>Egypt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Event</td>
<td>8/29/2010</td>
<td></td>
<td>Terrorist</td>
<td>0</td>
<td>Egypt</td>
<td>Egypt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Attempted</td>
<td>9/13/2010</td>
<td></td>
<td>Terrorist</td>
<td>0</td>
<td>Egypt</td>
<td>Egypt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>War</td>
<td>8/19/2010</td>
<td></td>
<td>Terrorist</td>
<td>0</td>
<td>Afghanistan</td>
<td>Afghanistan</td>
<td>Western</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>Event</td>
<td>8/29/2010</td>
<td></td>
<td>Terrorist</td>
<td>0</td>
<td>Afghanistan</td>
<td>Afghanistan</td>
<td>Western</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>Assassination</td>
<td>9/19/2010</td>
<td></td>
<td>Terrorist</td>
<td>0</td>
<td>Cairo</td>
<td>Egypt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>Attack</td>
<td>9/19/2010</td>
<td></td>
<td>Terrorist</td>
<td>0</td>
<td>Egypt</td>
<td>Egypt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>Event</td>
<td>8/29/2010</td>
<td></td>
<td>Terrorist</td>
<td>0</td>
<td>Cairo</td>
<td>Egypt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>Event</td>
<td>8/29/2010</td>
<td></td>
<td>Terrorist</td>
<td>0</td>
<td>Cairo</td>
<td>Egypt</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An explosion occurred in the
Le Bateau Submarine de Paris
An explosion occurred in
an unorganized fight in Paris.
1. Iran Focus - Iran accused of hiding secret nuclear weapons site ... 
Suspicions have been growing that Iran has a secret military nuclear research site, by Iran's Revolutionaries. Check www.iranfocus.com, module shows article.php?storyid=7656 - 13k - 

2. Iran Nuclear - Tehran concentrating on building 5000 centrifuge ... 
Engineer Ali Rezaei who is involved in nuclear work in the Defense Ministry under the nuclear organization's nuclear research center. - 5000 centrifuges. - 17k - 

3. Iran Focus - IAEA report backs Iran's claims on group's revelations ... 
They were identified as Mohsen Fakhrizadeh, Fereydoon Abbasi, Mansour Asgari, and Mohsen Kheirkhah. - 17k - 

4. Godlike Productions - UFOs, Conspiracy and Lunatic Fringe - Discuss it all at Godlike Productions - 'We put the FUN back into ufology'. - 17 - 

5. calgal's qalma ... 
Calgal's qalma ... 

6. Walter Dimmock - A World Viewpoint Commentary On Today's Culture ... 
The regime's most senior nuclear experts such as Mohsen Fakhrizadeh work under the control of Iran's Revolutionary Guards. - 13k - 

7. el peligro islamista - June 2006 
Zarimán 27 - el proyecto nuclear secreto militar iran... ... - by Iran's Revolutionary Guards under the control of Iran's Revolutionary Guards. - 9k - 

8. Iran's 'Nuclear University' conceals research [Archive] - Rapture ... 
Other key figures include Mohsen Fakhrizadeh, 45, a nuclear engineer who worked on the International Atomic Energy Agency (IAEA) project. - 9k - 

9. NCI - Smuggling of Beryllium - 9.1.05 
Conan Mohsen Fakhrizadeh, one of the most important nuclear experts of the clerical regime. - 9.1.05 - 

10. Soon - Gold & Silver Forum 
Nuclear experts working for the International Atomic Energy Agency in Vienna are... - 82k - 

Jaharzadeh Press Conference Statement - 87k -
### Entity Analysis Wizard

#### Entity Analysis Options

Press the 'Finish' button to start the entity analysis.

- [x] Person Recognition
- [x] Date Recognition
- [x] Telephone Number Recognition
- [x] Email Recognition
- [x] Token Recognition
- [x] IP Number Recognition
- [x] Organisation Recognition
- [x] Geographical Recognition
- [x] URL Recognition
- [x] Number Recognition
- [x] IP Names Recognition

#### Minimum Entity Threshold

<table>
<thead>
<tr>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

#### Language

- Guess Language
<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Entity Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORG</td>
<td>GMT Mullah Security Forces</td>
<td>12</td>
</tr>
<tr>
<td>ORG</td>
<td>IRNA news agency</td>
<td>12</td>
</tr>
<tr>
<td>ORG</td>
<td>This university</td>
<td>12</td>
</tr>
<tr>
<td>ORG</td>
<td>Wall Street</td>
<td>12</td>
</tr>
<tr>
<td>ORG</td>
<td>Jaber Ibn Hayyan Laboratory</td>
<td>12</td>
</tr>
<tr>
<td>ORG</td>
<td>Atomic Energy Organization</td>
<td>7</td>
</tr>
<tr>
<td>GEO</td>
<td>IR</td>
<td>12</td>
</tr>
<tr>
<td>ORG</td>
<td>Iran Scientific and Industrial Organization</td>
<td>12</td>
</tr>
<tr>
<td>ORG</td>
<td>laboratory of the Atomic Energy Agency</td>
<td>12</td>
</tr>
<tr>
<td>ORG</td>
<td>laboratory of the Atomic Energy Organization</td>
<td>12</td>
</tr>
<tr>
<td>ORG</td>
<td>Malek-Ashtar University</td>
<td>12</td>
</tr>
<tr>
<td>ORG</td>
<td>Scientific and Industrial Organization</td>
<td>12</td>
</tr>
</tbody>
</table>

**Related Documents**

1. [Document](http://www.irnwatch.org/privateviews/NCRI/perspex-ncri-neutrino...)
   - Entity Count: 4
   - Document Count: 4

**Filters**

- Geo Location
- Date
- Number
- Organisation
- People
- URL
- Email
- IP Name
- IP Number
- Telephone
- Entity Name
- None

**Actions**

- View Web Page
- View RSS Item
- Delete Entity
- Add Entity to ignore list
PROJECT TO BUILD NEUTRON INITIATOR, THE TRIGGER TO FISSION CHAIN REACTION FOR NUCLEAR BOMB

PRESS CONFERENCE BY MOHAMMAD MOHADDESSIN FOREIGN AFFAIRS COMMITTEE CHAIRMAN NATIONAL COUNCIL OF RESISTANCE OF IRAN February 3, 2005

As most of you know, a nuclear bomb consists of three main parts: nuclear fuel, launching systems and a neutron initiator that is needed to trigger a fission chain reaction. In the past few years, the National Council of Resistance of Iran has revealed the Iranian regime’s efforts to acquire nuclear fuel and facilities used for uranium enrichment. Last November, here in Paris, I revealed information on two new nuclear sites that are involved in laser enrichment. In December, my colleagues in London exposed information about the top-secret Ghadr missile that is capable of carrying nuclear warheads. Today, however, I would like to expose the secret project pursued by the regime to build the neutron initiator to trigger a chain reaction for a nuclear bomb. This renders false the mullahs’ claims that they are only seeking a peaceful use of nuclear technology. The Iranian regime is producing Polonium-210, Beryllium and Neutron Generators. They have concealed these activities from the IAEA. I will now explain some of the details regarding these activities. As we speak, Tehran has already succeeded in using Beryllium in conjunction with Polonium-210 for large-scale laboratory testing purposes, and is getting very close to the point of industrial production. Polonium-210 emits Alfa source, which in conjunction with Beryllium metal, produces neutron source that serves as the trigger for fission chain reaction for a nuclear bomb explosion. The Ministry of Defense in charge of building the neutron initiator. Brigadier General Dr. Seyyed Ali Hosseini Tash, deputy Defense Minister, is the official in charge producing weapons of mass destruction in the Ministry of Defense. He is, among other things, responsible for producing the neutron initiator for the nuclear bomb. The production of Beryllium and polonium 210 is being carried out under the supervision of Hosseini Tash. He has all the capabilities of the Iranian Atomic Energy Organization at his disposal. Producing Polonium-210 by irradiating Bismuth metal, Iranians have succeeded in transmuting it into Polonium-210. Tehran has lied to the IAEA that it has not produced Polonium-210 in the past 12 years, since 1993. It has also failed to...
Case Study

• Wednesday 20\textsuperscript{th} September- UK Home Secretary Reid heckled at visit to East London Muslim Community.
Reid heckled during Muslim speech

Home Secretary John Reid has been heckled during a speech about targeting potential Muslim extremists.

He was interrupted by activist Abu Izzadeen, who said he was “furious” about “state terrorism by British police”.

In his speech, Mr Reid asked Muslim parents to keep a close eye on their children and act if they suspected they were being radicalised by extremists.

It was his first speech to a Muslim audience since becoming home secretary.

The protester, also known as Omar Brooks, denies being a member of the banned al-Ghuraba group.

He accused the minister of being an "enemy" of Islam before he was led from the building by police and stewards.
Abu Izzaden
Information about this person was last updated on Saturday, November 19, 2005.

Names
Abu Izzadeen (en)

Analysis over time
Timeline
Timeline [en] for 09/2006

Quotes from
There are no recent quotes in the chosen language.

Quotes about
There are no recent quotes in the chosen language.

Important notice about quotes:
All quotes both by and from, are derived by an automatic algorithm with no human intervention. This means that sometimes mistakes WILL occur. Users are therefore strongly advised to check the validity of quotes from the original news report. Neither the JRC or the European Commission accepts any responsibility for the content of these pages.
OSInt links the Entity to the London Bombings

London bomber
Once a friendly Christian, he now backs the bombers

BY NICOLAWOOLCOCK AND SEAN O'NEILL

Two faces, two converts - two Muslim extremists in Britain

A BRITISH Muslim convert has emerged as successor to Omar Bakri Mohammad as the leader of a radical group that wants Britain ruled by Islamic law.

The Times has obtained transcripts of Omar Brooks, now known as Abu Izzadeen, preaching holy war and discussing killing Tony Blair in a recent sermon in London. Abu Izzadeen had previously described the...
Future Challenges

• News Explorer is quantized in 24 hour chunks

• Real-Time information extraction
  – Real time clustering, story tracking
  – Entity recognition
  – Geocoding
  – Event Extraction

• Live Situation Monitoring
2 NATO soldiers, 40 insurgents killed in Afghanistan (2009)

Location: AF: Kandahar
Impact: 2 Deaths
Description:
Soldiers killed in two separate attacks in southern part of country; nationalities not specified

Links: Go to Load Article - Go to EMM Cluster

Last update: 07/09/2007 12:01 Running for 1:27 (Hours Minutes)
Conclusions

• Security Application are a strong growth area for Web Mining
• OSINT tools are much in demand
• News Monitoring/Aggregators – many new entrants
• NEWS/Information syndication is important for the mobile web.
• Link analysis in news just beginning
• Mining news archives like EMM has high potential
Thank You