

A Visualization Dashboard for Muslim Social Movements

by

Shreejay Nair

A Thesis Presented in Partial Fulfillment  
of the Requirements for the Degree  
Master of Science

Approved March 2012 by the  
Graduate Supervisory Committee:

Hasan Davulcu, Chair  
Partha Dasgupta  
Arunabha Sen

ARIZONA STATE UNIVERSITY

May 2012

## ABSTRACT

Muslim radicalism is recognized as one of the greatest security threats for the United States and the rest of the world. Use of force to eliminate specific radical entities is ineffective in containing radicalism as a whole. There is a need to understand the origin, ideologies and behavior of Radical and Counter-Radical organizations and how they shape up over a period of time. Recognizing and supporting counter-radical organizations is one of the most important steps towards impeding radical organizations.

A lot of research has already been done to categorize and recognize organizations, to understand their behavior, their interactions with other organizations, their target demographics and the area of influence. We have a huge amount of information which is a result of the research done over these topics. This thesis provides a powerful and interactive way to navigate through all this information, using a Visualization Dashboard. The dashboard makes it easier for Social Scientists, Policy Analysts, Military and other personnel to visualize an organization's propensity towards violence and radicalism. It also tracks the peaking religious, political and socio-economic markers, their target demographics and locations. A powerful search interface with parametric search helps in narrowing down to specific scenarios and view the corresponding information related to the organizations. This tool helps to identify moderate Counter-Radical organizations and also has the potential of predicting the orientation of various organizations based on the current information.

To Mummy, Papa, Chotu and Anusha

## ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my advisor Dr. Hasan Davulcu, for his guidance and support. He has always been a source of encouragement and inspiration. I also want to thank Dr. Arunabha Sen and Dr. Partha Dasgupta for being a part of the thesis committee and for taking the time to review this thesis.

I would like to thank Sukru, for his valuable technical insights. His suggestions for design and implementation of the components were very helpful. I would also like to thank Sedat, Dananjayan and all the members of the CIPS lab for their help and support.

There aren't enough words to thank Anusha, who has always been by my side with her love, encouragement and occasional kicks in the back to keep me on track.

I want to thank my family for their constant love and support through all these years.

Finally I would like to thank all my friends here at ASU for making this a memorable stay.

# TABLE OF CONTENTS

	Page
TABLE OF CONTENTS . . . . .	iv
LIST OF FIGURES . . . . .	v
CHAPTER . . . . .	1
1 INTRODUCTION . . . . .	1
1.1 Motivation . . . . .	1
1.2 Document Outline . . . . .	1
2 MINERVA RESEARCH . . . . .	3
3 SYSTEM ARCHITECTURE AND COMPONENTS . . . . .	5
3.1 Data Collection . . . . .	5
3.2 Document Database . . . . .	7
3.3 Scale Construction and Baseline Knowledge . . . . .	8
3.4 User Interface . . . . .	10
4 Visualization Dashboard . . . . .	12
4.1 Search and Navigation Widget . . . . .	13
4.2 Trends and Timeline Widget . . . . .	15
4.3 Quadrant Widget . . . . .	20
4.4 Demographics and Markers Treemap . . . . .	22
4.5 Heat Map . . . . .	23
5 SCENARIOS AND FINDINGS . . . . .	28
5.1 Radical Organizations' Trends . . . . .	28
5.2 C-Quadrant Organizations' Trends . . . . .	28
5.3 A-Quadrant Organizations' Trends . . . . .	30
5.4 B-Quadrant Organizations' Trends . . . . .	31
6 Summary . . . . .	34
7 Future Work . . . . .	35
REFERENCES . . . . .	37

## LIST OF FIGURES

Figure	Page
2.1 The quadrants model . . . . .	4
3.1 An overview of system architecture . . . . .	6
3.2 A snippet of the processed data in XML . . . . .	7
3.3 Radical subset of organizations and keywords, sorted according to aggregate row values. . . . .	9
3.4 Counter-radical subset of organizations and keywords, sorted according to aggregate row values. . . . .	10
3.5 Radical subset of organizations and keywords . . . . .	10
3.6 Counter-radical subset of organizations and keywords . . . . .	11
4.1 A sample snapshot of the web application . . . . .	12
4.2 Political markers for the period between 2000-2006 . . . . .	16
4.3 Markers Categories . . . . .	17
4.4 Schema file various categories of markers . . . . .	18
4.5 Schema file showing positiveSentiment and negativeSentiment field . . . . .	18
4.6 Increase and decrease in radicalism denoted by red and blue lines in quadrant . . . . .	21
4.7 Radical and Counter Radical fields in SOLR schema . . . . .	21
4.8 Tag clouds in Navigation widget . . . . .	22
4.9 Demographics and Markers Treemap . . . . .	24
4.10 Heat Map - shows the intensity map based on the number of times a location is mentioned in the corpus . . . . .	25
4.11 Heat map showing the locations mentioned for organization NU and location USA . . . . .	26
4.12 Schema file showing entries for location SOLR database . . . . .	27
5.1 Trend of radical markers . . . . .	28
5.2 Consistent rise of FPI on the radical scale . . . . .	29

Figure	Page
5.3 "Ahmadiyya" peaking during the period 2006 - 2010 . . . . .	29
5.4 "Khilafah" ideology of Hizb ut- Tahrir . . . . .	30
5.5 Decline of HTI on radical scale . . . . .	31
5.6 Counter Radical markers associated with CounterRadical organizations	32
5.7 Radical markers associated with CounterRadical organizations . . . . .	33

## Chapter 1

### INTRODUCTION

The Visualization Dashboard, which is the outcome of this thesis, can be used extensively to understand the origins of an organization, their inclination towards various ideologies and practices, their propensity towards radicalism and violence over a period of time, their target audience and members. It also provides a powerful parametric text search, giving the user a quick and easy way to search for information and narrow down to specific scenarios and events.

#### 1.1 Motivation

This thesis is part of the Minerva Research “Finding Allies for the War of Words: Mapping the Diffusion and Influence of Counter-Radical Muslim Discourse”.

Radicalism is the belief in the ideology which favors profound changes to the current social status. Radicalism may also support the use of violence to bring about a dramatic change in the existing social order. This research aims to map counter-radical and radical concepts and trends, assess the impact, popularity and spread of voices and content. One of main goal of this research is to have the ability to identify a counter-radical organization using their textual discourse.

As a part of this research a lot of information was collected and there was a need to visualize this information in a fast and easy manner. The visualization dashboard provides an interactive way of visualizing all this information.

#### 1.2 Document Outline

The rest of the document is arranged as follows:

Chapter 2 briefly describes the research goals of the Minerva Research project and how this thesis contributes to this research.

Chapter 3 describes the System Architecture which includes details about data collection, pre-processing, information extraction using NLP tools and parametric



search.

Chapter 4 gives an overview of the Visualization Dashboard and explains the details of its various components.

Chapter 5 discusses use cases with specific scenarios related to radical and counter-radical organizations.

Chapter 6 gives a brief summary and possible utilization of the dashboard.

Chapter 7, Future work, describes the possible new features and improvements, both to the interface and the back-end processes.

## Chapter 2

### MINERVA RESEARCH

Muslim extremism is a global threat and there are a variety of strategies being applied to counter and contain it. “The War of Words”<sup>1</sup> research, which is a part of the Minerva Initiative<sup>2</sup> aims to understand, categorize and recognize radical as well as counter-radical organizations and their dynamic interactions and often overlapping behavior towards ideologies and practices.

The topics, which have been studied as a part of this project includes:

- (a) Identifying organizations and collecting their textual discourse by crawling various websites [4]. Primary sources for this data were the organizations’ websites and blogs.
- (b) Identifying most frequent “marker” phrases corresponding to people, location names, demographics and social, political, economic, and religious issues from the crawled text data [4].
- (c) Sentiment Analysis on the extracted markers showing polarity of organizations towards these markers [4].
- (d) Cluster the organizations based on their polarity towards various markers [4].
- (e) An orthogonal model to classify organizations into four different quadrants based on their **Change Orientation** - what degree of change does the movement seek to achieve and **Change Strategy** - To what degree does the movement advocate violence. A quadrant system was developed, as shown in Figure 2.1, which showed the organizations on two dimensions, namely, Counter Radical/Radical and Violent/Non-violent [15]. (more details in section 3.3 and 4.3) .

---

<sup>1</sup><http://minerva.dtic.mil/funded.html>

<sup>2</sup><http://minerva.dtic.mil/>

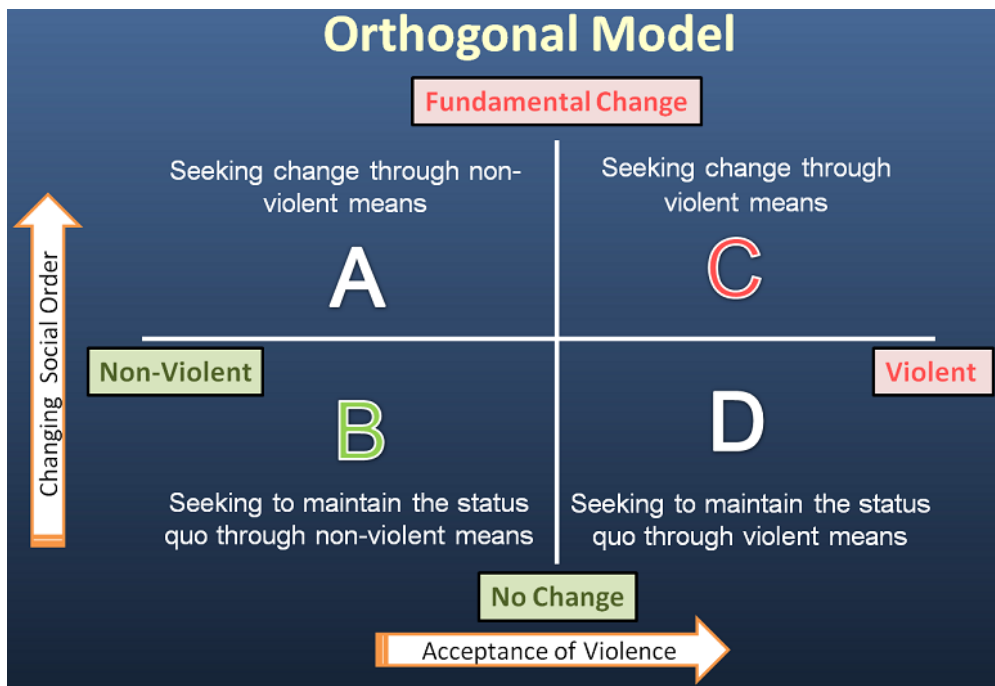


Figure 2.1: The quadrants model

The visualization dashboard gathers a lot of information, which has been made available as a result of the research on the above topics. The dashboard utilizes this information and provides it to the user, with a interactive web interface and a powerful parametric text search.

## Chapter 3

### SYSTEM ARCHITECTURE AND COMPONENTS

This section briefly explains the proposed framework, which encompasses all the processes and components of the Visualization Dashboard. An overview of the System Architecture is given in Figure 3.1. It consists of four major components:

- (1) Data Collection: performs web crawling, keyword extraction and sentiment analysis.
- (2) Document Database: stores the information extracted from the above processes.
- (3) Scale Construction and Baseline knowledge: contains the algorithm generating the social scales automatically based on keywords in text. Baseline knowledge consists of expert rankings and rankings generated using two other baseline methods, one based on simple sorting and another based on principal component analysis [7].
- (4) User Interface: Contains an interactive website which displays important information pertaining to radicalism/violence level of organizations, demographics, trends and location information on a heat map.

The subsections from 3.1 to 3.4 will briefly explain these processes. (More details can be found in [14], sections 3.1 to 3.5)

#### 3.1 Data Collection

The first step in data collection is identifying a set of radical and counter-radical organizations. Social scientists from our research group identified a set of 23 Indonesian religious organizations. The variance in the organizations is based on a bipolar **Radical** vs. **Counter-radical** scale [15]. The methods discussed in the paper can be easily modified to build other relevant scales. The next step was to crawl

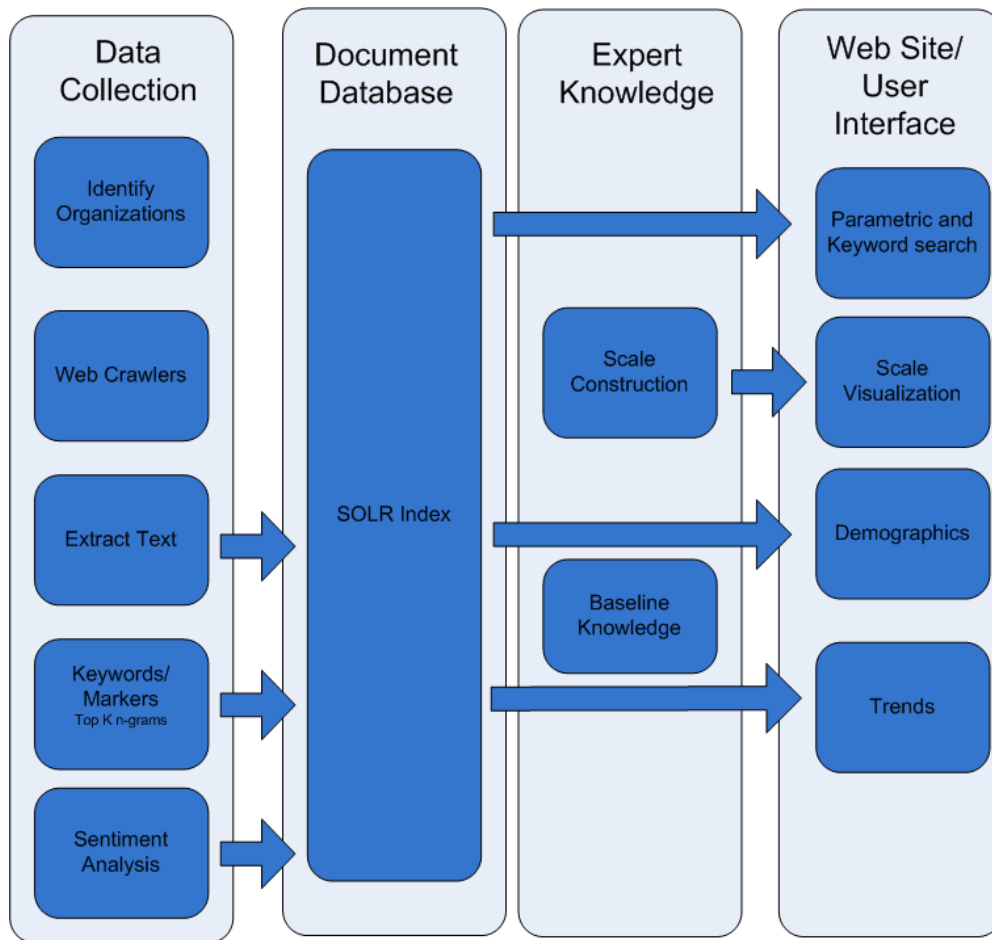


Figure 3.1: An overview of system architecture

websites related to these organizations. Here we focused only the blogs and official websites of these organizations. A total of 37000 articles, published between 2001 and 2011, were downloaded from these websites. A REGEX filter was used to remove the HTML tags and extract only the text between these tags. For each organization's text corpus, the top-k n-grams were extracted. These top-k n-grams were combined and presented to social scientist for feature selection. Further processing on the text data involves an NER which automatically extracts the Names, Locations and Persons from this text. Sentiment analysis is also conducted on the features (selected by social scientists), using an Indonesian POS tagger. Other information extracted from the text includes, the date on which article was written, the title of article, and the URL from which this article was extracted. All this data is

```

<field name="source">LKIS</field><field name="type">CounterRadical</field>
<field name="URL">http://lkis.or.id/index.php?option=com_content&amp;view=article&amp;</field>
<field name="date">Monday, 14 December 2009 09:44</field>
<field name="content">Waktu | 28 - 31 Januari 2009
Lokasi | Hotel Bintang Fajar, Yogyakarta
<field name="title">Amina Wadud: Intensif Workshop</field>
<field name="positiveSentiment">asasi manusia</field><field name="positiveSentiment">budaya</field>
<field name="LOCATION">Dunia Islam</field>
<field name="LOCATION">Indonesia</field>
<field name="PERSON">Usaha</field>
<field name="PERSON">Tema</field>
<field name="positiveSentimentENG">human rights</field><field name="positiveSentimentENG">culture</field>
<field name="event_dt">2009-12-14T00:00:00Z</field>
<field name="CR">politik,asasi manusia,pluralisme,toleransi</field><field name="CRENG">politics,human rights,plu
<field name="R">kekerasan</field><field name="RENG">violence</field><field name="allmarkers">yogyakarta</field>

```

Figure 3.2: A snippet of the processed data in XML

also converted to ENGLISH using machine translation. This data is stored in XML files. A snippet from one of these XML files is shown in Figure 3.2. These XML files serve as a backup also, in case of issues with the database or index corruption.

### 3.2 Document Database

The visualization dashboard requires a varied amount of data, to display on its widgets. Each widget requires a separate amount and type of data from the information we have collected. We needed a database which would allow us to execute multiple queries on huge datasets without affecting the retrieval speed.

Most of the data we collected does not have any structure. We are also adding a lot of new fields during pre-processing based on the requirements. Due to the changing structure of the data, it was not practical to index this data using traditional RDBMS. We selected SOLR<sup>1</sup> an enterprise level search platform. SOLR provides an easy way to index large amounts of data without affecting the retrieval speed. It also provides a flexible query syntax, which can be used to extract very specific data. SOLR provides a lot of options while indexing and searching, which can be used to get more relevant results. For example, SOLR can be easily con-

<sup>1</sup><http://lucene.apache.org/solr/>

figured to search not only a particular keyword, but also the stem or synonym of the keyword. It has many advanced search features. We specifically used the features related to faceted navigation, range queries (for extracting document belonging to a time range), and nested queries. SOLR has APIs<sup>2</sup> in almost all modern programming languages like Java, Python and JavaScript, which makes it very easy to integrate with an existing system. We used `ajax-solr`<sup>3</sup>, an AJAX based JavaScript library, to communicate with SOLR. AJAX communication lets us send multiple queries simultaneously. This makes sure that the dashboard continues to load as and when the various queries have been processed, instead of waiting for all queries to finish executing.

A part of an existing SOLR database<sup>4</sup>, was used in this thesis. A lot of changes were incorporated into this database. This included changes to the schema file<sup>5</sup> of SOLR for addition of new fields. SOLR comes with a jar file<sup>6</sup>, which indexes data from XML files, mentioned in section 3.1. Once new fields are added it is re-indexed again, which takes a few minutes for a collection of around 50,000 documents.

A python program was written which had the function of adding new fields or modifying existing ones. A detailed description of the changes for each of the widgets is given in sections 4.1 to 4.5 , under the heading “**Changes to SOLR**”.

### 3.3 Scale Construction and Baseline Knowledge

Scale construction is the main idea behind the quadrant widget which displays the trajectory of organizations on a radical/counter-radical scale. Scaling is the process where both social movements (subjects) and their socio-economic, political, reli-

---

<sup>2</sup>Application Programming Interface

<sup>3</sup><https://github.com/evolvingweb/ajax-solr>

<sup>4</sup>This database was used to build a Parametric Search website on the Indonesian data [13]

<sup>5</sup>schema file is similar to table description in RDBMS. More details on SOLR schema can be found at <http://wiki.apache.org/solr/SchemaXml>

<sup>6</sup><http://lucene.apache.org/solr/tutorial.html> - The tutorial on this webpage gives more details on how to use this jar file for indexing

RADICAL																														
	quran	islamic teachings	Shari'a	infidel	military strength	Kufr	Caliph	prophet muhammad	prophet muhammad	Tabligh	sharia	corruption	Shirk	paganism	die shaheed	street preaching	obligation for Shari'a	Mujahideen council	the Zionist Jews	Islamic thoughts	pornography	jihad council	mujahideen	suicide bomb	communists	the idolaters	Sharia enforcement	violence	Islamic state	
Org 9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	0	1	0	0	
Org 7	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	1	0
Org 1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0	1	
Org 3	1	1	1	1	0	0	0	1	1	1	0	1	0	0	1	0	1	1	0	1	1	0	0	0	0	0	0	0	1	0
Org 4	1	1	1	1	0	1	0	0	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
Org 8	0	0	1	0	1	1	1	0	0	1	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	1
Org 2	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Org 5	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Org 6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Org 10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 3.3: Radical subset of organizations and keywords, sorted according to aggregate row values.

gious beliefs, goals and practices (items) are scaled simultaneously on the same scale.

The first step in constructing the scale is the selection of a candidate set of keywords. The top 100 keywords for each organization were selected from the list of top-k n-grams<sup>7</sup>. Social scientists then manually selected keywords which belonged to the categories **social, politics, economics and religion**. This process gave us 790 keywords, out of which 29 and 26 were selected by experts to be included in the radical/counter-radical scale. The frequency of each keyword (radical and counter-radical) is extracted from the document corpus and a response table is built. Figure 3.3 and 3.4 show sample response tables. These response tables are given as input to a *Rasch Model building* algorithm.

The algorithm creates a metric to validate the fitness of the model and rankings of the organizations and keywords. Figures 3.5 and 3.6 show the relative positions of the organizations and keywords on the latent scales.

The quadrant widget uses this scale to plot the organization’s trajectory on the quadrant.

<sup>7</sup>These are the same top-k keywords mentioned in section 3.1



COUNTER-RADICAL	politics	equality	Hindu	pluralism	state	tolerance	gender justice	human rights	Indonesian Islam	religious freedom	intercultural dialogue	constitutional rights	democracy	election	Christian	non-Muslims	civil society groups	liberal	secularism	activists	Jewish	peace building	conflict prevention	freedom of expression	homosexual	multicultural education
Org 12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	1	0	0
Org 21	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	0	0	1	0	0	0	0	0	0
Org 15	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0
Org 11	1	1	1	0	1	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0	1
Org 17	1	0	1	1	1	0	0	1	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Org 14	1	1	0	0	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Org 16	0	1	1	1	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Org 19	1	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0
Org 18	0	1	1	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Org 20	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Org 13	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Org 22	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Org 23	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 3.4: Counter-radical subset of organizations and keywords, sorted according to aggregate row values.

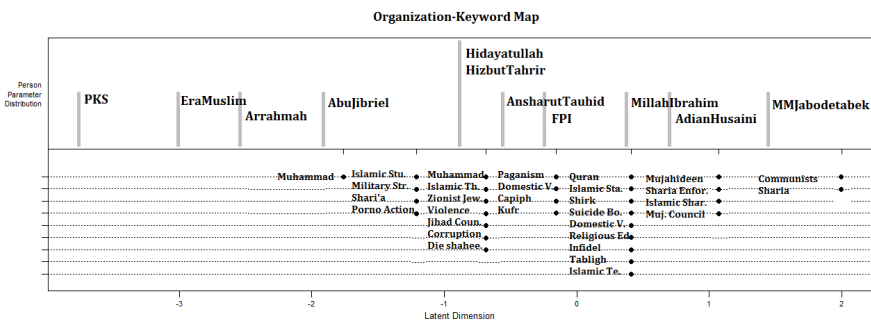


Figure 3.5: Radical subset of organizations and keywords

Rankings are also generated using expert knowledge. A tool<sup>8</sup> was built which would capture the rankings given by experts. Two other baseline methods were also used for ranking, one based on simple sorting and another based on principal component analysis [7]. Details about the entire process can be found here [14].

### 3.4 User Interface

We needed an intuitive way to access all the information we had extracted. We also wanted to give the user the ability to do complex queries on this data using a graphical interface.

<sup>8</sup>This tool was built by Sukru Tikves, as a part of [14]

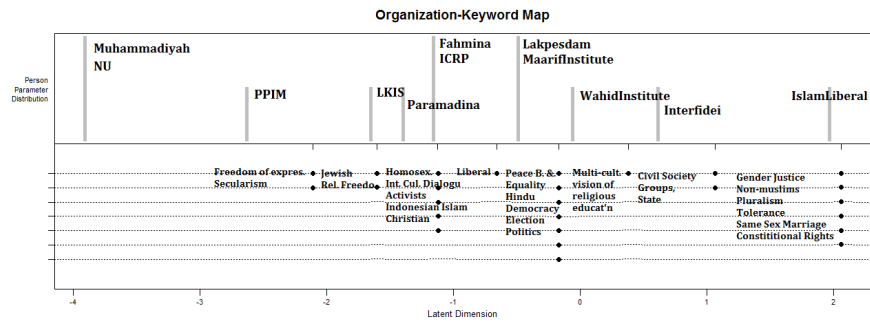


Figure 3.6: Counter-radical subset of organizations and keywords

A web interface was created which had the Parametric Search as its driver. We also added some widgets like treemaps and annotated timeline, to give a better visualization of the data. The parametric search provides a keyword search, which along with the faceted navigation, makes it very easy for the user to narrow down to specific results. We also provided some other useful options such as, timeline widget, which allows selecting results belonging to a specific time range and a markers menu, which lets the user see only specific categories of markers.

## Chapter 4

### Visualization Dashboard

This chapter explains the various widgets of the dashboard. It explains the purpose of each widget, its construction and its usage. A sample screenshot of the dashboard is given in Figure 4.1.

It comprises of the following components:

- Search and Navigation widget (top left): This allows filtering of results using parametric queries and keyword based search. We also have a heat map, which can be accessed by clicking on the tab “Heat Map”.
- Quadrant widget (top right)<sup>1</sup>: For the given result set, this widget displays the corresponding organizations on a two dimensional scale (radicalism axis and violence axis).

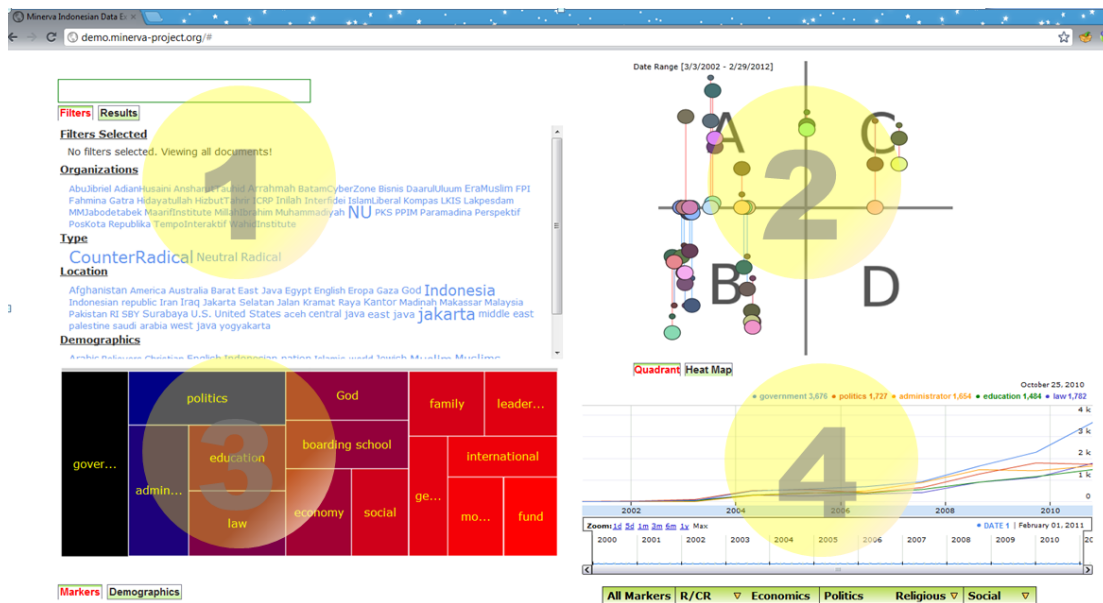


Figure 4.1: A sample snapshot of the web application

<sup>1</sup>select tab **Quadrant**

- Treemap widgets (bottom left): This widget displays the demographics and top markers for the particular user query<sup>2</sup>.
- Timeline and Peaking markers (bottom right): This widget provides a visualization of the markers trends on a timeline. The timeline widget also lets the user control the time range of the results of a particular query.

The navigation of the dashboard starts with the Navigation widget (top-left). The user can filter down the corpus using the powerful text search. Faceting on fields such as Demographics and Location can further narrow down the results. The next subsections will provide a detailed explanation of the widgets, how they work, which other widgets control them and the queries being used by them.

#### 4.1 Search and Navigation Widget

As a part of Minerva project, [13] showed a powerful way to search through the Indonesian corpus using faceted navigation. The document corpus for this search was indexed using SOLR. We made a few changes to the index and the SOLR queries to include the extra fields which were required for the other widgets<sup>3</sup>. A lot of existing queries were modified to improve performance. For example, selecting any one of the parameters in the filter tab, would trigger a query which was fetching many fields in a particular document. For queries which returned a few thousand documents, this data was running into a few megabytes in size. This was useful for the parametric search, but not all widgets needed all these fields. We changed the queries to return only those fields which were required by the widgets of the dashboard. More details about the changes in SOLR index and schema, changes to existing queries and addition of new queries will be discussed in this section and through sections 4.2 to 4.5.

---

<sup>2</sup>select the tab **Demographics** or **Markers** to toggle between the two treemaps.

<sup>3</sup>refer section 3.2

This widget consists of two tabs **Filters** and **Results**. “Filters” tab contains various facets and a search box. The facets provided are:

- Organizations
- Type of Organization (Radical or Counter-Radical)
- Location
- Demographics

Users can also use the search box to do a keyword based search or a combination of keyword and facet search. The “Result” tab can be used to see the articles which match these search criteria.

**Keyword based search:** When a keyword is typed into the search box, a SOLR query is used without any other parameters. A query which processes a text search is shown below for a text search of the word “prayer”.

#### Listing 4.1: Example of a single keyword search

```
http://xx.xx.xx.xx:xxxx/solr/select?q=text%3Aprayer&wt=json
```

The above query would return all the documents which contain the word “prayer”. The entire document with all its fields will be returned. This is a huge overhead and slows down the application considerably. SOLR gives us an option to return only specific fields.

For the above query to display the tag clouds, only the faceted information of the fields “Demographics”, “Location”, “Organizations” and “type” is required. This can be achieved by a query as follows:

#### Listing 4.2: Example of faceted query

```
http://xx.xx.xx.xx:xxxx/solr/select?facet.enum.cache.minDf=35
&rows=5&q=%3A*&facet=true&facet.field=source&
facet.field=type&facet.field=demographicsENG&
facet.field=LOCATIONENG&facet.limit=35&facet.mincount=1&
f.source.facet.limit=35&json.nl=map&fq=text%3Aprayer&wt=json
```

Every time a user selects a new facet or types a new word in the search box, the above query is modified by adding the corresponding facet or word to the query. The original query is preserved and the facets are added as another parameter.

**Control:** The Navigation widget controls the treemaps, the timeline and peaking markers, and the quadrant. None of the other widgets affect the display of the navigation widget.

**Changes to SOLR:** We changed most of the queries which were returning all the documents for a particular query. Instead we returned only those fields which were relevant for the rest of the widgets. No changes were made to the SOLR index for this widget.

## 4.2 Trends and Timeline Widget

The dashboard is provided with a Timeline widget, which displays the trends of the most frequent markers on a timeline. When the dashboard loads the first time, the five most frequent markers are shown by default. The trends are shown using Google Annotated timeline<sup>4</sup>.

The 790 keywords (markers) mentioned in 3.3, are categorized into the following groups (Figure 4.3):

---

<sup>4</sup><http://code.google.com/apis/chart/interactive/docs/gallery/annotatedtimeline.html>

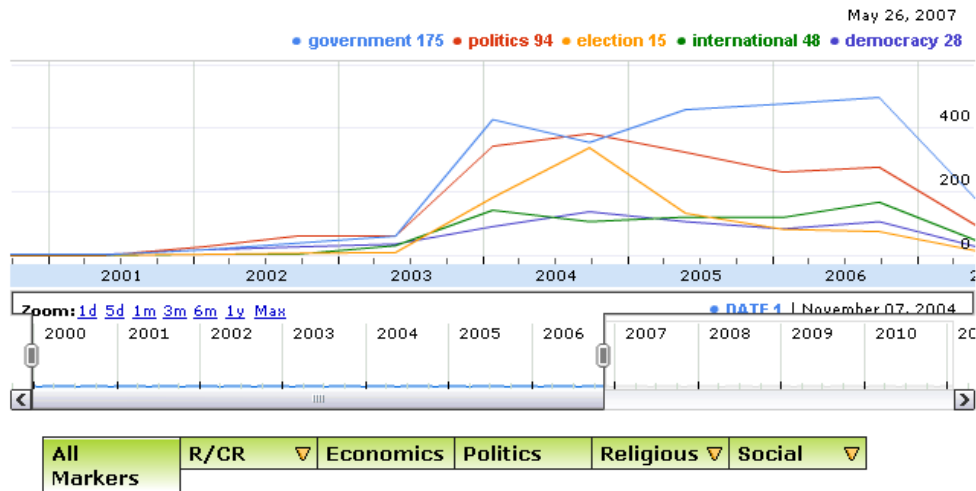


Figure 4.2: Political markers for the period between 2000-2006

- R/CR
- Economics
- Politics
- Religious
- Social

Some of these categories were further divided into sub-categories. These categories were manually tagged by social scientists. We wanted to utilize this feature so that the user could look at the trends for a group of markers. It is especially helpful when a user wants to see a particular type of marker (example "politics" marker like "democracy"). Showing the top 5 from the 790 markers might not show any political markers. This is where the markers menu is useful. Figure 4.2 shows the top 5 political markers between 2000-2006 and how they fared during each year.

When the visualization dashboard loads for the first time, no category is selected (which means all 790 markers are considered). The time range of the result is split into ten equal parts and a query is executed which extracts the top 5

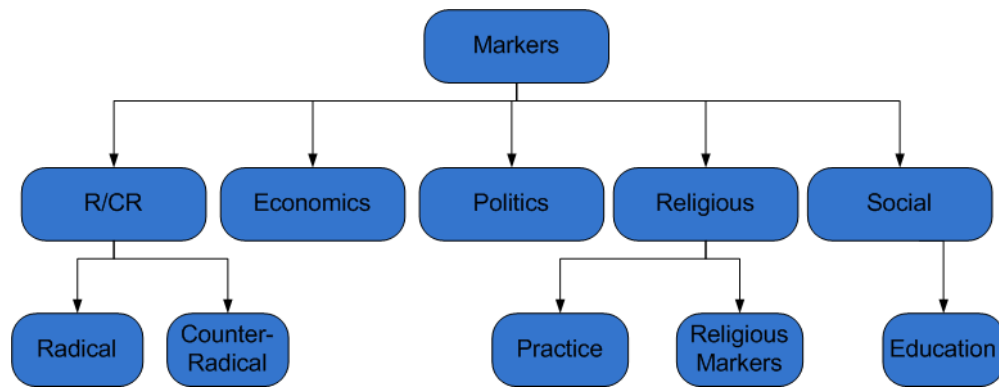


Figure 4.3: Markers Categories

markers (among 790 markers) for each time range based on its frequency. These frequencies are then plotted on the annotated timeline to show the trends. When a sub category of markers is selected, only the trends timeline changes. Ten queries are again executed, but this time they are executed for the particular category of markers.

**Changes to SOLR:** A lot of new fields were added to SOLR schema file to accommodate this information. A snippet of the XML document containing the various categories is shown in Figure 4.4.

A python program was written to change the XML files and add these new fields. The markers were present in the SOLR index in the PositiveSentiment and NegativeSentiment fields (Figure 4.5). Instead of searching for the keywords in the text again, the python program would go through each XML document and select the markers from PositiveSentiment and NegativeSentiment. Before extracting each keyword, the program would check if it has already been extracted (to prevent duplicates).

A CSV file containing the keywords and the corresponding categories was also read by the python program. The extracted keywords were then run through this CSV file and based on the categories new fields were created in the XML file and the keywords were placed in this new field.



```

<!-- Additional fields for markers -->
<field name="allmarkers" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="rmarkers" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="crmarkers" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="rcrmarkers" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="economics" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="politics" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="practice" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="relmarkers" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="belief" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="issues" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="religiousall" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="education" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="social" type="string" indexed="true" stored="true" multiValued="true"/>

```

Figure 4.4: Schema file various categories of markers

```

<field name="positiveSentiment" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="negativeSentiment" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="positiveSentimentENG" type="string" indexed="true" stored="true" multiValued="true"/>
<field name="negativeSentimentENG" type="string" indexed="true" stored="true" multiValued="true"/>

```

Figure 4.5: Schema file showing positiveSentiment and negativeSentiment field

**Control:** The peaking markers is controlled by the navigation widget, the timeline widget and the markers menu.

**Selecting a Time Range:** The timeline widget also enables the user to select various time ranges. This is especially useful when a user/analyst wants to observe the movement of an organization over a period of time. In section 5.2, we look at FPI, a known radical organization [5][12] and the increase/decrease in it's radical discourse over a period of ten years. To play this scenario an analyst can select various time ranges like 2000-2002, 2002-2006 and so on. The timeline widget re-

stricts the results from the Navigation widget. When the dashboard loads for the first time, the timeline widget queries SOLR, and gets the oldest and most recent article dates. These two dates are used to create the Minimum and Maximum range of the timeline. When a user selects a particular range, SOLR is queried again, using the previous SOLR query and the new time range given by the user. For example, if the user selects the facets Organization = "HizbutTahrir", the timeline re-draws itself and now the start and end date in the timeline range is 2007 to 2010. These new dates are extracted from the set of results returned by the query:

Listing 4.3: Query using facets - "source: HizbutTahrir"

```
http://xx.xx.xx.xx:xxxx/solr/select?
facet.enum.cache.minDf=35&rows=5&q=*%3A*&facet=true&
facet.field=demographicsENG&facet.limit=35&
facet.mincount=1&f.source.facet.limit=35&
json.nl=map&fq=source%3AHizbutTahrir&wt=json
```

Now if the user selects a time range of 2008-2009, the same query above is sent to SOLR, with an added parameter of start date = 2008 and end date = 2009.

Listing 4.4: Same query as in listing 4.3 but with start and end date selected by user.

```
http://xx.xx.xx.xx:xxxx/solr/select?q=source:HizbutTahrir
event_dt%3A%5B2008-10-14T00%3A00%3A00Z%20TO%202009-11-7
T00%3A00%3A00Z%5D&facet=true&facet.limit=1000&
facet.mincount=1&rows=0&facet.field=demographicsENG
```

**Control:** The quadrant widget, demographics and markers treemap and the peak-ing markers widget are affected by changing the timeline. The results tab in the

search widget is not changed.

### 4.3 Quadrant Widget

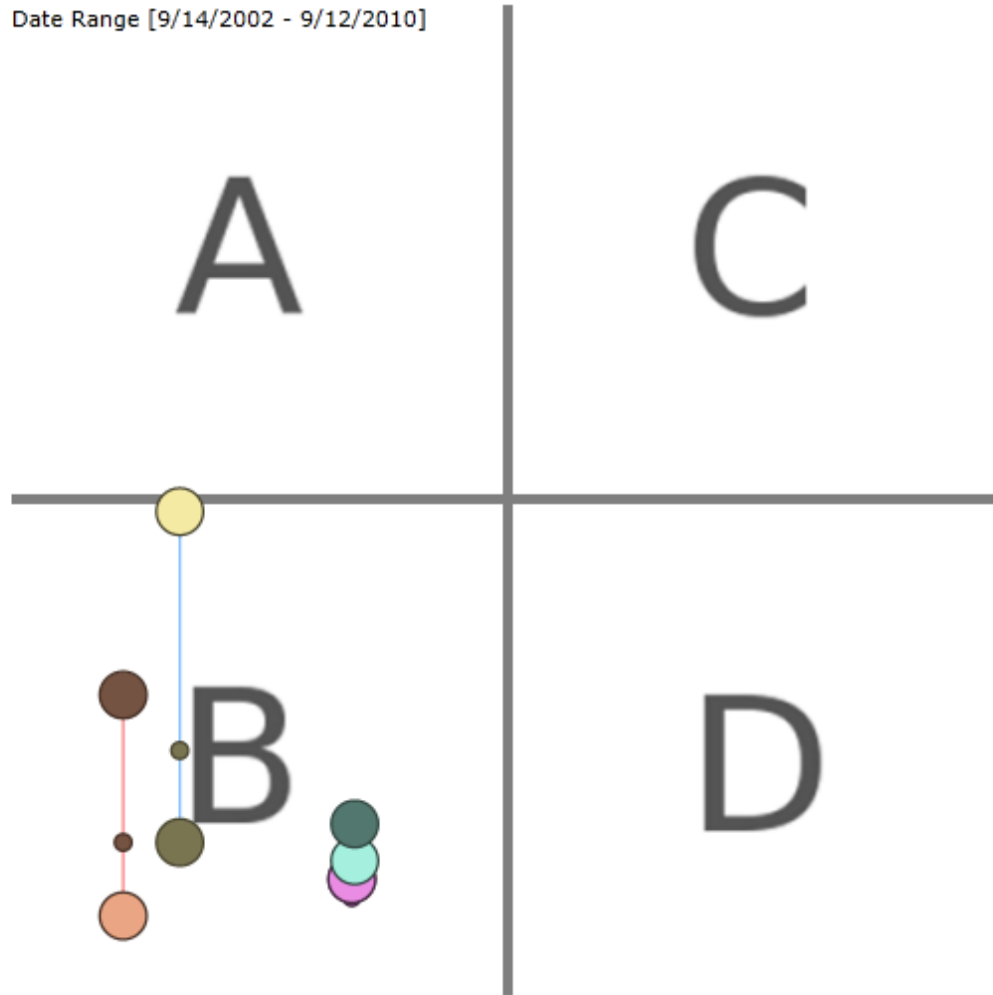
This widget plots the trajectory of all the organizations on the two dimensional scale. The number of organizations and their trajectory depends on the results of the user query. The radical/counter-radical axis (R/CR) is calculated in real time using the subset of organizations and the time range of the current selection. The trajectory of an organization on the quadrant is shown as a color coded path. The beginning of the trajectory is shown by a light colored circle, the mid point is shown by a smaller circle, and the end point is shown as a dark colored circle. A red line between the circle denotes a rise in radicalism for that particular organization and a blue line denotes the opposite. The smaller circle is useful to visualize the overall movement of the organization.

For example, an organization might show an increase in radical activities and then a rapid decrease. The smaller circle helps in capturing this overlap on the quadrant. Figure 4.6 shows the increase and decrease in radicalism of 3 organizations. The Violence/Non-violence (V/NV) axis is retrieved from expert opinions. In the future this scale will also be calculated dynamically.

**Changes to SOLR:** As explained in section 3.3, a set of 26 ad 29 radical and non radical markers were selected. The markers were used to build response tables. These response tables were built using the frequency of occurrence of these markers in each organization's corpus. To accommodate this, we changed SOLR index and added new fields to the schema called "R" and "CR" as shown in Figure 4.7 (we also added the English versions of these called "RENG" and "CRENG"). The python program mentioned in section 4.2 was used for indexing these new fields.

**Control:** The quadrant widget is controlled by the navigation widget and the timeline widget.

Date Range [9/14/2002 - 9/12/2010]



**Quadrant** **Heat Map**

Figure 4.6: Increase and decrease in radicalism denoted by red and blue lines in quadrant

```
<field name="CR" type="string" indexed="true" stored="true"/>  
  <field name="CRENG" type="string" indexed="true" stored="true"/>  
  <field name="R" type="string" indexed="true" stored="true"/>  
  <field name="RENG" type="string" indexed="true" stored="true"/>
```

Figure 4.7: Radical and Counter Radical fields in SOLR schema



Figure 4.8: Tag clouds in Navigation widget

#### 4.4 Demographics and Markers Treemap

The filter tab of the search widget uses tag cloud as shown in Figure 4.8. Each tag cloud represent a specific value of an attribute. For example, the tag cloud “Indonesia” represents one of the values of the attribute “Location”. Each tag cloud serves two purposes:

- Selecting the tag cloud would narrow down the results by selecting only those documents which have the value of the tag cloud. For example, selecting the demographics:“Cleric” (circled in black) will change the query and retrieve a subset of the only those articles which have “Cleric” as demographics.
- The tag cloud also gives a visual representation of the number of times the particular value occurs. Continuing the above example of demographics:“Cleric”, we can see that the number of times “Cleric” occurs in documents is larger than

most of the other values such as “Believers” or “Christian”.

Although tag clouds are very informative, we needed a clearer visualization for the markers and demographics. A treemap is a visualization of hierarchical structures in a limited space<sup>5</sup>. It is very useful in visualizing and comparing various values of a particular attribute. Unlike tag clouds the treemap takes only a limited amount of space irrespective of the number of values. For the visualization dashboard we decided to use Google Treemaps<sup>6</sup> which is also a part of the Google Visualization library<sup>7</sup>. There are two treemaps, **Markers** and **Demographics**. They can be selected using the corresponding tabs below the treemap as shown in Figure 4.9. Both treemaps show the top 10% of all the markers present in the result set.

**Changes to SOLR:** This widget did not need any additional changes in SOLR. It used the new fields added in section 4.2.

**Control:** This demographics treemap is controlled by Navigation widget and the timeline widget. The markers treemap is controlled by navigation widget, the timeline widget and the markers menu.

#### 4.5 Heat Map

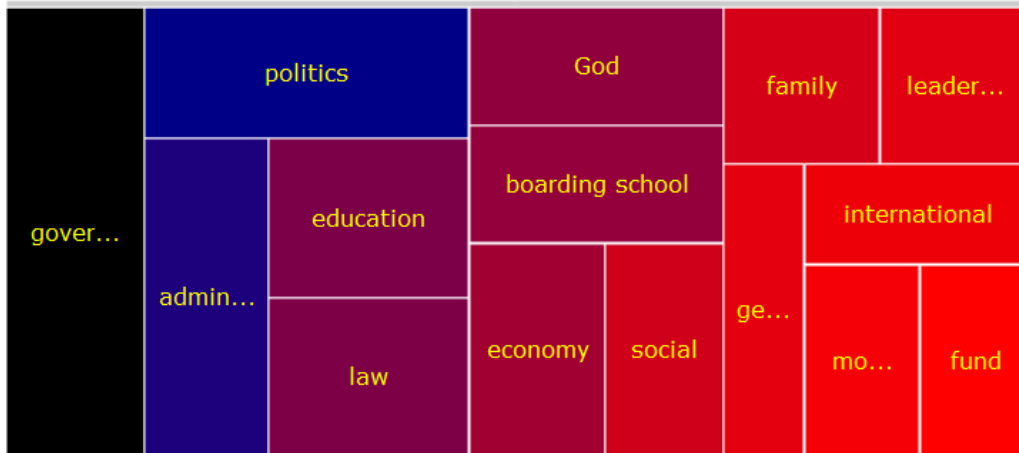
As mentioned in section 3.1 during pre-processing of data, locations were also extracted, using NER. This was very useful in searching and narrowing down to articles which mentioned a particular location. For example, an analyst might want to see the discourse of an organization (example: NU) for a particular country (example: United States). We saw a need to visualize the intensity of an organization's actions in various locations. A heat map, which is essentially a density map, su-

---

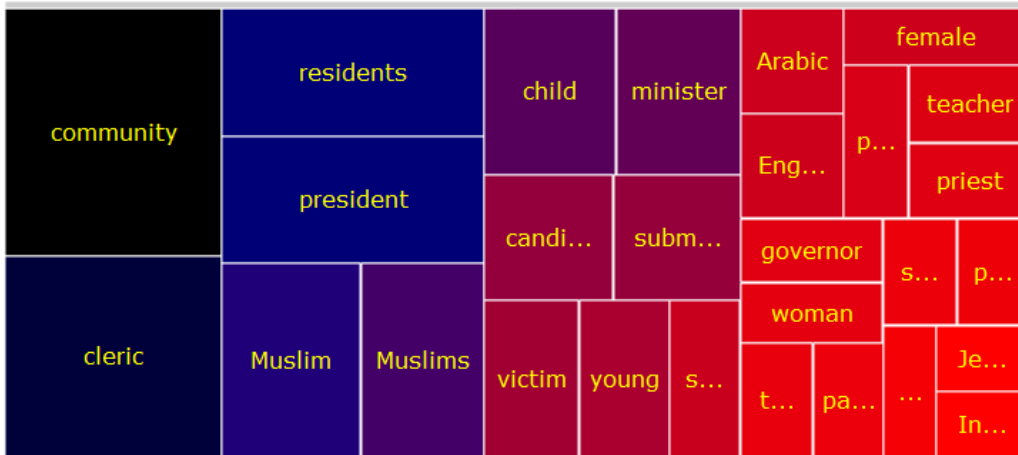
<sup>5</sup><http://www.cs.umd.edu/hcil/treemap/>

<sup>6</sup><http://code.google.com/apis/chart/interactive/docs/gallery/treemap.html>

<sup>7</sup><http://code.google.com/apis/chart/interactive/docs/gallery.html>



Markers Demographics



Markers Demographics

Figure 4.9: Demographics and Markers Treemap

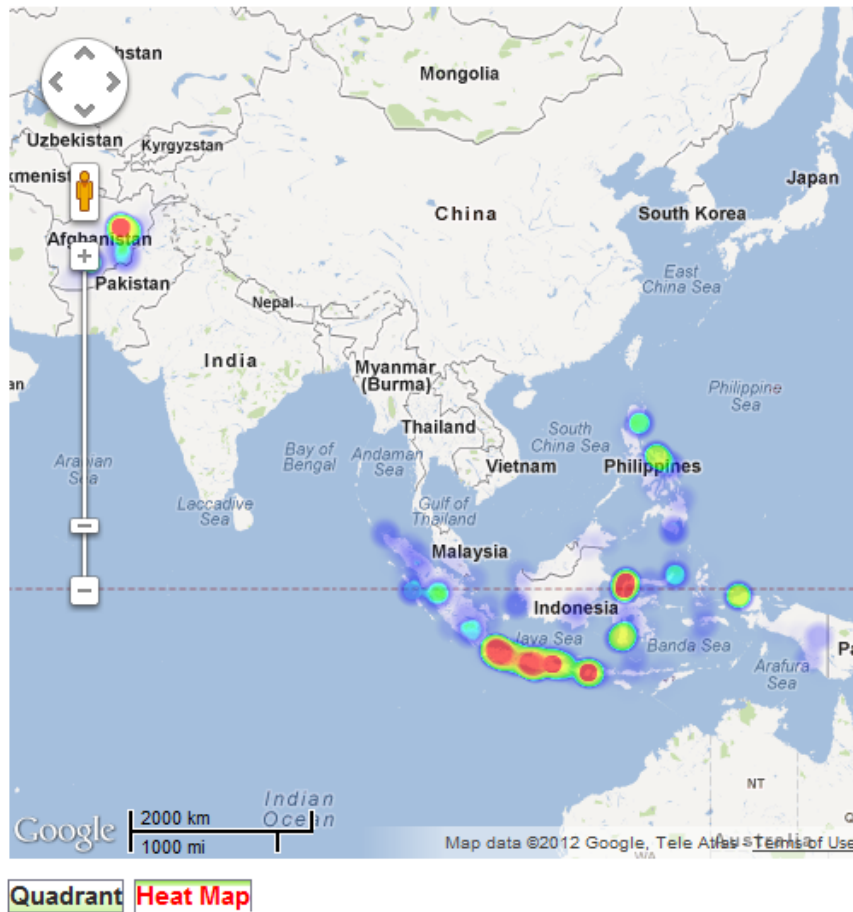


Figure 4.10: Heat Map - shows the intensity map based on the number of times a location is mentioned in the corpus

perimposed on top of a geographical map, provides an excellent visualization of locations on a map. Using the location information we had, a heat map was generated, which showed the number of times a location was mentioned in the result set. As shown in Figure 4.10 we can see that most of the intensity is centered around South East Asia (Indonesia) and Afghanistan. This is expected since our data belongs to Indonesian organizations.

The heat map widget was implemented using a JavaScript heat map library [2] which uses Google Maps<sup>8</sup>. Using google maps, gave us the flexibility to exclude any map related processing or installation on the server end. The heat

<sup>8</sup><http://maps.google.com/>



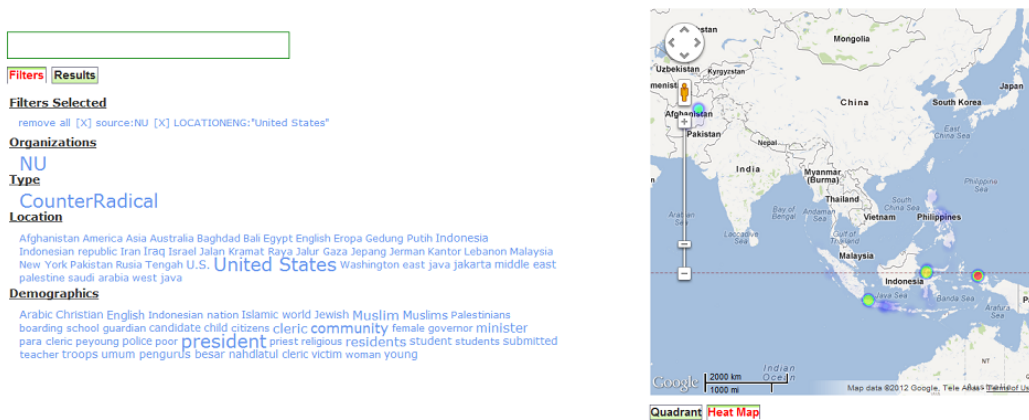


Figure 4.11: Heat map showing the locations mentioned for organization NU and location USA

map works by showing the intensity of a particular location using various colors. The color red is used to show high density. A screenshot of the map, without any user search parameters is shown in Figure 4.10 The data is sent to heapmap.js from the query result. Every change in timeline or search widget retrieves location data also and this is in turn reflected on the heat map.

**Changes to Solr:** To accommodate this feature, geo-location data (Longitude/Latitude) was added, for each of the document indexed in SOLR. We obtained the longitude/latitude data from [1] which contains the longitude and latitude information for all the locations in the word. This data was first indexed into a separate SOLR index (unrelated to the main SOLR index used for this dashboard). We selected SOLR instead of SQL due to its ease of setup and fast retrieval time for text queries. A python program reads the tab-separated input file. For each location, the corresponding location name, country and longitude/latitude are indexed as one document. Figure 4.12 shows the schema entries for these fields. The python program mentioned in section 4.2 reads each XML file, extracts the locations from each file, looks up the corresponding longitude/latitude from the geo-location SOLR index, and adds this as a new field in the XML files. Figure 4.12 shows these new

```
<field name="id" type="string" indexed="true" stored="true" required="true" />
<field name="region" type="string" indexed="true" stored="true" required="true" />
<field name="country" type="string" indexed="true" stored="true" required="true" />
<field name="fullname" type="string" indexed="true" stored="true" required="true" />
<field name="longitude" type="string" indexed="true" stored="true" required="true" />
<field name="latitude" type="string" indexed="true" stored="true" required="true" />
```

Figure 4.12: Schema file showing entries for location SOLR database

entries.

**Control:** This widget is controlled by navigation and timeline widget.

## Chapter 5

### SCENARIOS AND FINDINGS

#### 5.1 Radical Organizations' Trends

In this scenario we analyze both violent and non-violent radical organizations. Our web application shows the ideologies that these organizations are propagating. We can see<sup>1</sup> the most prominent markers associated with these radical organizations. Markers such as "infidel", "sharia" and "violence" show an increasing trend between 2001 and 2010. A very strict interpretation of "Sharia" is used by radical organizations to justify their actions [17][6]. "Sharia" peaks during this period as shown in Figure 5.1.

#### 5.2 C-Quadrant Organizations' Trends

We now analyze Front Pembela Islam (FPI), an Islamic organization in Indonesia established in 1998. FPI is well known for its violent acts [5][12] justified by a strict interpretation of Sharia [3]. Our documents for FPI ranges between 2000 - 2010.

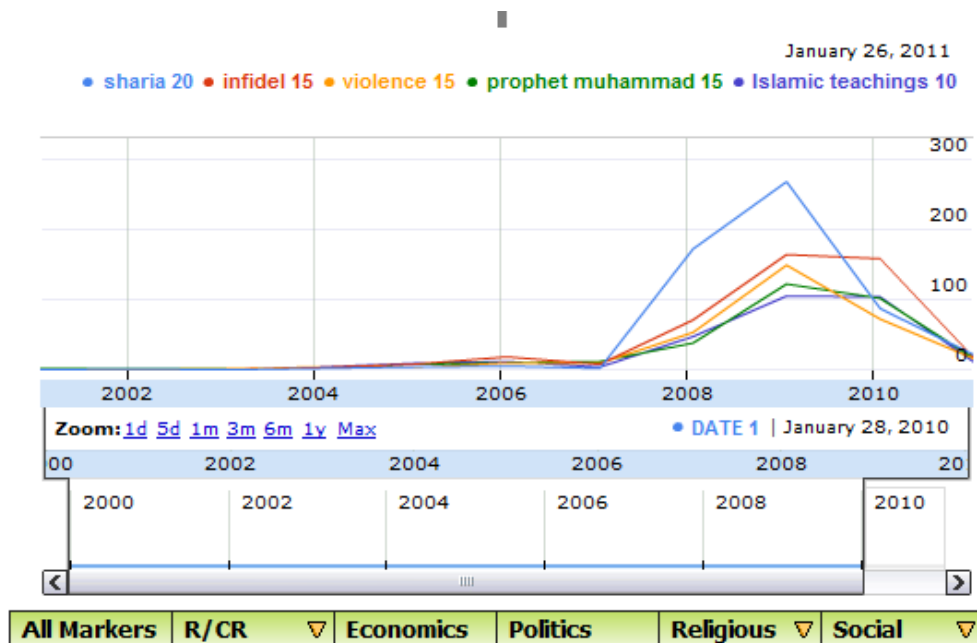


Figure 5.1: Trend of radical markers

<sup>1</sup>Select the filter "Radical" from the search options and then in the Markers Menu select [Religious → Radical Markers]

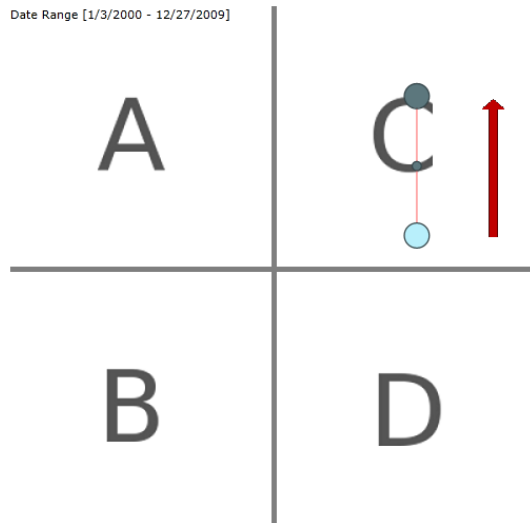


Figure 5.2: Consistent rise of FPI on the radical scale



Figure 5.3: "Ahmadiyya" peaking during the period 2006 - 2010

Using our web application's plots of the movement of FPI in the C Quadrant, we found that FPI consistently rised higher on the radical scale as shown in Figure 5.2. We selected the following time ranges, **2000 - 2003**, **2002 - 2006**, **2006 - 2010** and analyzed the trends of various markers associated with FPI. There was a substantial increase in the intensity of various

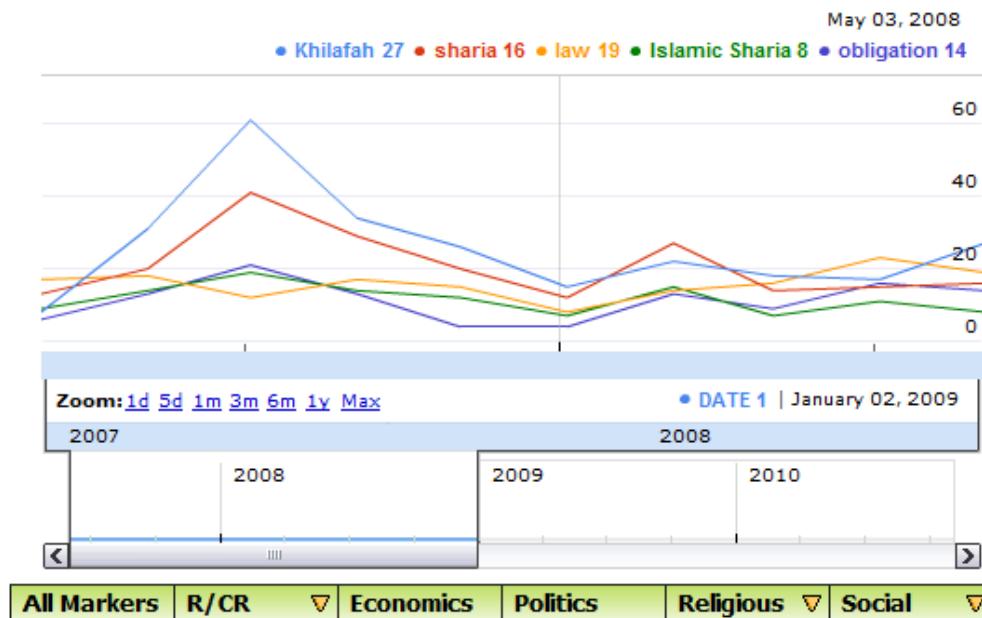


Figure 5.4: “Khilafah” ideology of Hizb ut- Tahrir

radical markers such as “infidel”, “Mujahedin”, “pornography”.<sup>2</sup> Since 2006, we also saw a steep increase in the frequency of marker “Ahmadiyya”, as shown in Figure 5.3, which indicates FPI’s increased opposition to this heretical sect [11].

### 5.3 A-Quadrant Organizations’ Trends

We analyze Hizb ut-Tahrir also known as HTI(Hizb ut-Tahrir Indonesia), a radical organization widely believed to be non-violent [16], which has been active in Indonesia since 1982 [10]. Between 2007 - 2009, our web application shows various radical and non-radical markers associated with this organization.

Radical	Non-Radical
“Sharia”, “Prophet Muhammad”, “Infidel”, “Caliph”, “Violence”	“Politics”, “Indonesian Islam”, “Election”, “Liberal”, “Democracy”

<sup>2</sup>Select “Radical” and “FPI” from the filters, then select the time range 2002-2006 or 2006-2010, then select “radical” markers under “R/CR” menu.

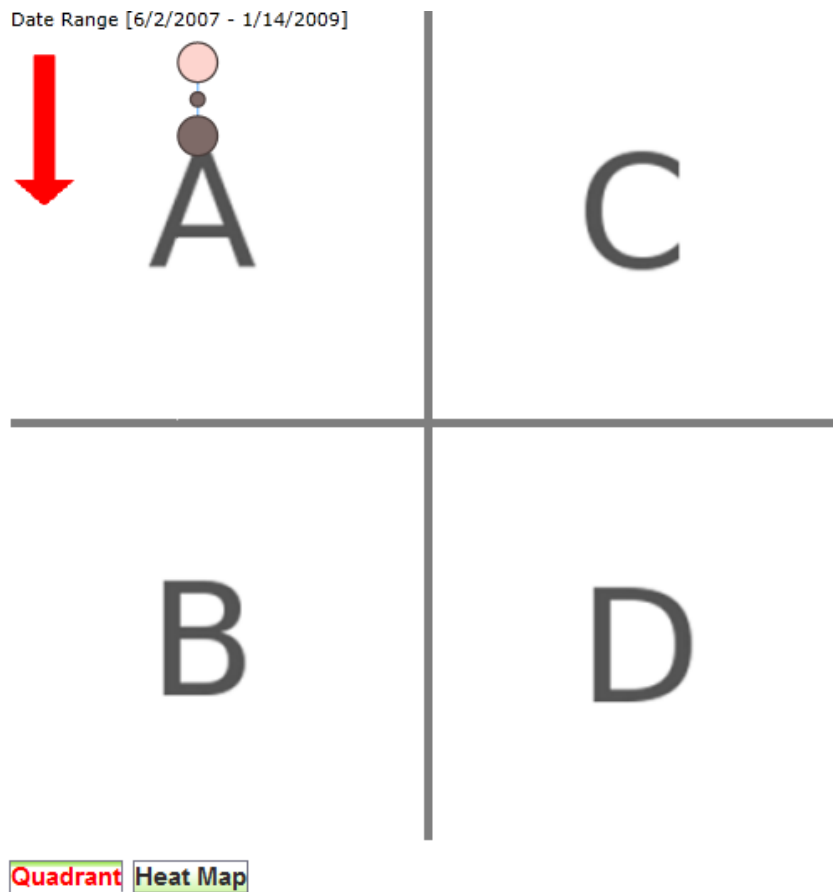


Figure 5.5: Decline of HTI on radical scale

During the same period we see a steady increase in the frequency of the radical marker "Sharia". This is consistent with one of HTI's goals of implementing Sharia in Indonesia [6]. Hizb ut-Tahrir openly propagates the ideology of **Khilafah**, which believes in unification of all Muslim countries as a single Islamic State [18][9]. Figure 5.4 shows "Khilafah" as the most prominent marker<sup>3</sup> in Hizb ut-Tahrir's discourse. By looking at the Quadrants widget we can infer that, HTI has been moderating its narrative as shown in Figure 5.5.

#### 5.4 B-Quadrant Organizations' Trends

In this scenario we discuss the trends of Counter Radical organizations like NU and DaarulUluum. We also show an interesting scenario on the topic of "Suicide

<sup>3</sup>Select "Hizb ut-Tahrir" and "radical" from filters. Select the time range 2007-2009. The markers can be seen by selecting the options of Markers Menu [Religious -> Religious Markers]

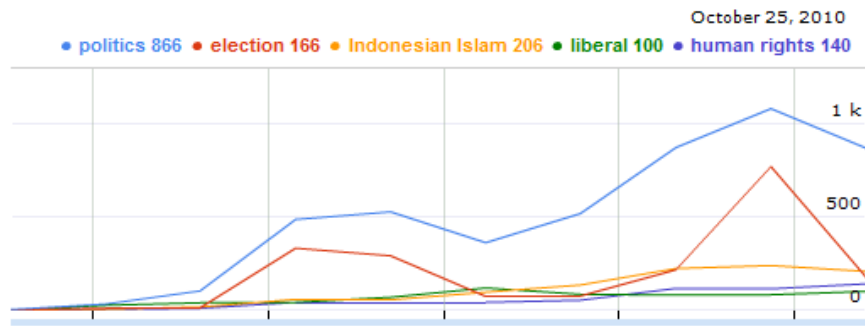


Figure 5.6: Counter Radical markers associated with CounterRadical organizations

Bombing" using the keyword based Navigation widget.

The "counter radical" markers<sup>4</sup> associated with these organizations are: "politics", "election", "Indonesian Islam", "liberal", "human rights". These organizations support democracy and elections, which is shown by the high frequency of the markers "politics" and "election". Their narrative has local interpretation of Islam at its core, which is shown by the marker "Indonesian Islam" (Figure 5.6).

On analyzing the occurrences of radical markers<sup>5</sup> in B-Quadrant, we find that Counter Radical organizations are very vocal against all of radical markers (Figure 5.7). One of the interesting radical markers is "Suicide Bombing". Most of the Counter Radical organizations are against suicide bombings [8]. We will now demonstrate how combination of parametric and keyword search, and various widgets in the web application can help reveal opposition to "Suicide Bombing" by counter-radical organizations.

Searching for the text "suicide bombing", we see that one of the related markers is "ideology". Adding the keyword "ideology" to the search filter reveals a new set of markers including the "sin" keyword. Adding "sin" to our search, we obtain a set of matching documents. One of the top matches, is titled "Mengapa Saya

<sup>4</sup>Select CounterRadical filter in the search option, then from the Markers Menu select [R/Cr → Counter Radical]

<sup>5</sup>In the Markers Menu select [R/CR → Radical]

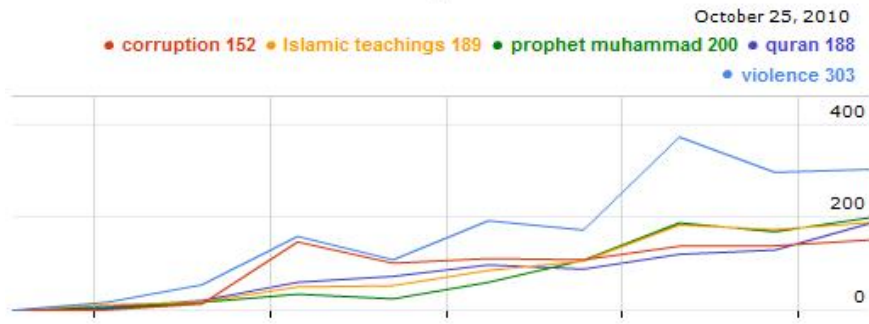


Figure 5.7: Radical markers associated with CounterRadical organizations

Berubah?" (English translation: "Why I changed?")<sup>6</sup>. This article is by a reformed terrorist, debunking the misinterpretation of the jihad related verses used by violent groups.

<sup>6</sup><http://islamlib.com/id/artikel/mengapa-saya-berubah/>



## Chapter 6

### Summary

The Visualization Dashboard provides theoretical insights into the complexities arising from the dynamic interactions of divergent and evolving Islamic radicalisms and counter-radicalisms by capturing their overlap, movement and interactivity. Based on a portable multi-disciplinary methodology, our project transcends parochial and binary frames about friends and enemies, and it provides a new set of social and computational tools for mapping the trajectories of social movements, and emerging threats and opportunities.

Knowing the centers of radical and moderate discourses and their constituencies will enhance the effectiveness of US defense and security policies and programs by allowing a targeted and focused approach. The findings will enhance irregular warfare capability with respect to radical Islam by documenting the landscape of counter-radical discourse. The project will provide information which will help the DoD determine whether a particular group qualifies as “moderate” or how to identify the moderates.

## Chapter 7

### Future Work

- (1) **Calculating V/NV scale dynamically:** Currently this scale is created from inputs given by social scientists and is fixed. In future, it can be calculated dynamically like the R/CR scale. Work is in progress to get the relevant data for this scale.
- (2) **Scalable Document database:** In the future, we are expecting a lot of data coming in from various sources, like government reports on organizations/countries, news feeds and blogs/official websites of organizations. SOLR can handle millions of records without affecting the retrieval time or performance. To make SOLR scalable, the index can be split into shards<sup>1</sup>.
- (3) **Pre-processing using SOLR:** Currently most of the pre-processing on the text is done by a set of scripts. Some of these pre-processing functions, like stemming are already available in SOLR<sup>2</sup> and can be utilized during the indexing. This will reduce the time taken for pre-processing.
- (4) **Hierarchies in Demographics and Markers treemaps:** A treemap is used to represent hierarchical entities. In the future the demographics can be further split based on which country it belongs to. For example, once we have the data we can see which country had more number of “students” associated with a particular organization. Similarly we anticipate an increase in the markers and their categories. To make sure that important markers in a particular category are not missed, we can nest markers inside categories.
- (5) **Improving the timeline range:** Though the timeline tool is very intuitive, there might be occasions when a user wants to select a particular date, say for exam-

---

<sup>1</sup>More information on distributed search using SOLR - <http://wiki.apache.org/solr/DistributedSearch>

<sup>2</sup>These functions are known as Analyzers, Tokenizers and Token filters - <http://wiki.apache.org/solr/AnalyzersTokenizersTokenFilters>

ple, the day elections are held in a country. A user might want to give an exact start and end date. This can be slightly cumbersome in the current timeline. Start and end date fields can be added and a calendar can be presented to the user to select dates from it.

## REFERENCES

- [1] Complete Files of Geographic Names for Geopolitical Areas from GNS. <http://earth-info.nga.mil/gns/html/namefiles.htm>.
- [2] Heatmap.js - JavaScript Library for HTML5 canvas based heatmaps. <http://www.patrick-wied.at/static/heatmapjs/>.
- [3] Terrorist Organization Profile: Front for Defenders of Islam. [http://www.start.umd.edu/start/data\\_collections/tops/terrorist\\_organization\\_profile.asp?id=4026](http://www.start.umd.edu/start/data_collections/tops/terrorist_organization_profile.asp?id=4026). [Online; accessed 21-November-2011].
- [4] Hasan Davulcu, Syed Toufeeq Ahmed, Sedat Gokalp, M'hamed H. Temkit, Tom Taylor, Mark Woodward, and Ali Amin. Analyzing sentiment markers describing radical and counter-radical elements in online news. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing, SOCIALCOM '10*, pages 335–340. IEEE Computer Society, 2010.
- [5] Frank Frost, Ann Rann, and Andrew Chin. Terrorism in southeast asia. <http://www.aph.gov.au/library/intguide/FAD/sea.htm>, November 2010. [Online; accessed 21-November-2011].
- [6] N. Hasan. Islamic militancy, sharia, and democratic consolidation in post-suharto indonesia. 2009.
- [7] I.T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics, 2002.
- [8] Malang. NU chairman deplores suicide bombing attempt. [http://www.nu.or.id/page/en/dinamic\\_detil/15/28282/News/NU\\_chairman\\_](http://www.nu.or.id/page/en/dinamic_detil/15/28282/News/NU_chairman_)

- deplores\_suicide\_bombing\_attempt.html, November 2006. [Online; accessed 22-November-2011].
- [9] Mohamed Nawab Mohamed Osman. Reviving the caliphate in the nusantara: Hizbut tahrir indonesia's mobilization strategy and its impact in indonesia. *Terrorism and Political Violence*, 22(4):601–622, 2010.
- [10] Mohamed Nawab Mohamed Osman. Preparing for the caliphate. *The Asian Studies Association of Australia's E-Bulletin*, pages 14–16, SEP 2011.
- [11] Rahmat and Markus Sihaloho. FPI vows to disband ahmadiyah 'whatever it takes'. <http://www.thejakartaglobe.com/home/fpi-vows-to-disband-ahmadiyah-whatever-it-takes/423477>, February 2011. [Online; accessed 21-November-2011].
- [12] Olivia Rondonuwu and Sunanda Creagh. Opposition grows to indonesia's hardline fpi islamists. <http://in.reuters.com/article/2010/06/30/idINIndia-49777620100630>, June 2010. [Online; accessed 21-November-2011].
- [13] Srinivasa R. Sanaka. Faceted search and browsing of indonesian text collection using shallow parsing techniques, 2010. 1483356; 52231251; 9781124359663; 2222678801; Sanaka, Srinivasa Raviteja; Copyright ProQuest, UMI Dissertations Publishing 2010; 2010; 821237901; 66569; English; M3: 1483356; M1: M.S.
- [14] Sukru Tikves, Sujogya Banerjee, Hamy Temkit, Sedat Gokalp, Hasan Davulcu, Arunaba Sen, Steven Corman, Mark Woodward, Shreejay Nair, Inayah

Rohmaniyah, and Ali Amin. A System for Ranking Organizations using Social Scale Analysis. submitted to journal., January 2012.

- [15] Sukru Tikves, Sujogya Banerjee, Hamy Temkit, Sedat Gokalp, Hasan Davulcu, Arunabha Sen, Steven Corman, Mark Woodward, Inayah Rochmaniyah, and Ali Amin. A system for ranking organizations using social scale analysis. In *EISIC*, pages 308–313. IEEE, 2011.
- [16] Ken Ward. Non-violent extremists? Hizbut Tahrir Indonesia. *Australian Journal of International Affairs*, 63(2):149–164, 2009.
- [17] Hasyim Widhiarto. Radical groups urge Bekasi administration to implement sharia law. <http://www.thejakartapost.com/news/2010/06/27/radical-groups-urge-bekasi-administration-implement-sharia-law.html>, June 2010. [Online; accessed 21-November-2011].
- [18] Yamin Zakaria. A Global Caliphate: Reality or Fantasy? <http://usa.mediamonitors.net/content/view/full/91207>, November 2011. [Online; accessed 21-November-2011].