12th ICCRTS

"Adapting C2 to the 21st Century"

Developing a Horizon Scanning System for Early Warning

Topics: C2 technologies and system, sensemaking, collaboration

Foo Kok Jin, Wong Rong Fah Ni De En, Leong Ming Wei
Leong Hoe Wai

Defence Science and Technology Agency National Security Coordination Center DSO National Labs

POC: Foo Kok Jin
Defence Science Organisation
1 Depot Road, #03-01J Singapore 109679
Phone: (+65) 63734559, Fax: (+65) 6275-2910
E-mail: fkokjin@dsta.gov.sg

Abstract

In Singapore we have embarked on a Horizon Scanning Initiative to provide the concept, theory and technology enabler for making sense of the vast amount of available data for early warning. Our system is premised on three principles: (a) Technologies are to augment Human and team in the sensemaking process and the needs to b) move beyond data sharing to perspective sharing and (c) to move beyond mere search towards data analytic and data structuring services. We have completed the development of a Service Oriented Based Horizon Scanning Architecture (SOSA). SOSA enables collaboration across agencies and allows data and tool in different agencies to be treated as web services that are discoverable, sharable and can be orchestrated for analysts' consumption. The perspective sharing services allow multiple analysts from different agencies to meta-tag and add comments to incoming data set which could then be visualised and can serve to amplify outliers and allow users to avoid getting blind-sided through premature convergence. The data structuring services enable building of system map with associated consistency matrix and to perform automated morphological analysis. This paper will explain the imperatives of the Horizon Scanning Initiatives, SOSA and the perspective sharing and data structuring services.

Introduction

The complexity of issues and tempo of activities handled by the Singapore Public Service has increased substantially over the past years. Due mainly to a) an increase in overall scales of activities, b) increase in inter-connectivity, and c) interplay of different components exhibiting emergent behaviour. In this regard, two recent examples stand out clearly. One is the declining birth rate of Singapore. To address this, a Government review panel government had to look at issues handled by multiple departments. These included employer demands, workplace benefits and employee wages, family cohesion and values, manpower for national defence, and rising educational status of women. Another example was addressing the threat of Avian Flu spreading to Singapore, and planning for a possible human flu pandemic. This will involve co-ordinating the actions of our foreign affairs, human health, veterinary health, housing, and transport departments.

Ops Impetus for Early Warning

Given the complex, international and cross domain nature of the issues that we faced, it is no longer feasible to focus on past trends and assuming the future will be similar, or by adopting an agency-specific, or worse still, purely reactionary approach to emerging issues. This is especially so given Singapore's inherent vulnerability both economically and socially. This small country-state of around 700 km^2, is home to several oil refineries, and provides highly developed infrastructure to maintain its status as a major air and sea transport hub. Its population comprises 4 main races and over 5 religions. While it stands as a symbol of stability and competent governance, it is also a prized target for terrorists, which if skilled, can easily exploit the vulnerabilities. With this in mind, we have embarked on a Risk Assessment and Horizon Scanning (RAHS) Initiative, which will enable the Public Service to make sense of the vast increase in amounts of data and know-how, to provide early warning of arising strategic threats, such as pandemics or terrorist attacks, so that anticipatory action can be taken to remove or mitigate the impact.

The initiative will employ technology to:

a) Augment human efforts in sensemaking processes, and to also adequately manage the volume of data without missing out on weak signals.

b) Enable collaboration and information exchange across various Public Service departments, to meet the challenge of cross-sector issues, as well as the increasing speed at which they arise.

Through this, the initiative will also ensure optimal allocation of limited Public Service resources, by allowing departments to focus efforts on identified low probability but high impact issues.

Survey of orginisations

We have carried out studies of various organisations that have carried out risk assessment and horizon scanning functions, and found that scenario planning and massive data analytics are the two main methods used.

This traditional Scenario Planning method has been employed by Shell International Petroleum Company before the 1973 oil shock, to allow the company to anticipate the rise and subsequent fall of oil prices. Also, in the mid-1980s, Shell created scenarios that focused on the future of the Soviet Union because that country was a major competitor in the European gas market.[1]

[1] Details in Chp 13 "Scenarios", Futures Research Methodology CD-ROM v2.0 on, by the Millennium Project.

Siemens AG has utilized a form of scenario planning as part of its technological forecasting technique, named "Pictures of the Future". This technique involves selecting a suitable time frame in the future, and generating comprehensive scenarios taking into account various aspects such as socio-political environments and new customer needs.[2]

The Millennium Project, a global participatory futures research think tank for global issues, has also employed scenario planning together with the Delphi participatory method, to produce a large range of scenario sets in various domains such as Demographics and Human Resources, Environmental Change and Biodiversity, Technological Capacity, Governance and Conflict, and International Economics and Wealth.[3]

In the public sector, the UK Horizon Scanning Centre, under the Department of Trade and Industry (DTI), helps identify future issues and trends relevant to the entire public policy spectrum, often employing scenarios and expert consultation in the process. Its aim in such work is to feed into cross-government priority setting and strategy formation. Its work has been used by the Health and Safety Executive to inform scenarios on the future of workplace health and safety.[4]

The Global Public Health Intelligence Network (GPHIN) was developed for WHO through collaboration with Health Canada in 1996, and functions as a secure Internet-based early warning system that gathers information about potential public health threats on a 24/7 basis. It functions as a web crawler, conducting data mining from various sources, with a focus on infectious diseases and human safety issues. Acting on keywords assigned by Health Canada's Laboratory Centre for Disease control (LCDC) and keywords associated with public, environmental and animal health, it monitors over 10,000 online sources, such as newspapers and biomedical sources.[5]

The Australia and New Zealand Horizon Scanning Network (ANZHSN) is an initiative under the Australian Government Department of Health and Ageing (DoHA), that utilizes internet scanning and stakeholder consultation to provide advance notice of significant new and emerging technologies to health

---

[2] Details at Siemens' website: http://www.siemens.com/index.jsp?sdc_p=ft4mls3u20o1156534n1156534i1168864pFEcz2&sdc_sid=21842775797&

[3] Details at the AC/UNU Millennium Project website: http://www.acunu.org/millennium/environscen.html

[4] More info at the DTI website at: http://www.gnn.gov.uk/environment/fullDetail.asp?ReleaseID=251912&NewsAreaID=2&NavigatedFromDepartment=False

[5] More background at the Carleton University website at: http://www.carleton.ca/jmc/cnews/12031999/f4.htm

departments in Australia and New Zealand, and to exchange information and evaluate the potential impact of emerging technologies on their respective health systems.[6]

Principles of the RAHS System

All methodologies and technologies have their pro and cons. The RAHS System in Singapore thus aims first to define the key principles and provide an inclusive architecture that not only includes a comprehensive suite of methods and technologies based on existing work and concepts still being developed, but also be flexible enough to cater for future methods and technologies. The three key principles of our systems are:

a) Technologies are to augment Human and team in the sensemaking process as we believe that Human is the ultimate sense maker.

b) We need to move beyond data sharing to perspective sharing

c) We need to move beyond mere search towards data analytic and data structuring services

Based on the principles our systems aims to provide four enablers to the users:

a) Enable collaborations across agencies: RAHS will link up various agencies and provides an architecture with applications that enable people from different agencies to collaborate. In developing the RAHS system, multiple agencies have been involved; from the requirement-gathering phase, to constant user feedback in line with the agile development methodology, to the deployment of the operational system as eventual users of the System.

b) Enable the reduction of time spending on search John Poindexter, chief architect of the DARPA Terrorism Information Awareness Programme, advocates the use of computer tools to augment human performance[7]. One of the ways in which this can be achieved is through the use of analytical tools, that reduce the search time required by analysts to increase the amount of analysis time spent on relevant information. To this end, the RAHS System implements a range of analytical tools to reduce the amount of reading time required by analysts

[6] Details at the DoHA website at: http://www.horizonscanning.gov.au/

[7] Report to Congress regarding the Terrorism Information Awareness Programme

and implements a software to automate the running of pre-determined workflow.

c) Enable the creation and use of Models for sense making and monitoring Based on the Naturalistic Decision Making framework developed by Gary Klein[8], experts, when dealing with complex situations and faced with high levels of uncertainty, rely on mental models and pattern matching against these models to allow them to discern subtle clues and do just the right thing in situations where novices fail. This hypothesis is further substantiated and popularised by subsequent publications from Klein[9] as well as by Malcolm Gladwell[10]. The RAHS System explores the use of various models as different means of data structuring to make explicit mental models, and improve pattern matching and weak signal detection.

d) Enable the sharing of perspectives Knowledge management guru, Dave Snowden argues the need for and value of perspectives sharing when operating in the complex domain, in order to gain new perspective on the situation rather than to rely on entrained patterns from past experience[11]. He describes the complex domain as one where "emergent patterns can be perceived but not predicted", thus greatly diminishing the value of "expert opinion based on historically stable patterns of meaning".

System Overview

We have successfully completed the development of a Service Oriented Based Horizon Scanning Architecture (SOSA). SOSA enables collaboration among human analysts in the different agencies. It also enables data and tool in different agencies to be consumed as services so that they are discoverable and can be orchestrated. The architecture sits on a physical network connecting the various agencies involved in our pilot phase of the project. On top of the network, key components of the SOSA are a generic, service-based flexible architecture called SEFAR (Service-Oriented Flexible Architecture) and a collaborative environment. The RAHS system also provides a wide range of services such as data services, application services and visualisation services.

SEFAR allows service discovery and sharing of data, algorithms and visualization in a multi-agencies environment, and provides the orchestration

[8] *Sources of Power: How People Make Decisions*
[9] *The Power of Intuition: How to Use Your Gut Feelings to Make Better Decisions at Work*
[10] *Blink: The Power of Thinking Without Thinking*
[11] *The new dynamics of strategy: Sense-making in a complex and complicated world*

layer which enables workflow configuration on the fly. A key benefit of SEFAR is to allow system resources (data, tools and visualization) to be better managed and utilised. In addition, users are given greater flexibility to fine tune or create new workflow to cater for changing business requirement.

Overview of SEFAR Architecture

SEFAR is designed to be generic, extensible and conforms to industry-support standards. Thus, architecture consideration includes: Interoperability, Reusing existing applications, Scalability, Security and use of industry supported standards

The basic elements in the SEFAR architecture are:

a) Workflow MMI Display : The workflow MMI allows user to compose (i.e. aggregate) Web services. The output of this module will be in a XML format, termed as the workflow XML. The workflow XML describes the relationship between data, tools and visualisation services. This information is then passed to the orchestration layer for further processing. The workflow MMI display executes at the client-side.

b) Orchestration Layer. This module is divided into two separate modules. They are XML Mapper and XML Actor. The XML Mapper is responsible for parsing the workflow XML into events. This chain of events is then passed to the XML Actor for service initialisation and invocation. The XML Actor is responsible for service synchronisation and mediation. The orchestration module itself is a web service residing on server side.

c) Service Container. A service container is a network node hosting Web services. The service container resides at the server-side. Service Containers contain web services (data, tools and visualisation) and sit in a multi-agencies environment.

d) Visualisation Display. Visualisation display allows computed results to be displayed at the client-side within the Eclipse RCP (Rich Client Platform) for viewing and analysis work.

e) Utility Servers. Utility servers are made up of UDDI Registry Server and Message-Queue Broker Server. The UDDI Registry Server keeps track of all registered Web services. The Message- Queue Broker Server provides message exchange facilities between web services. The security module ensures SSL transaction between actor in orchestration

layer and the web services. Quartz scheduler is used for the scheduling of workflow.

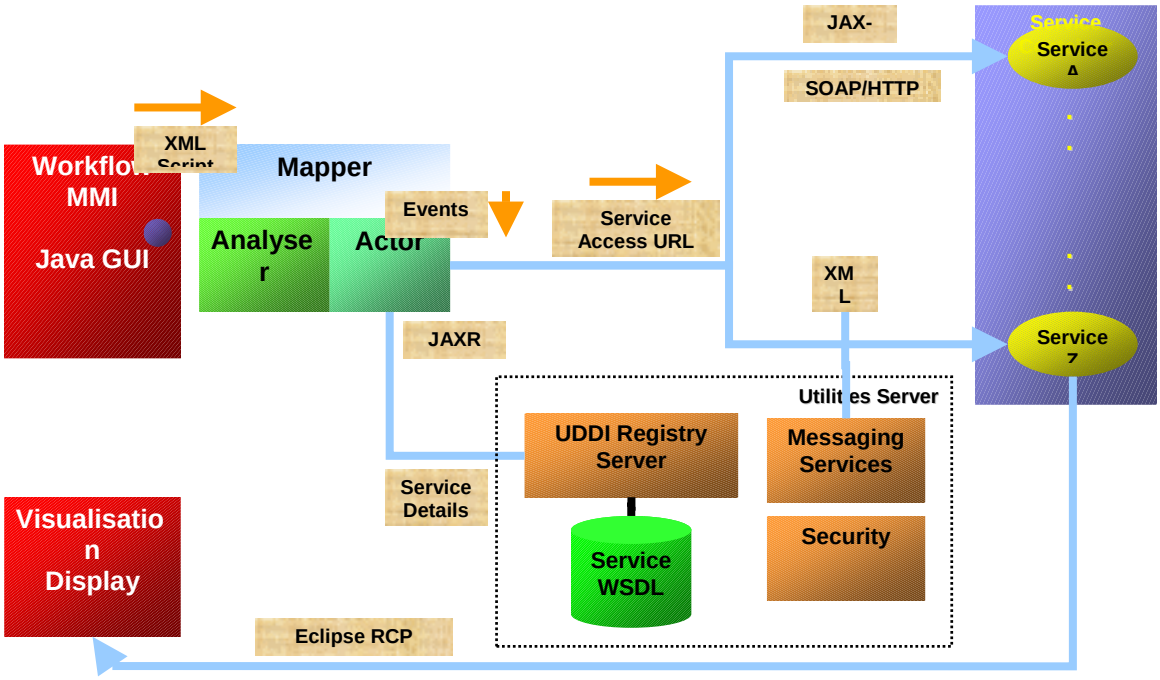

Figure 1: SEFAR Architecture

A simple scenario of a typical user's workflow, refer to Figure 1:

a) A user orchestrates end user workflow using the workflow MMI display.
b) The workflow MMI converts the graphical workflow into an XML script.
c) The XML Mapper takes in the XML script understands it and breaks it down into events.
d) Events are then passed to the XML Actor for further processing.
e) The XML Actor takes in the events and performs service discovery via the UDDI Registry Server to determine services end-points (service access URL).
f) The XML Actor initiates service synchronisation. The Message Queue (JMS) is responsible for XML messages exchange between services.
g) All the web services are residing at the server-end. Servers that host web services sits in a multi agencies environment. There are no data and web services at the client-end.

Results computed by services are pushed back to a client via a message queue subscription by the visualisation display on the client-end. The plug-in of the visualisation display sits within Eclipse RCP framework on the client-end. Figure 2 illustrate an example of how an end-user uses SEFAR workflow MMI. User loads a workflow which he has orchestrated and scheduled previously. The workflow automates and monitors the end user work process of performing

searches on the various data sources, moving the articles into project folders and performs entity, timeline and network analysis on the consolidated articles. SEFAR orchestration layer and utilities servers mediate the messages pass between the different web services in the workflow. User can choose to perform further analysis on the system should he discover interesting information from the analytical visualisations.

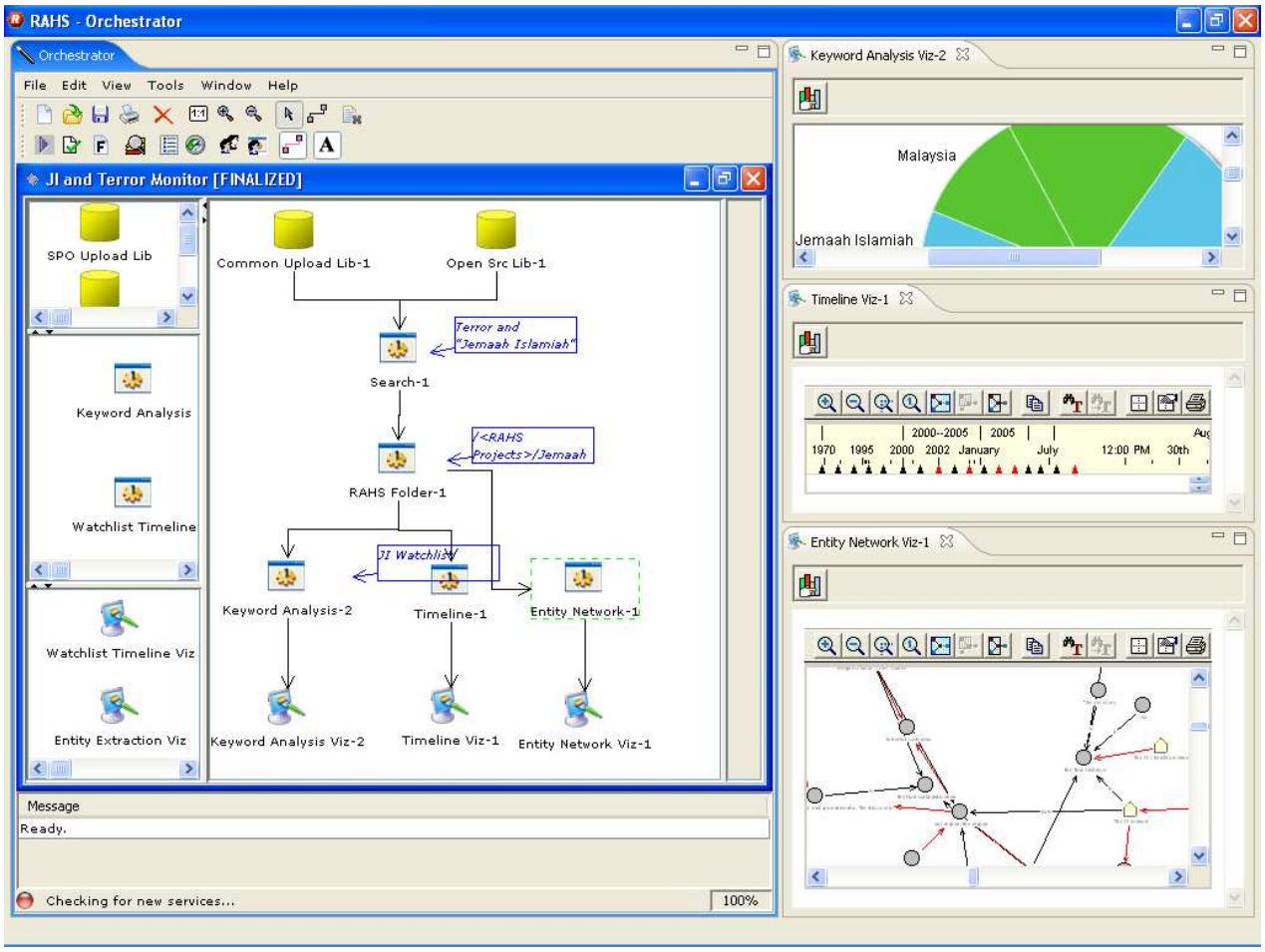

Figure 2: A typical service scenario

Advance Data Analytic

RAHS system empowers analysts with a suit of tools to help them process large amount of information. The system is designed to empower the analysts without limiting their intuition; not everyone processes information in the same way. RAHS gives the user the flexibility to apply these analytical tools

in any order in support of the analytic process. Description on the analytical tools will be covered in this section.

Search and Advance Search Tool

RAHS allows Analysts to search for articles within its repository. Advanced Search provides powerful features for fine-tuning text searches. Analysts can select different search modes, select specific libraries and filter search through fielded queries. There are three primary search modes: Concept, Pattern, and Boolean. Concept search enables related words or concepts that may be relevant to search query. It utilises a knowledge base containing word meanings, syntax, word variations, and relationships between words. These defined relationships between words link them together in a "semantic network". In pattern search, query terms are expanded to include terms with similar spellings. Boolean search produces exact matches and is a fast way to look up articles without going through the majority of the semantic network. If an analyst enter multiple terms, he must use the appropriate operators (e.g. AND, OR, NOT, etc.). The search engine also enables the three different search modes to be mixed within a single search query.

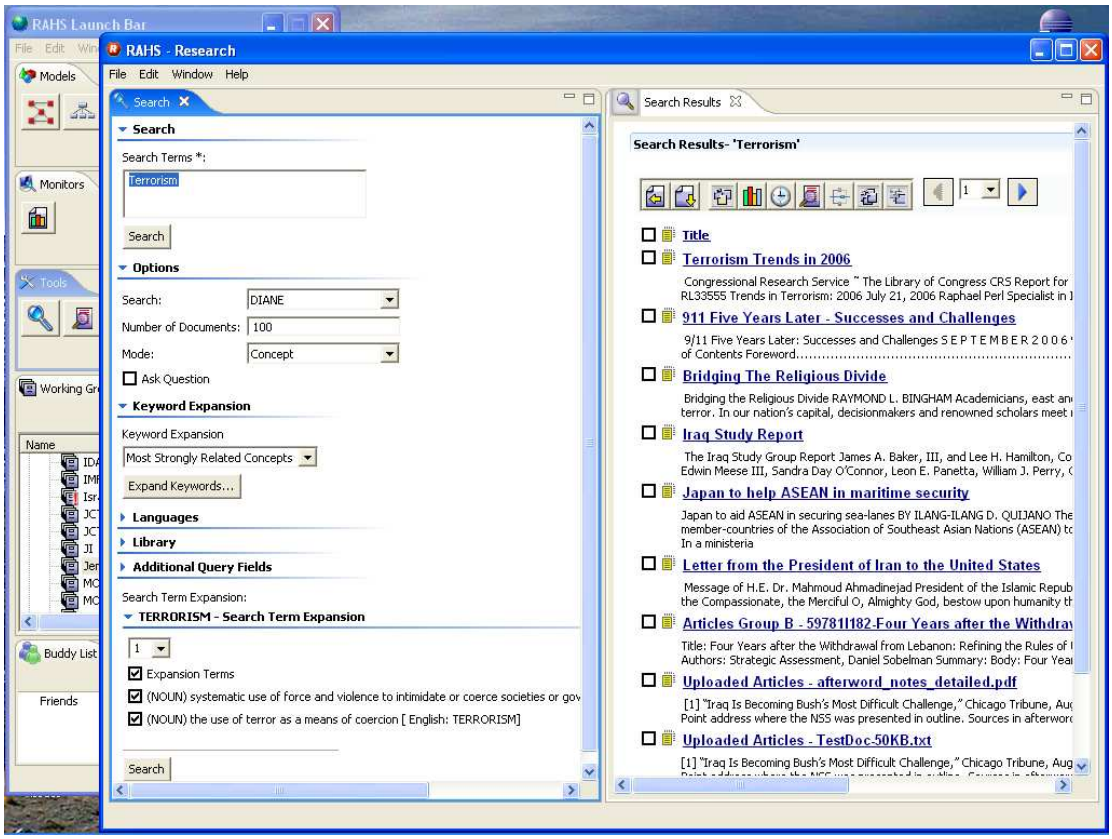

Figure 3 : Screen Shot of Search Tool

Summarisation Tool

A common problem faced by every analyst today is reading through a collection of articles, after finding out that only a couple of those really fall into the topic of research. Imagine all the time wasted on reading irrelevant information. The time can be better utilised for more meaningful analysis and findings. A multiple document summariser reduces the amount of reading by picking up key sentences that best summarise a single article or a collection of articles.

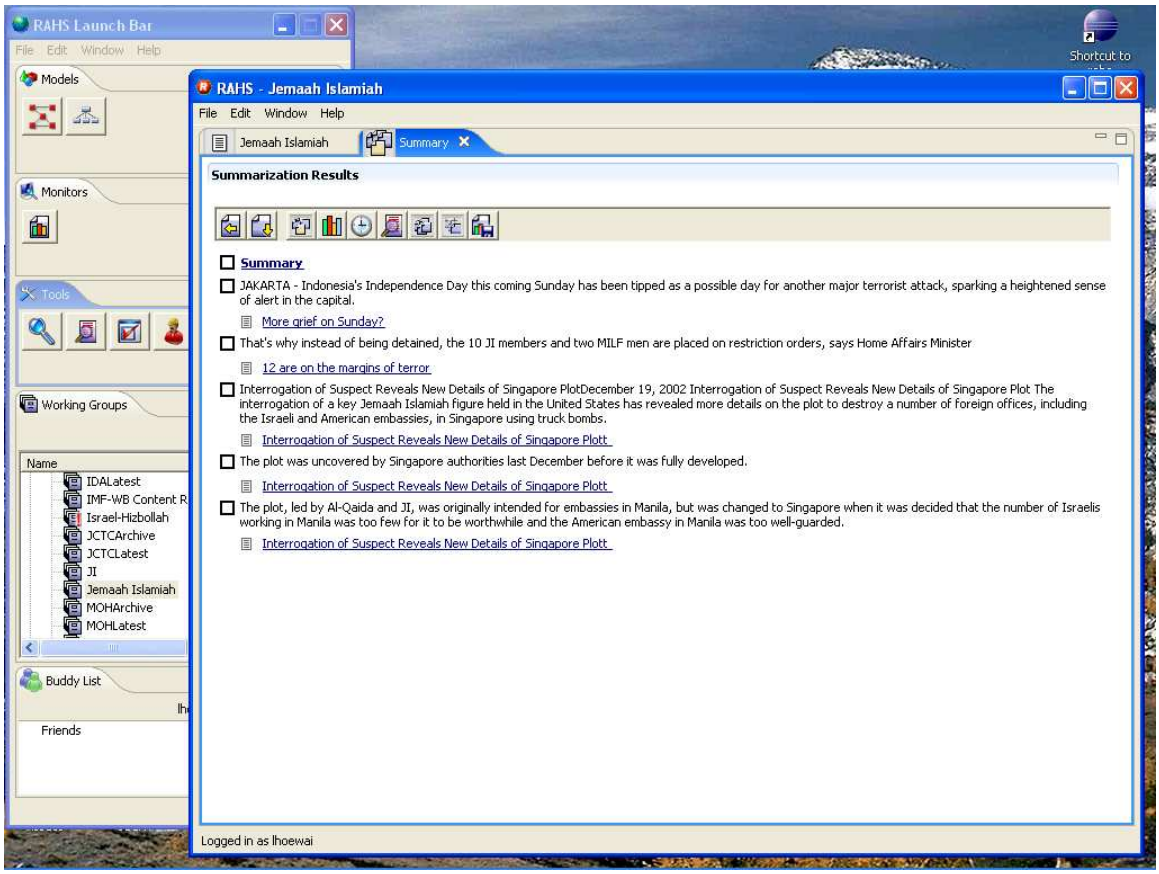

Figure 4 : Screen Shot of Summarisation Tool

Entity Analysis Tool

Entity Analysis is a process where locations, persons, and organisations, dates and times, monetary amounts and percentages are extracted. The evaluation is based on automatic word alignment of the speech recognition output through natural language processing of the English grammar and language.

The **Temporal Analysis tool** is a variation, which shows references to entities extracted from entity analysis tool over time. The time is associated with the publication dates of the document if it exists; otherwise, the date the article entered the RAHS system is used. The **Cross-sectional Analysis tool** is another variation that shows the entities that are mentioned in the same article as the entity that you are holding constant.

The **Keyword Analysis tool** is similar to entity analysis such that it counts the occurrences of "entities" on a set of documents. The major differences are that:

a) Keyword analysis is a process where user-defined "terms" or "phrases" are extracted from the set of documents. These user-defined terms are defined in a watch list.

b) Keyword Analysis is able to process "terms" in its native language (This assumes the English "terms" have the corresponding synonyms defined in native language in the watch list).

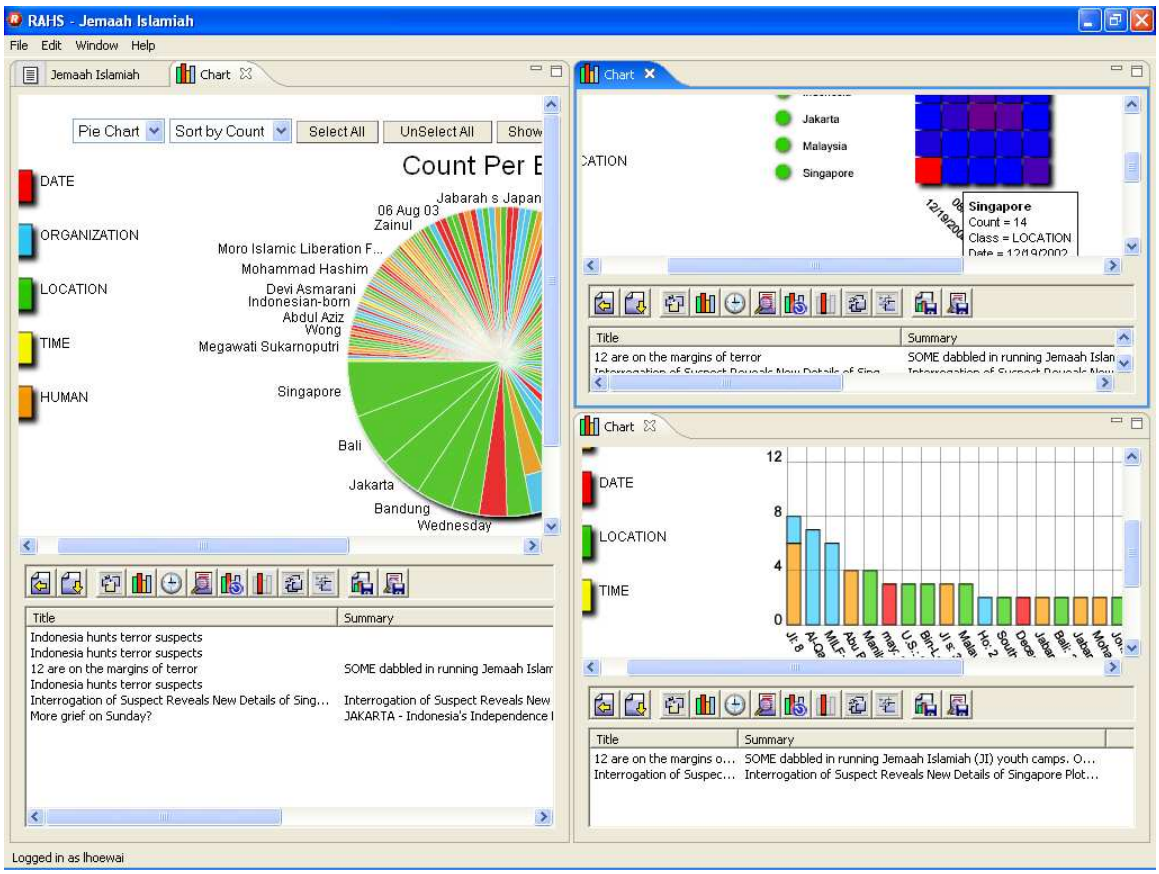

Figure 5: From clock wise, screen Shot of Entity Analysis (Pie Chart), Temporal and Cross-sectional analysis Tool on Eclipse RCP

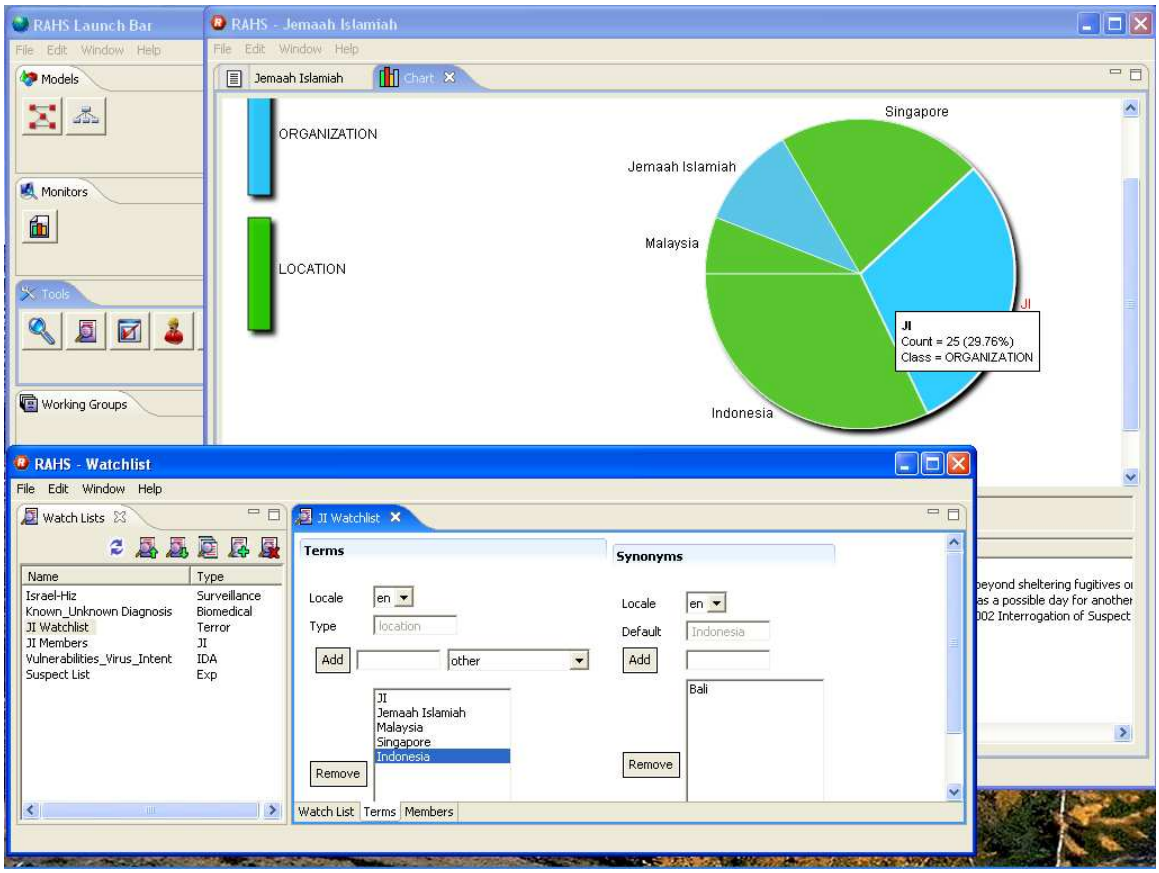

Figure 6 : Screen Shot of Keyword Analysis and Watchlist wizard

Timeline Analysis Tool

Timeline analysis extracts events (with a best estimate of when these events occurred) from a set of articles, using a set of user-supplied query terms. The analysis tool also assigns ranking scores to these events based on how significant these events are to the query terms. Rules are used to detect date expressions in sentences and resolve them to absolute dates using the creation dates as references. For example, “today” will be resolved to the article creation date, “September” to mean the last month of September before the article’s date, and “Sunday” to be the last Sunday before the article’s date. Such dates may be a single day, a whole month or a whole year. In the case where no date expression is detected in the entire sentence, the date is taken to be the date of creation of the article. Note that such rules only apply well to news articles where the reported events in the document source are happenings that occurred recently. The assumptions these rules have made do not apply well to other article types such as biographies, literary writings or historical texts.

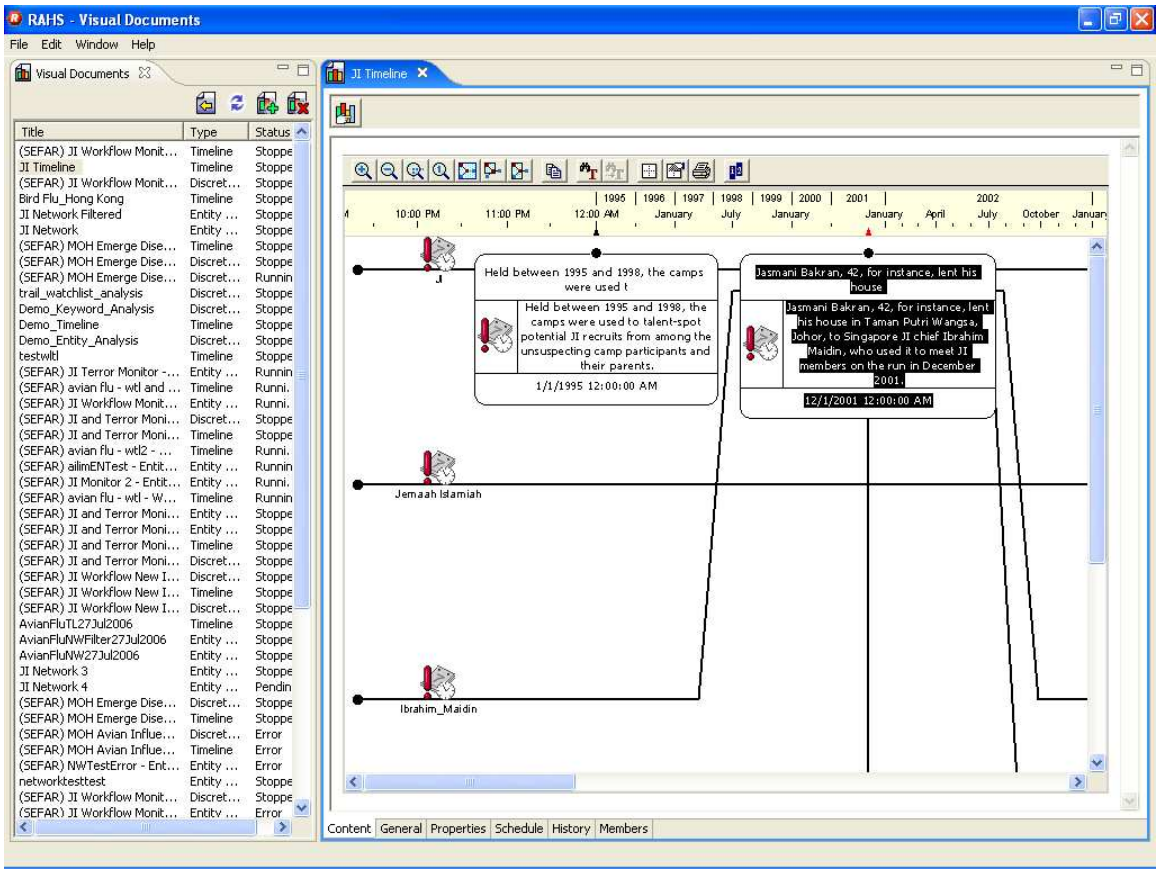

Figure 7 : Screen Shot of Timeline Analysis Tool

Network Analysis Tool

An entity is an object that can be a human, a location, an organisation or even a date. A link between two entities is the relationship between them. Whenever a person reads an article, he forms a mental model of how the entities mentioned in the article relate to each other. An entity network tool captures this relationship map in the form of a graph. RAHS has integrated a highly sophisticated entity and link extraction engine to linguistically process and present the captured information in a visually appealing format. As a result, analysts are able to apply link analysis and reveal knowledge that is obscured by the thousands of words.

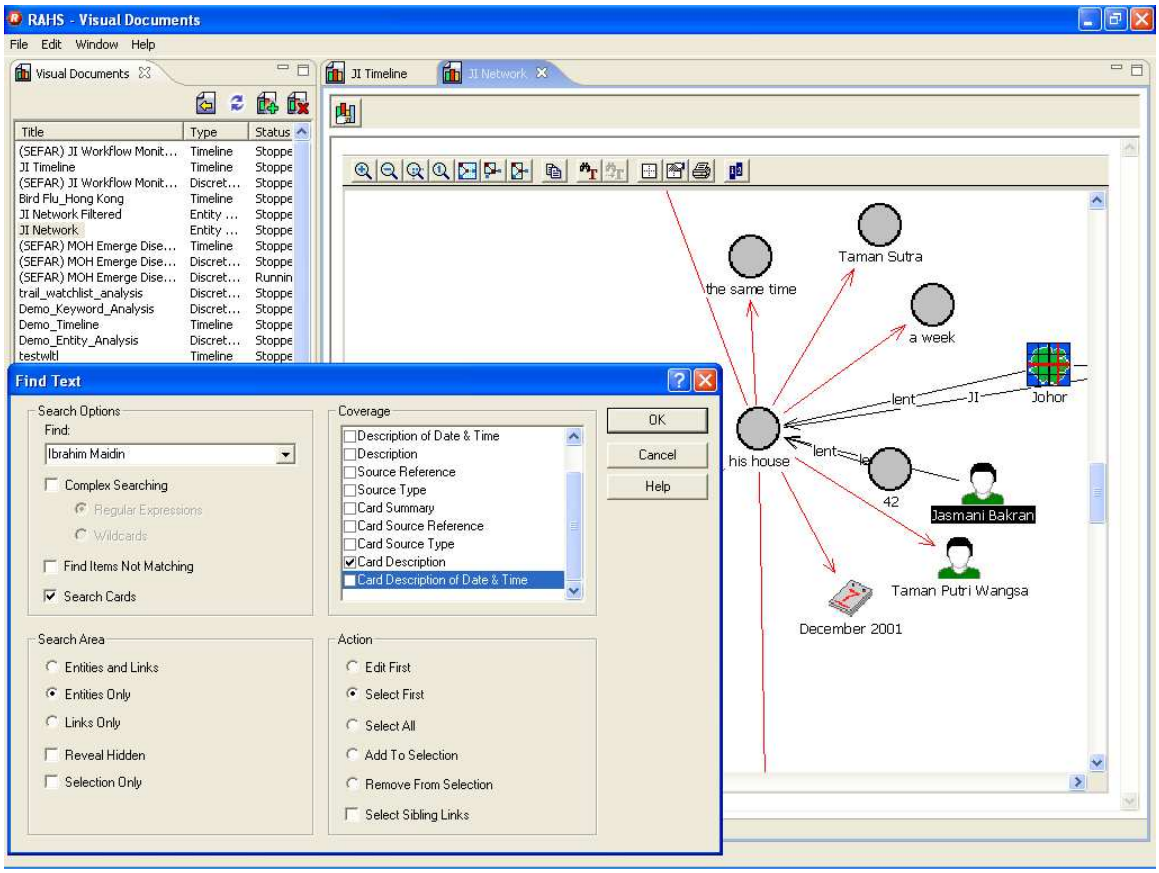

Figure 8 : Screen Shot of Network Analysis Tool

Automatic Categorisation Tool

As the frequency at which analysts place articles into their project, and as the number of folders in your project increases, they may find it increasingly time consuming to route articles into the appropriate folders. To ease this process, RAHS provides a personalised categorise that can automatically route articles and place into the project into the relevant folders. For each folder that has categorisation enabled, RAHS will remember which articles the analysts have placed into and moved out of the folder. The categorise will try to deduce each analyst preferences for which articles belong in which folder, so that in future it can automatically route documents into their respective folders.

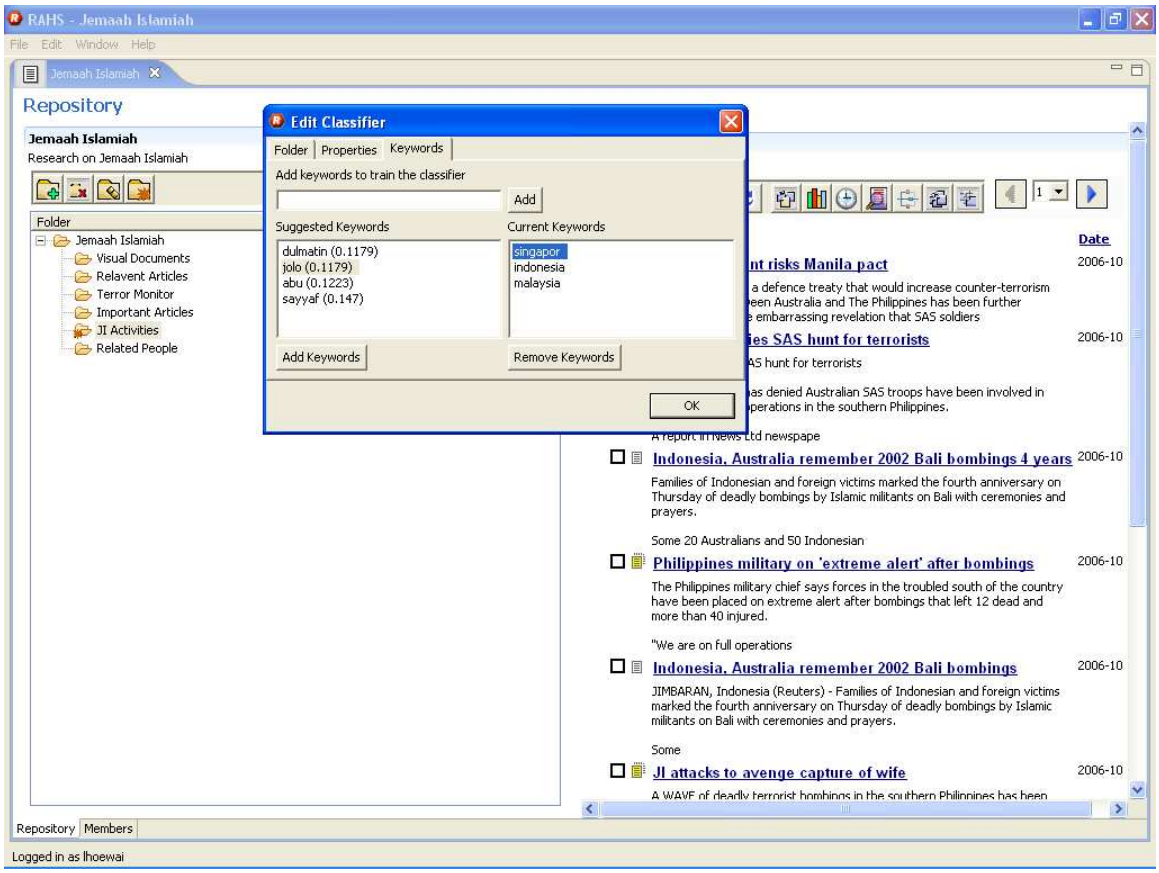

Figure 9: Screen Shot of Automatic Categorisation Tool

Clustering Tool

Clustering is an analysis process that arranges a collection of articles so that similar articles are grouped together, and articles in different groups are clearly distinct from one another. Such a group is called a cluster. The automated or manual analysis of information process can be used for:

a) Identifying hidden relationships between groups of objects from a search result to help you refine your search query.

b) Easing the process of browsing to find similar or related information from a search result to help analysts get to the information

c) Finding unique topics within a collection of articles. The unique topics might lead analysts to new trends or patterns that have so far not been mentioned in other articles

Duplicate Document Detection tool is a variation of Clustering. The process detects near duplicate documents based on the correlation of a group of words, which appears frequently in a document collection. The engine provides a user-defined parameter called Similarity Threshold to refine the results.

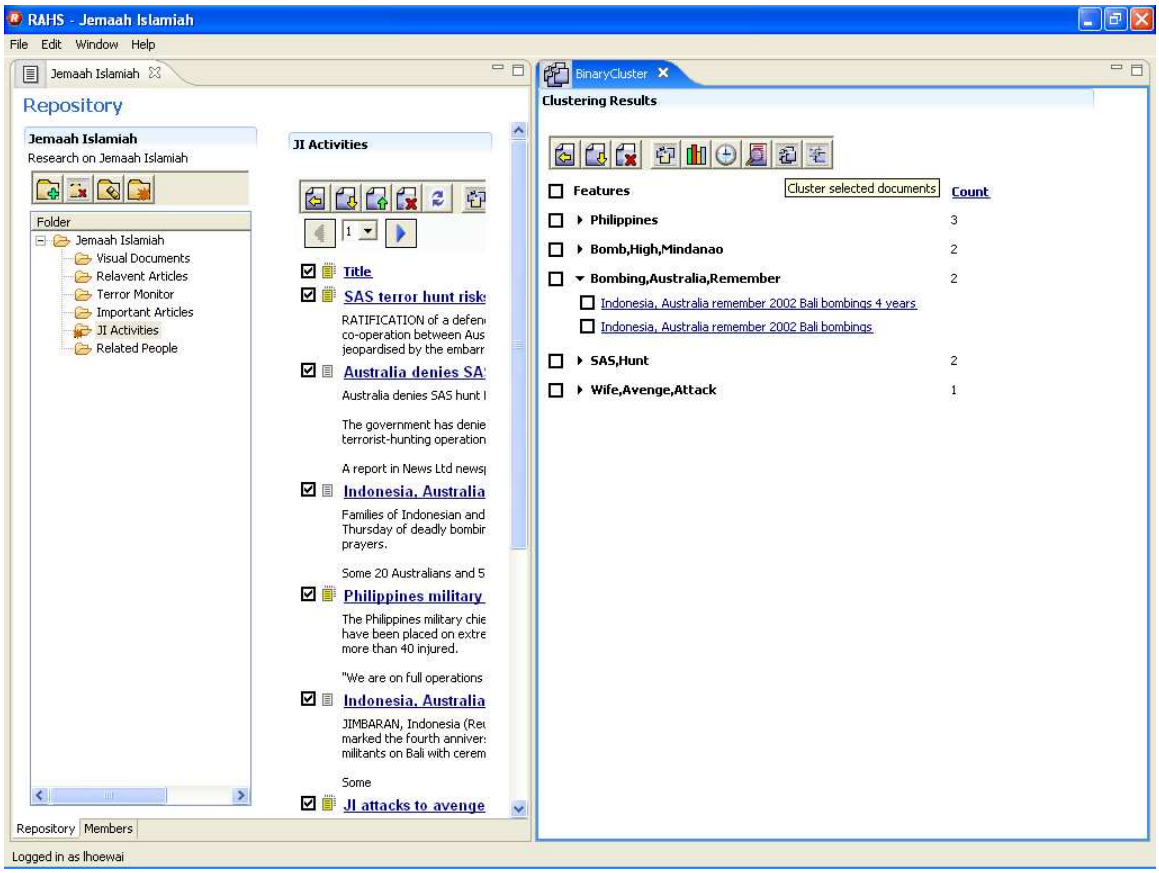

Figure 10: Screen Shot of Clustering/Duplicate Document Detection Tool

Model creation and monitoring

We are currently integrating tools to provide perspective sharing and data structuring capabilities into SOSA. In data structuring, we are incorporating capabilities to enable analyst to build model and enables collaborative modelling effort and thus allow users to connect across silos and challenge previous assumptions. Model monitoring capability will also be provided to enable matching with incoming data stream and to allow the human team to explore the interpretation and implication of the data.

Model Creation Tools

Our system provides a tool based on System Thinking to create system model called System Map to depict the user's understanding of the factors or components of the system/situation understudy as well as the relationships between these factors. In a System Map, the factors/components are represented as nodes and the relationships are depicted as links. RAHS allows the System Map to be created either individually or collaboratively. The System Map helps the users clarify their understanding of the system and serves as a basis for in depth discussion and sense making. A set of tools, such as one to help find out the most influenced and the most influential factors, are provided to help the

users derive insight about the system from the System Map. Fig 11 shows a System Map and its Influenced/Influential Map.

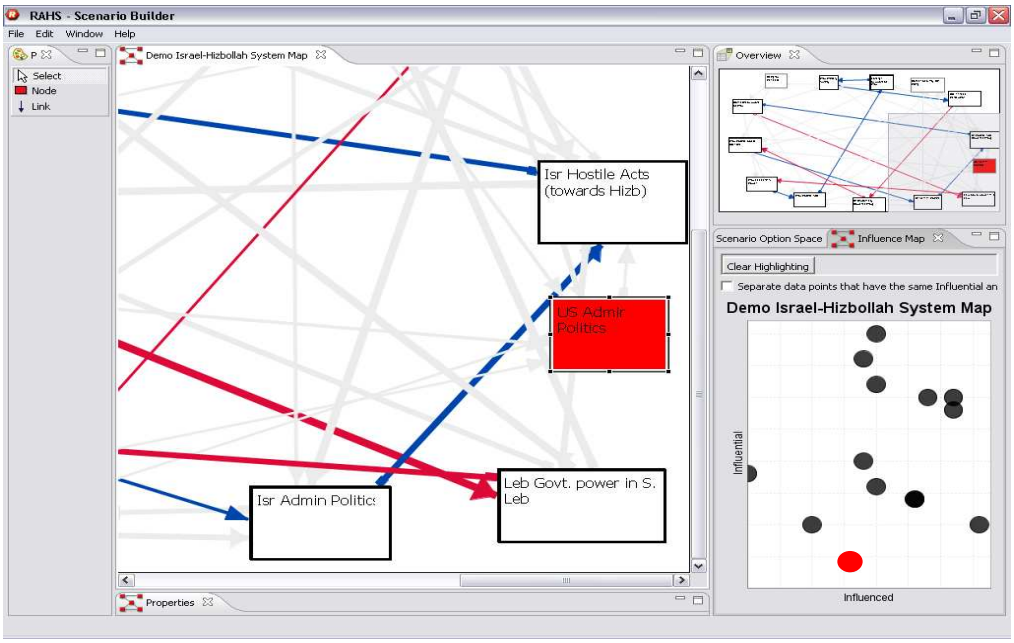

Figure 11: A System Map and its Influenced/Influential Map

We can further enter the states for each factor in the System Map and the pair-wise consistency between each pair of states of two different factors. After the users have entered the options for the factors and the consistencies, the RAHS system will be able to carry out Morphological Analysis and generate possible scenarios based on the user understanding of the situation.

Monitoring

To monitor the situation, user can also create filters for each of the factors so that incoming data stream can be extracted out and match to each of the factor. If the input data indicate a situation different from the model depicted by the System Map and the consistency constraint, the users can improve the model. The data input can also help us to monitor the developer of the situation and narrow down the possible scenarios that can happen in the near future.

Perspective Sharing

RAHS also allows the users to enter their perspective and share their perspective across the whole network. User can enter their perspective of a data item by meta-tagging the articles. The meta-tag not only facilitates retrieval of the data but when the meta-tags of all the relevant data are viewed together it

will help the users to discover outliers. The meta-tag of the other users will provide the user with other perspectives, thus prevent the user from getting blind-sided through premature convergence.

Indexing Tool

RAHS provides indexing software for the analysts to meta-tag the articles. Four types of meta-tags are provided: Filters, Questions, Comments and Keywords. Filters and Questions are meta-tags whose meanings are understood by every one involved in the sharing. Comment and Keywords are freeform meta-tags that are user specific. User can select the Filters and the extent to which each Filter best describe a piece of data by select the range of value for each selected Filter. Each Question has a set of predefined options. User can select the Question and the option best describe the data. Comment meta-tag allows the users to enter freeform text. Keyword meta-tag allows entry of multiple user specific words or sentences relating to the data.

Perspective Visualization

The meta-tags entered are representations of perspective. RAHS provides a set of visualization tools specially designed to help the users analyze the perspectives and data. Six types of visualizations are provided to the users. One has a web based user interface that shows the number of data indexed by each Filter and each option in the Questions. Another allows the users to compare multiple pieces of data concurrently. Some visualisations enable the user to see the meta-tags based on combinations of multiple Filters and options.

Lessons Learnt

A spiral development approach was adopted for the development of the system and a test bed was created for the users to try out the new software and concepts. Training and exercise were also carried out to familiarisation the users with the new tools and processes and to encourage the formation of informal networks across different agencies. Some of the lessons learnt during the development of the system are as follows:

Conceptual vs. Operational. It was a challenge translating the concepts into system design. Many of the techniques had shown success in workshop setting, but it was difficult to translate them into tools, which could be easily used for day-to-day usage without heavy facilitation. We sought an agile development approach, iterating between concepts, technical implementation

and user prototyping. An underlying principle adopted was to design the system such that it could be used without an expert knowledge of the underlying theories and concepts.

Broad-based vs. Agency Specific. It was immensely challenging implementing a common language for perspective sharing. There are a few existing approaches to establishing such a folksonomy. Our approach was to conduct a workshop process with the different user agencies to determine a set of common indices that the agencies could use in the context of national security. This approach is less technology dependent, but there will be the need to regenerate the common indices with the inclusion of new agencies.

User Engagement There is a distinct difference between functionality and usability. Implementing a rich interface is just as important as delivering functional capabilities to the end-user. The development focus should be properly communicated to users. We have focused primarily on capability development in the early stages of the system, and also involved the eventual end-users at an early stage to create ownership and buy-in. Subsequently, there was a need to manage user expectations on the turn-around time to incorporating usability enhancements as well as functionality required by specific agencies only.

Security vs. Functionality There was a trade-off between security and functionality. A horizon scanning system requires the information richness of the open-source. Nevertheless, the need for operational security has meant that the project team has built the system in a secured intranet, with a one-way channel for open source information.

Conclusion

The RAHS system is a network of people, tools and data. It adopts a wide range of methodologies to cater for different situations. Its open architecture enables it to evolve continuously to connect with new agencies and incorporate new concepts and technologies. The system will be deployed by end 07. The same architecture can be applied in an open network to leverage on a wider number of people, expertise and data both locally and internationally to provide early warning on National Security threats.