W.I.S.D.Or.

Web Intelligence for Scaling Discourse of Organizations

by

Sukru Tikves

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2016 by the
Graduate Supervisory Committee:

Hasan Davulcu, Chair
Arunabha Sen
Huan Liu
Mark Woodward

ARIZONA STATE UNIVERSITY

August 2016

ABSTRACT

Internet and social media devices created a new public space for debate on political and social topics (Papacharissi 2002; Himelboim 2010). Hotly debated issues span all spheres of human activity; from liberal vs. conservative politics, to radical vs. counter-radical religious debate, to climate change debate in scientific community, to globalization debate in economics, and to nuclear disarmament debate in security. Many prominent 'camps' have emerged within Internet debate rhetoric and practice (Dahlberg, n.d.).

In this research I utilized feature extraction and model fitting techniques to process the rhetoric found in the web sites of 23 Indonesian Islamic religious organizations, later with 26 similar organizations from the United Kingdom to profile their ideology and activity patterns along a hypothesized radical/counter-radical scale, and presented an end-to-end system that is able to help researchers to visualize the data in an interactive fashion on a time line. The subject data of this study is the articles downloaded from the web sites of these organizations dating from 2001 to 2011, and in 2013. I developed algorithms to rank these organizations by assigning them to probable positions on the scale. I showed that the developed Rasch model fits the data using Andersen's LR-test (likelihood ratio). I created a gold standard of the ranking of these organizations through an expertise elicitation tool. Then using my system I computed expert-to-expert agreements, and then presented experimental results comparing the performance of three baseline methods to show that the Rasch model not only outperforms the baseline methods, but it was also the only system that performs at expert-level accuracy.

I developed an end-to-end system that receives list of organizations from experts, mines their web corpus, prepare discourse topic lists with expert support, and then

ranks them on scales with partial expert interaction, and finally presents them on an easy to use web based analytic system.

DEDICATION

Dedicated to my late father, who is no longer with us today.

ACKNOWLEDGMENTS

I would like to thank Dr. Fatih Gelgi, Dr. Syed Toufeeq Ahmed, Dr. Sujogya Baranjee, Dr. Sedat Gokalp, Nyunsu Kum for the support during my student years at ASU and my adviser Dr. Hasan Davulcu for his wisdom and contribution.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter 1

RESEARCH OVERVIEW

1.1 Introduction

Being able to asses information on radical and moderate actors in a geographic area is an important research topic for national security. *Radicalism* is the ideological conviction that it is acceptable and in some cases obligatory to use violence to effect profound political, cultural and religious transformations and change the existing social order fundamentally. Muslim radical movements have complex origins and depend on diverse factors that enable translation of their radical ideology into social, political and religious movements. In (Crelinsten 2002) Crelinsten states that "both violence and terrorism possess a logic and grammar that must be understood if we are to prevent or control them". Therefore, analysis of Muslim radical and counter-radical movements requires attention to the global, national and local social, economic and political contexts in which they are located. Similarly, in the Islamic context, counter-radical discourse takes various different forms; discursive and narrative refutations of extremist claims, symbolic action such as ritual and other religious and cultural practices, and Islamic arguments for pluralism, peaceful relations with non-Muslims, democracy, etc. The most effective counter-radicals are likely to be religiously conservative Muslims. Effective containment and defeat of radicalism depends on our ability to recognize various levels of radicalization, and detection of counter-radical voices.

As our initial work we developed a framework and a measure for impact forecasting of events in a news stream. We proved the viability of our approach using a SVM based forecaster on six months of NYT corpus - consisting of 16,852 articles. We experimented with different feature selection and ranking algorithms including standard frequency based methods, as well as a new method named ImpactRank. Our ImpactRank based forecaster performed as the best feature ranking technique while providing a graph suitable for browsing and identifying the most influential topics, entities and inter-relationships going into its impact predictions.

As a contribution we introduced the ImpactRank network, which is based on TermRank method introduced by Gelgi et. al. (Gelgi, Davulcu, and Vadrevu 2007). This eigen-vector based algorithm provides both a feature selection method for SVM classifier, performing on the best possible level, and also a social relationship network inference tool for use in investigation research.

We worked with social scientists on our team to come up with an orthogonal model comprising of two primary dimensions. Both dimensions, (i) radical/counter-radical and (ii) violent/non-violent, are characterized as latent, partial orders of discrete beliefs and practices based on a generalization of item order in Guttman scaling (Guttman 1950) using a Rasch model (Andrich 1988). A true Guttman scale is a deterministic process, i.e. if a social movement subscribes to a certain belief or practice, than it must also agree with all lower order practices and beliefs on the scale. Of course, such perfect order is rare in the social world. The Rasch model provides a probabilistic framework for Guttman scales to accommodate for incomplete observations and measurement errors.

In our prior work (Tikves et al. 2011) we showed that both counter-radical and radical movements in Muslim societies exhibit distinct combinations of perspectives on various social, political, and religious issues, and those perspectives can be mapped to a latent linear continuum, or a scale. The resulting model allowed us to measure the distance between organizations and movements over the underlying scale. It also facilitates tracking the ways in which movements and organizations change over time and space (Tikves et al., Accepted for publication).

A debate is defined as a formal discussion on a set of related topics in a public meeting, in which opposing arguments are put forward. Initially, we observe that given a certain topic, each organization's web site mostly discusses their own perspectives related to that topic, and occasionally discusses others' perspectives, relating them back to their own perspectives. As a case study of an ongoing large scale online debate, we utilize the discourse found in the web sites of 10 *radical*, and 13 *counter-radical* Indonesian religious organizations - comprising a total of 37,000 articles dating from 2001 to 2011. Radicalism (Woodward et al. 2010) is the ideological conviction that it is acceptable and in some cases obligatory to use violence to effect profound political, cultural and religious transformations and change the existing social order. Counter-radicals oppose violent social and political movements.

Given the complex nature of the task, such as regional differences in local cultures, beliefs and practices, and in the absence of readily available high accuracy parsers, highly structured religio-social ontologies, and information extraction systems; we decided to devise a multi-lingual non-linguistic text processing pipeline that relies on only statistical modeling of keyword frequency and co-occurrence information. However, we designed the system to be able to incorporate additional information extracted from the text, if available. For example, named entity recognition (NER), machine translation, and GIS based location look up information are part of the user interface presentation.

In (Tikves et al. 2011), we utilized a simple term frequency - inverse document frequency (TF-IDF) Hartigan and Wong 1979 based technique to generate a large candidate list of topics and perspectives for inclusion in scaling analysis. Top 100 n-grams from each organization's web site were collected into a list of candidate keywords. Next, we asked social scientists to scan this list manually, and identify all significant keywords belonging to social, political, economic, and religious perspectives. During this process, social scientists on our team assessed a total of 790 candidate keywords; of which 29 and 26 were selected by experts for inclusion in the radical and counter-radical scaling analysis respectively.

Upon analyzing the results of this study, we have identified that automatically generating the items of the radical and counter-radical scales would be an important contribution to the research. For example, among the included scale items were phrases like "religious education". However, reaching that item from a seed topic (like "education"), instead of manual selection would be desirable. This would not only decrease the expert intervention in scale generation, it would also provide us with useful perspective of organizations on these topics aligned with the underlying scale. In order to explore this idea, we have developed methods for perspective analysis built upon previous findings of the scaling research.

Our next contribution was the development of automated perspective discovery techniques which would contribute to the understanding of features (i.e. social, political, cultural, religious beliefs, goals, and practices) shared by one side of a debate, and by those opposing them. Secondly, we show that, our perspective discovery algorithms not only identify larger number of relevant features - compared to the semi-automated process, but also yield a higher accuracy scale of radicalism vs. counter-radicalism.

We've designed a web based system to visualize this orthogonal model. The web tools provided by the system allows drilling down on specific data, and plotting the trends and trajectories of organizations on a timeline. It consists of several modules: an off-line web mining, and data processing pipeline, two web services for application logic, and an AJAX based presentation layer. The web based interface built for this study can be accessed through the web site at http://demo.minerva-project.org[1]. We present several scenarios with this tool in Section 2.5.

[1]This research was supported by US DoDs Minerva Research Initiative Grant N00014-09-1-0815, Project leader: Prof. Mark Woodward, Arizona State University, and the project title is "Finding Allies for the War of Words: Mapping the Diffusion and Influence of Counter-Radical Muslim Discourse".

Chapter 2

SOCIAL SCALING WITH RASCH MODELS

2.1 Problem Definition

The primary goal of this part of the study is to build a semi-automated method to rank religious organizations from a certain geographical region on a range of scales using their web sites' corpora. The efficacy of the generated model is evaluated by comparing it against baseline methods and expert level performance. In addition to accomplishing these goals, we also present an end-to-end system architecture, tools for gathering expert wisdom, an analysis framework for scaling fata, and a graphical user interface design to facilitate faceted search and browsing of this corpus.

2.2 Introduction of Guttman Scaling and Rasch Model

In social science *scaling* is a process of measuring and ordering entities called *subjects* based on their qualitative attributes called *items*. In general, subjects are requested to respond to surveys conducted by means of structured interviews or questionnaires. Items are presented to the subjects in form of questions. Statistical analysis of the response of the subjects on the questions about items are used in *scaling* the subjects. Some of the widely followed scaling procedure in social science surveys are Likert scale (Likert 1932), Thurnstone scale (Thurstone 1928), and Guttman scale (McIver and Carmines 1981). In Likert scale subjects indicate their magnitude of agreement or disagreement about an item (from strongly agree to strongly disagree) on a five to ten point scale. On the other hand Thurnstone scales is a formal method of ordering the attitudes of the subjects towards the items. Guttman scaling procedure orders both the subjects and the items simultaneously with respect to some underlying cumulative continuum. In our research we adopted the Guttman scaling process to rank the organizations based on their response on the radical and counter-radical keywords.

2.2.1 Guttman Scaling

A Guttman (Guttman 1950) scale presents a number of items to which each subject is requested to provide a dichotomous response, e.g. agree/disagree, yes/no, or 1/0. This scaling procedure is based on the premise that the items have strict orders (i.e., the items are presented to the subjects ranked according to the level of the item's difficulty). An item "A" is said to be "more difficult" than an item "B" if any subject answering "yes" on item "A" implies that the subject will also answer "yes" on item "B".

A subject who responds to an item positively is expected to respond positively to all the items of lesser difficulty. For example, in order to find out how extreme a subject's view is on Guttman scale, the subject is presented with the following series of items in question form: (1) Are you willing to permit immigrants to live in your country? (2) Are you willing to permit immigrants to live in your community? (3) Are you willing to permit immigrants to live in your neighborhood? (4) Are you willing to permit immigrants to live to your next door? and (5) Are you willing to permit your child to marry an immigrant? If the items form a Guttman scale, any subject agreeing with any item in this series will also agree with other items of lower rank-order in this series.

Guttman scale is a deterministic process and the score of a subject depends on the number of affirmative responses he has made on the items. So, a score of 2 for a subject in the above Guttman scale not only means he has given affirmative response to two of the questions or items, but also indicates that he agrees with two particular questions, namely the first and second.

Scores in Guttman scale can also be interpreted as the "ability" of a subject in answering questions sorted in increasing order of "difficulty". These scores when presented on an underlying scale, give us an ordering of the subjects based on their "ability" too.

The objective of our ongoing research is to order the Indonesian Islamic organizations based on their views on religio-social keywords which have an inherent ordering. For example, two such keywords are "Qur'an" and "Sharia". An organization supporting "Sharia" will also likely to "believe in Qur'an". So it makes sense to use Guttman scaling procedure to rank the organizations and their beliefs and practices. One drawback of Guttman scale is that it is deterministic and assumes a strict ordering of the items. In real world, it is difficult to order all the items in such a strict level of increasing difficulty, therefore perfect scales are not often observed in practice. Furthermore, many times, the order of the items are not known since they are not straightforwardly comparable. Also measurement errors might lead to responses that do not strictly fit the ordering. As a result we can no longer conclude deterministically that if a subject answers a question affirmative, whether she will be able to give affirmative answers to other questions of lower order in the same questionnaire. We use Rasch model to overcome this drawback by taking into account measurement error.

2.2.2 Rasch Model

Rasch model (Andrich 1988) provides a probabilistic framework for Guttman scales. In Rasch model, the probability of a specified binary response (e.g. a subject agreeing or disagreeing to an item) is modeled as a function of subject's and item's parameters. Specifically, in the simple Rasch model, the probability of a positive response (yes) is modeled as a logistic function of the difference between the subject and item's parameters. Item parameters pertain to the difficulty of items while subject parameters pertain to the ability of subjects who are assessed[2]. A subject of higher ability relative to the difficulty of an item, has higher probability to respond to a question affirmatively. Rasch models are used to assess the organizations degree of being radical or counter-radical based on the religio-social keywords (items) appearing in their rhetoric.

Rasch model also maps the responses of the subjects to the items in binary or dichotomous format, i.e., 1 or 0. Let Bernoulli variable X_{vi} denotes the response of a subject v to the item i, variable θ_v denotes the parameter of "ability" of the subject v and β_i denotes the parameter of "difficulty" of an item i. According to simple Rasch model the probability that subject v responds 1 for item i is given by

$$P(X_{vi} = 1|\theta_v, \beta_i) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)}$$

Rasch model assumes that the data under analysis have the following properties

[2]Ability in our study means the rank in the social scale, while difficulty means the topic weights in ranking

1. *Unidimensionality*: $P(x_{vi} = 1|\theta_v, \beta_i, \alpha) = P(x_{vi} = 1|\theta_v, \beta_i)$, i.e., the response probability does not depend on other variable
2. *Sufficiency*: sum of responses contains all information on ability of a subject, regardless which item it has responded
3. *Conditional independence*: for a fixed subject there is no correlation between any two items
4. *Monotonicity*: response probability increases with higher values of θ, i.e., subject's ability

Items with $s_i = \sum_v^n x_{vi}$ value of 0 or n, and subjects with $r_v = \sum_i^k x_{vi}$ value of 0 or k are removed prior to estimation, where n is the total number of subjects and k is the total number of items. Inferring the Rasch model from the data gives us an Item parameter estimate or a score for each item. Generally the estimation of β_i or score for a item i is calculated through Conditional Maximum Likelihood (CML) estimation (Pawitan 2001). The conditional likelihood function for measuring item parameter estimate is defined as

$$Lc = \prod_v P(x_{vi}|r_v) = \frac{\exp(-\beta_i s_i)}{\prod_r \sum_{x|r} \exp(-\beta_i x_{vi})}$$

where r represents the sum over all combinations of r items. Similarly maximum likelihood is used to calculate subject parameter estimation θ_v or score for each subject. Expectation-maximization algorithms (Hunter and Lange 2004) are used in implementing Conditional Maximum Likelihood (CML) estimation in Rasch model. We can also assess whether the data fits the model by looking at goodness of fit indices, such as the Andersen's LR-test. We used the eRm (Mair and Hatzinger 2007) package to run the Rasch models.

2.2.3 Implementing Rasch Model in the Text Mining Domain

We used Guttman scaling and Rasch model to find a ranking of 23 religious organizations based on extremity of their views are on radicalism and counter-radicalism. In our application, Rasch model *subjects* correspond to a group of religious organizations, and *items* correspond to a set of keywords for socio-cultural, political, religious *radical* and *counter-radical* beliefs, and practices. An organization responding "yes" to a feature means the organization exhibits that feature in its narrative, while an organization responding "no" to a feature indicates that the organization does not exhibit such a feature. *Difficulty* of an item translates to *strength* of the corresponding attitude in defining radical or counter-radical ideology of any organization. Similarly *ability* of a subject in this case means the *degree* of radicalism or counter-radicalism exhibited by an organization's rhetoric. Other works in text-mining domain such as sentiment analysis, have used Rasch model in their analysis (Drehmer, Belohlav, and Coye 2000). Details of keyword extraction and selection are presented in Section 2.3.2.

The fundamental advantage of Rasch model (and hence Guttman scale) in ranking is being able to asses both the subjects (organizations), and the items (keywords/features) with the same model. The Rasch model puts them both in the same latent dimension, allowing comparisons between both organizations and keywords at the same time (for example, having Organization A'a "ability" being higher than keyword K's "difficulty" would mean, it is likely that the Organization A will have keyword K in their corpus). However the absolute values of these scores should be assigned any meaning, since the model can be translated in the same dimension, without affecting any of the probabilities.

2.3 System Architecture

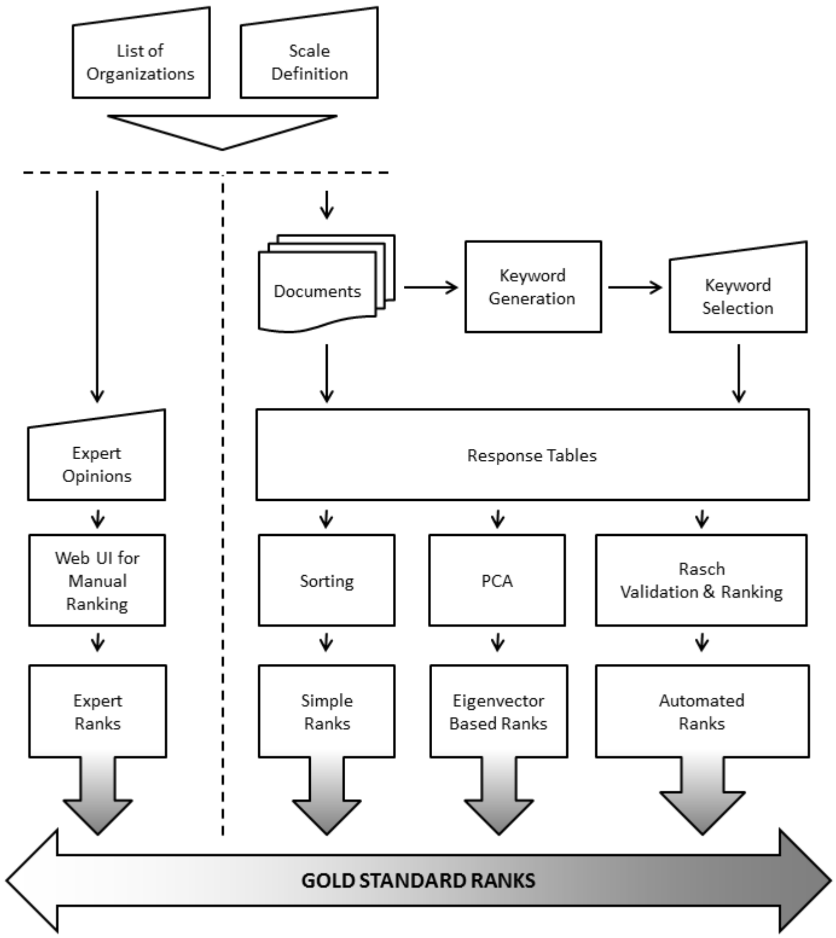

Figure 1. A model of the scale inference architecture.

A summary of the system architecture can be seen in Figure 2. The system is a composition of four components: a data gathering component, which does web crawling, and text extraction; a scale generation component, performing scaling algorithms; application services component, provides data to user interface components; and finally a web user interface, allowing the user to navigate the data.

The process consists of four main steps, where we first gather the text and key-

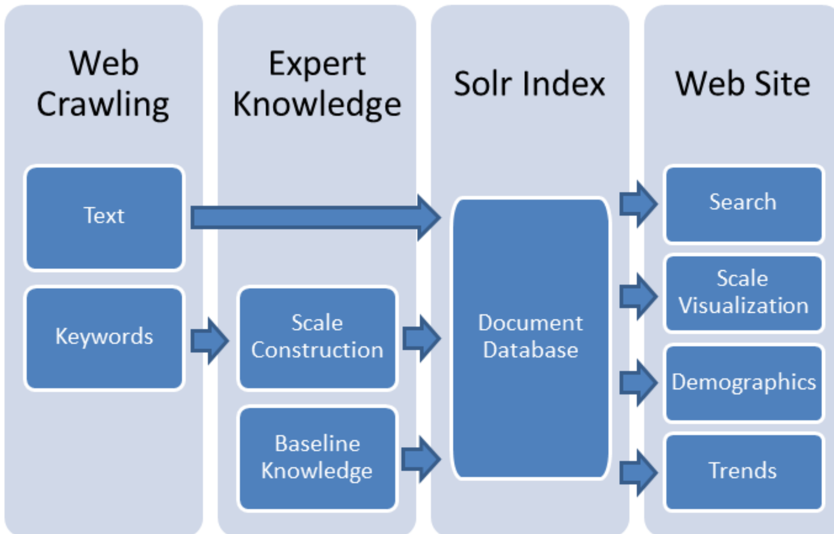

Figure 2. An overview of the system architecture.

word information through crawling the web sites of the organizations focused in the study, then expert knowledge is integrated to construct the scales, and provide a baseline. In the third step, all the documents are stored in a SOLR server instance, which is then presented to researchers in a web application.

2.3.1 Data Gathering

Initially, social scientists are invited to use their domain and area expertise to identify a set of *organizations*, and hypothesize any number of unipolar or bipolar *scales* that could explain the variance among their beliefs and practices. Next, a set of web crawling scripts are created for extraction of articles from those organizations' web sites. For each organization's corpus we extract their top-k n-grams, and a union of all these phrases are presented to experts for feature selection. Downloaded articles are then converted into XML structures, containing their original text, their set

```
<doc>
<field name="source">Muhammadiyah</field><field name="type">CounterRa
<field name="URL">http://www.muhammadiyah.or.id/Berita-Persyarikatan-
<field name="title">33 Rumah Sakit, dan 500 Tenaga Medis, Siap Dukung
<field name="PERSON">Demikian</field>
<field name="PERSON">Ahmad Muttaqin Alim</field>
<field name="ORGANIZATION">Seabad Muhammadiyah</field>
<field name="LOCATION">Mandala Krida</field>
<field name="LOCATIONENG">Mandala Krida</field>
<field name="date">29/06/2010</field>
<field name="event_dt">2010-06-29T00:00:00Z</field><field name="longl
<field name="content">Yogyakarta- Sebanyak 33 rumah sakit Muhammadiya

Demikian disampaikan Ahmad Muttaqin Alim, sekretaris tim kesehatan Mu
```


Figure 3. A portion of a document represented in the system

of keywords, and extracted information such as person, location and organization names using a named entity recognition (NER) tool for Indonesian language, and their machine translations into English.

An example document snippet is shown in Figure 3. Here the original input ($content$, $source$), and a sample of the automatically extracted information corresponding to $DATE$, $PERSON$, and $LOCATION$ can be seen. The corresponding XML versions for each input document are then stored in a document database for processing.

2.3.1.1 Social Scale Generation

Social scale generation is done by building *response tables*; a pair of tables for a bipolar scale, such as radical / counter-radical (R/CR), or a single table for a unipolar scale, by thresholding the occurrence frequencies of the selected keywords in the organizations' web corpus.

The scale generation architecture is shown in Figure 1. Here the flow of the processes and data can be seen as interactions between experts and automated modules. The system works as follows:

- Initially, area experts to identify a set of *organizations*, and hypothesize any number of unipolar or bipolar *scales* that could explain the variance among the beliefs and practices of the organizations.
- Next, we crawl and download the web sites of the organizations, and the system automatically *extracts the top-k candidate keywords* for consideration in the hypothesized scale. Social scientists screen the list of extracted keywords, and *select* the relevant ones for inclusion in further analysis.
- The system builds *response tables*; a pair of tables for a bipolar scale (such as radical / counter-radical R/CR), or a single table for a unipolar scale, by thresholding the occurrence frequencies of the selected keywords in the organizations' web corpus. See Figure 6 and Figure 7 for the response tables for the R/CR scale.
- The response tables are fed as input to the *Rasch Model building* algorithm. The algorithm produces a metric to *validate* the fitness of the model, and *rankings* of the organizations and keywords. Figures 4 and 5 shows the relative positions of the organizations and keywords on the latent scales. The algorithm also produces a metric to *validate* the fitness of the model.
- Two types of other information are collected for *evaluation* purposes. First, *expert rankings* of the organizations, using a graphical drag-and-drop expert opinion elicitation tool shown in Figure 11). Expert rankings are merged into a consensus *gold standard* of rankings. Next, two other computational baseline methods; one based on simple sorting, and another based on principal component analysis (Jolliffe 2002), are used to generate alternative *computational rankings* shown in Figure 12.

Figure 6 shows a sample of the response table for the radical subset of the corpus. A similar table is also built for the counter-radical subset. Figures 6 shows the relative positions of the organizations and keywords on the latent radical scales. A similar scale is also built for the counter-radical subset.

We fit the Rasch model on two datasets - (1) radical organizations with radical keywords and (2) counter-radical organizations with counter-radical keywords. We used the eRm package in R, an open source statistical software package[3], to fit a Rasch model to the dataset, and obtain the organizations' scores on the latent scale, which are the the subject parameter estimates (θ_v) discussed in previous section. The eRm package[4] fits Rasch models and provide subjects or organizations parameter estimates based on maximum likelihood estimation.

The automated scale of the organizations is formed by ranking the organizations according to their estimates on the latent scale. Not only we can provide the organization estimates but we can also assess whether the model fits the data by looking at several goodness of fit indices, such as the Andersen's LR-test.

Additionally the data for the *violence/non-violence* is gathered using a separately developed tool, by collecting the opinion of the experts. A future work will also include automated generation of this dimension, as well.

---

[3]http://cran.r-project.org/

[4]http://r-forge.r-project.org/projects/erm/

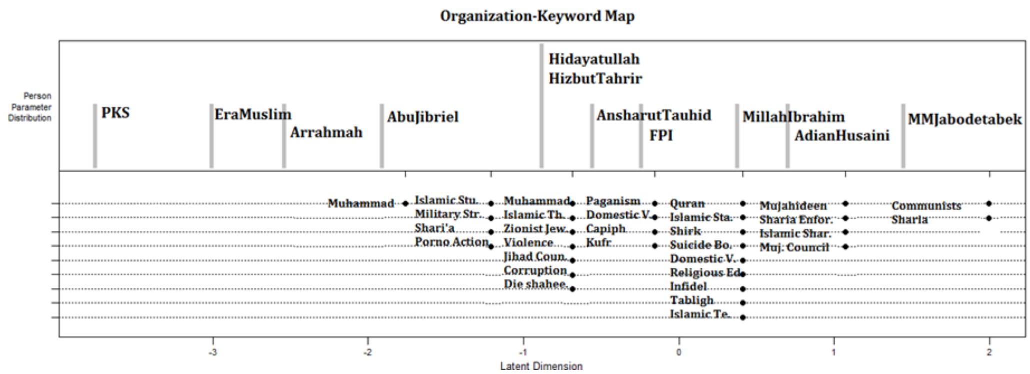

Figure 4. Radical subset of organizations and keywords

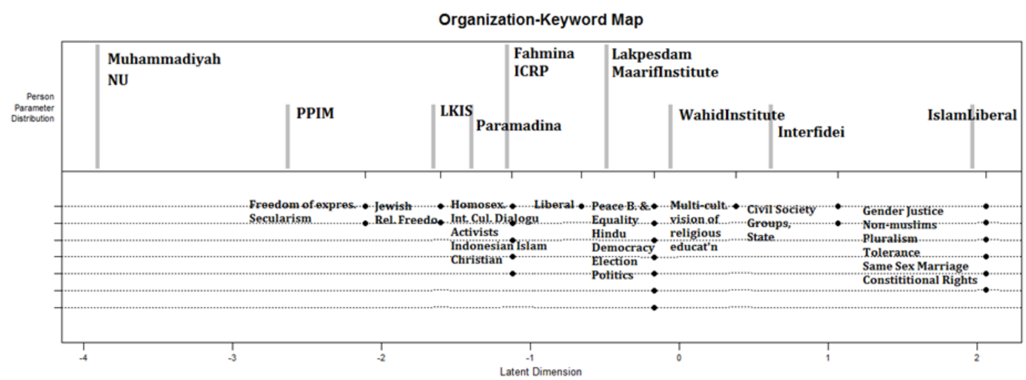

Figure 5. Counter-Radical subset of organizations and keywords

2.3.2 Keyword Extraction and Selection

In order to identify candidate keywords, one option was to translate the documents into English and apply readily available keyword extraction methods (W.B. Michael 2010). However it was preferable to preserve the original expression of the phrases in the original language. Hence, we utilized a non-linguistic technique that relies only on statistical occurrence, and frequency information.

In the initial version of our study, we utilized a non-linguistic technique that relies only on statistical occurrence, and frequency information.

Within each document, the words were separated by white space or punctuation marks. We considered each keyword to be an n-gram of one to three words. We treated each organization as one document and calculated the term frequency - inverse document frequency (TF-IDF) (Salton and Buckley 1988) values for every single n-gram mentioned by these organizations. Top 100 n-grams with highest TF-IDF values from each organization were used to generate a candidate list of topics that these organizations discuss most frequently. Next, we asked our team of experts to screen and manually select identify {social, political, economic, and religious} keywords corresponding to beliefs, goals and practices. During this process, our team of experts screened a total of 790 candidate keywords; and they selected 29 keywords for inclusion in the radical scale, and 26 keywords for inclusion in the counter-radical scale.

2.3.3 Response Table Extraction

Since no direct response is expected from organizations on the topics we are interested in, we utilized web information extraction techniques to build response tables. This would similate responsed to questionarries, by looking at their web corpora.

A threshold value for each keyword is calculated from the values in the related column. And then, each element was converted into a binary value by comparing it to the column's threshold. English translations of the keywords is presented for clarity in Figure 6 and Figure 7.

After identifying the keywords for the analysis, we needed to search the web site corpus of the organizations for the matching items. This yielded a term-document matrix.

This task was performed in a simple three step procedure; initially the occurrence

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	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	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	1	0	0	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	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	0	1	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	0	1	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
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

Figure 6. Radical subset of organizations and keywords, sorted according to aggregate row values.

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	1	0	1	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	0	1	0	0	1	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 7. Counter-Radical subset of organizations and keywords, sorted according to aggregate row values.

frequencies of particular keywords were counted within each organization's corpus, then a threshold matrix was calculated from the initial values, and finally a binary response matrix was generated by applying these thresholds to the initial values.

A response table is calculated based on the normalized frequency with which organizations voice various perspectives in their web sites. The normalized frequencies of perspectives for each organization are calculated by using formula 2.1. In formula 2.1 k is the perspective, o is the organization, and D_o is the entire document set for organization o.

$$f_{o,k} = \frac{|\{d \mid k \in d, d \in D_o\}|}{|D_o|} \tag{2.1}$$

The median frequency of each perspective is selected as a threshold. Organizations' normalized perspective frequencies and the threshold of each perspective are used to build a dichotomous [0/1] response matrix as the organizations' response table.

2.3.4 Model Fitting

We fit the Rasch model on two datasets - (1) radical organizations with radical keywords and (2) counter-radical organizations with counter-radical keywords. We used the eRm package in R, an open source statistical software package[5], to fit a Rasch model to the dataset, and obtain the organizations' scores on the latent scale, which are the the subject parameter estimates (θ_v) discussed in previous section. The eRm package[6] fits Rasch models and provide subjects or organizations parameter estimates based on maximum likelihood estimation.

The automated scale of the organizations is formed by ranking the organizations according to their estimates on the latent scale. Not only we can provide the organization estimates but we can also assess whether the model fits the data by looking at several goodness of fit indices, such as the Andersen's LR-test.

[5]http://cran.r-project.org/

[6]http://r-forge.r-project.org/projects/erm/

2.3.5 Application Services

We use two back end services in the application layer to present the data to the user interface. First, all the extracted textual information is stored in Apache SOLR[7], providing facilities like full text search, and faceting (Tunkelang 2009), using an AJAX interface. Additionally a WCF based scaling service is used to infer scales in real time. This particular service loads the response table, and the previously generated scale data, and estimates the R/CR scale for a subset of the input. Number of positive responses are interpolated on the scale to generate the scale, and the expert opinion is used for a static violent / non-violent (V/NV) scale. While the interpolation is based on a sufficient statistics, future work on speeding up Rasch model generation for real time use would be beneficial.

2.3.6 User Interface

The user interface is implemented as an interactive AJAX based application, using ajaxsolr[8] framework. In addition to the search and navigation capabilities provided with ajaxsolr, it also adds functional widgets for visualizing the organizations on a scale, mapping the intensity of the locations, displaying demographics trends, and so on. A more detailed discussion of the user interface is provided in Section 2.5.

The presentation of the scale, dynamically generated from filtered corpus subset, brought the following challenges:

[7] http://lucene.apache.org/solr/

[8] http://evolvingweb.github.com/ajax-solr/

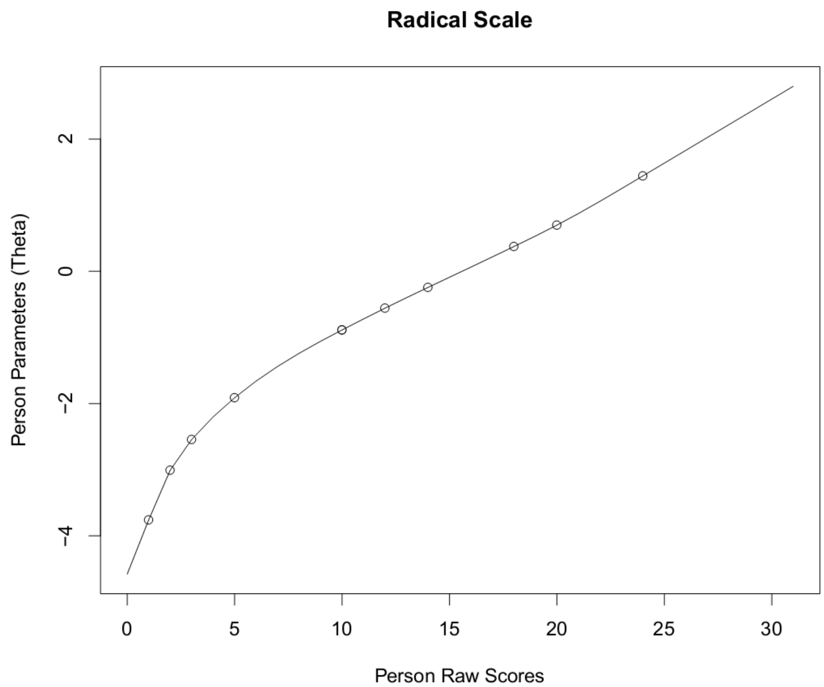

Figure 8. Radical Scale

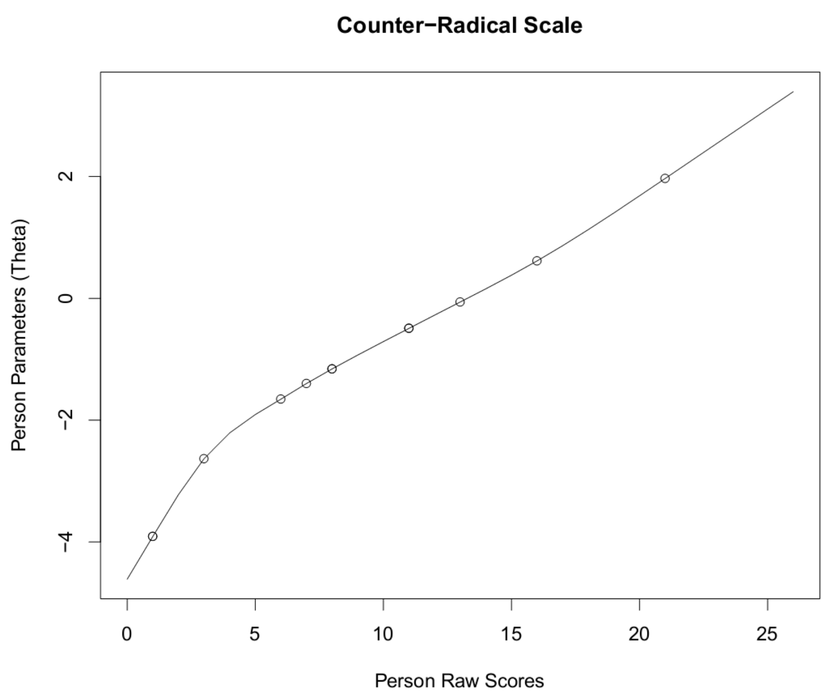

Figure 9. Counter-Radical Scale

- It would be preferable to plot the locations on a fixed range. However, the Rasch model is calculated for a scale on a latent range (Figures 8, and 9).
- Since this will be an interactive application, users would prefer to see almost instantaneous results. Yet, the eRm model generation is computationally expensive.

We resolve the first issue by uniformly scaling the ranges into [-10, 10], making it consistent with the inputs.

The second issue requires a more specific solution. We make use of the fact that the raw person scores pertaining to number of positive responses is a sufficient statistics for the Rasch model (Rasch 1961) to estimate scale values on the fly. Since we know the date range, and the selected organizations currently visible in the user interface, it's possible to quickly generate a response matrix for this subset of the data, and merge it with the previously known scale information to generate interpolated scale values.

The psuedo-code for the subset scale generation procedure is presented in Algorithm 1. The process starts with identifying the subset of documents in the $(start, end)$ date range (lines 2–5). Then the keyword frequencies, and thresholds are calculated for the entire set of organizations on this document subset (lines 6–14). Finally response tables for the subset of organizations is generated (lines 15–17), and then the sums need to be interpolated (lines 18–23), to be able to generate a scale on the $[-10, 10]$ range (line 24).

Here we've opted to include all the organizations in threshold calculations. This is because, the radical or counter-radical activity intensities are always measured relative to the other organizations participating in the same time period. However, while

Algorithm 1 Subset scaling algorithm

procedure Subset-Scale($D, O, start, end, scale$)
  $D' \leftarrow \{d \in D \mid Date(d) \geq start \wedge Date(d) \leq end\}$
  if $D' = \emptyset$ **then**
    return $\emptyset$
  end if

  for all $o \in O$ **do** ▷ Entire set of organizations
    $D'_o = \{d \in D' \mid Org(d) = o\}$
    for all $k \in Keywords$ **do**
      $f'_{o,k} = |\{d \in D'_o \mid k \in d\}| / |D'_o|$
    end for
  end for
  for all $k \in Keywords$ **do**
    $t'_k = Median(\{f'_{o,k}, o \in O'\})$
  end for

  for all $o \in O', k \in Keywords$ **do**
    $r'_{o,k} = f'_{o,k} > t'_k \rightarrow (t : 1, f : 0)$
  end for

  for all $o \in O'$ **do**
    $sum_o = \sum_{k \in Keywords} r'_{o,k}$
  end for
  for all $o \in O'$ **do**
    $S_o = Interpolate(sum_o, scale, -10, 10)$
  end for
  return S
end procedure

the scale is based on all the organizations, only the ones specifically asked will be presented to the user.

2.4 Experimental Evaluation

In order to measure the relation of the generated perspectives to the underlying scale, we have performed a series of experiments designed to compare their scaling capabilities to the gold standard ordering done by the experts.

2.4.1 Indonesian Corpus

The corpus domain is the online articles published by the web sites of the 23 religious organizations identified in Indonesia, in the Indonesian language. These sources are the web sites or blogs of the identified think tanks and organizations. As discussed in the introduction, each source was classified as either radical or counter-radical by the area experts. We downloaded a total of 37,000 Indonesian articles published in these 23 web sites, dating from 2001 to 2011. For each web site, a specific RegEx filter was used to strip off the headers, footers, advertising sections and to extract the plain text from the Html code.

2.4.2 The Quadrants Model

Our project leverages the results of our previous work, which relied on social theory including Durkheim's research on collective representations (Durkheim 2004), Simmel's work on conflict and social differentiation (Simmel 2008), Wallace's writ-

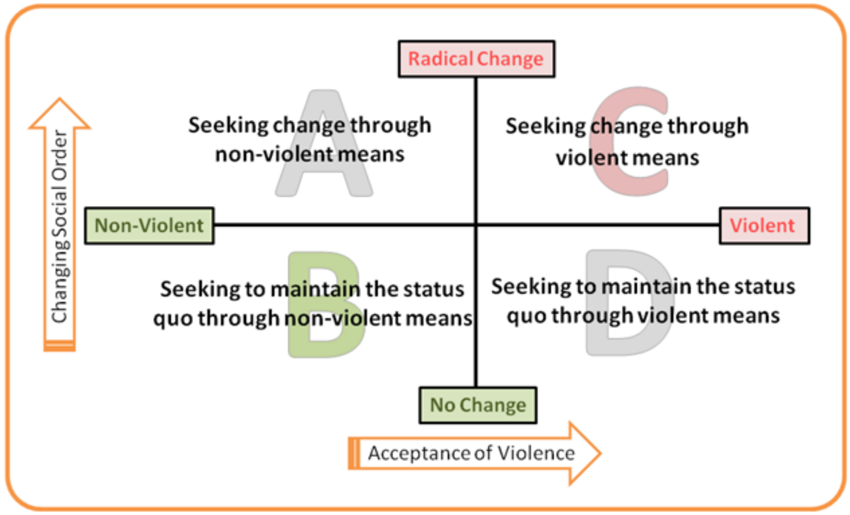

Figure 10. The quadrants model

ings on revitalization movements (Wallace 1956), and Tilly and Bayat's studies on contemporary social movement theory (Tilly 2004)(Bayat 2007). Our team has also developed, and is currently testing a theoretically based class model comprised of continuous latent scales. The first pair of scales focus on distinctions between the goals and methods of counter-radical and radical discourse, and capture the degree to which individuals, groups, and behaviors aim to influence the social order (Change Orientation) and the methods by which they attempt to do so (Change Strategies).

Quadrants model (see Figure 10) captures multiple social trends in four quadrants A, B, C, and D, and it makes the significant distinction between violent and not-violent dimensions of both radicalisms and counter radicalisms. Using the quadrants model, a researcher can locate organizations, individuals, and discourses in broader categories while still considering subtle differences between groups within categories. A researcher can document movement and trends from category to category, and identify points where movement is likely.

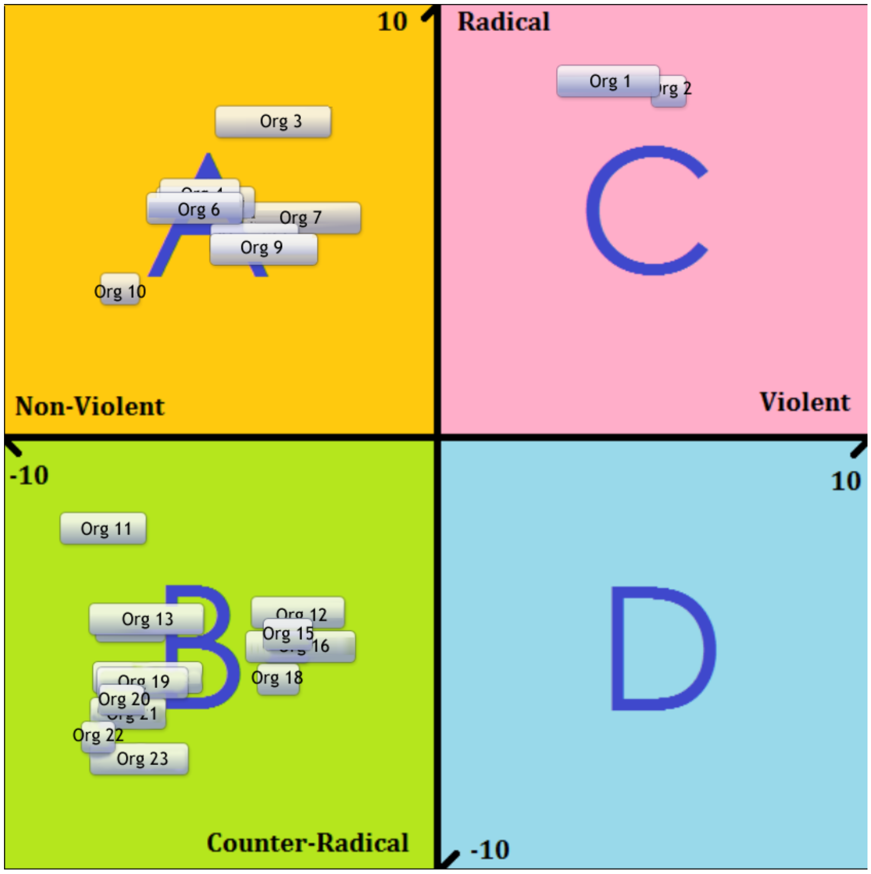

Figure 11. The visual interface of the expert opinion collector for manually placing the organizations on the two dimensional scale

2.4.3 Expert Opinion and Gold Standard of Rankings

We collaborated with three area experts, who collectively possess 35 years of scholarly expertise on Indonesia and Islam. We utilized a homegrown graphical drag-and-drop user interface to collect their opinion to build the gold standard of the rankings. A screenshot of this tool is shown in Figure 11.

Each expert separately evaluated and ranked the organizations in the dataset according to a two dimensional scale of radical / counter-radical (R/CR) and violent / non-violent (V/NV) axis. The consensus among the experts was high; since per item standard deviations among the experts' scores along the R/CR axis over a range of $[-10, 10]$, across all organizations were 2.75. The individual scores for each organization were combined and averaged to obtain the consensus *gold standard rankings* along the hypothesized R/CR scale.

The ranking discovered by the Rasch model fitting the corpus has been evaluated against the gold standard rankings of the organizations provided by the experts. We used two measures for evaluating the difference between two separate rankings, based on Spearman's footrule and Spearman's correlation coefficient. The original work utilized a mean displacement based measure as follows:

2.4.4 Evaluation Metrics

Given two discrete ordering functions G, and R, on the organization set O, the normalized displacement of a single organization is given as:

$$disp(G, R, O, o) = \frac{|G(o) - R(o)|}{|O|} \tag{2.2}$$

Here, O is the set of organizations, G and R are one to one mapping functions of rankings from set O to range $[1, |O|]$. Then overall $error$ measure for a given set of rankings was then defined as:

$$error(G, R, O) = \sum_{o \in O} \frac{disp(G, R, O, o)}{|O|} \tag{2.3}$$

For two exactly matching rankings, the $error(G, R)$ will be zero, whereas for two inversely sorted rankings it is expected to be 0.5 (when the size of O is even). Also a random ranking is expected to have a error of 0.375.

A work is in progress for building a publicly accessible expert opinion collection toolkit. The preliminary version can be accessed at: http://www.minerva-project.org/DataCollector.

2.4.5 Expert-to-Gold Standard Error

	Computational Rankings					Expert Rankings		
	Random	Sort	PCA	Rasch	Gold	Expert 1	Expert 2	Expert 3
Organization Rankings	21	9	1	6	1	2	7	1
	15	7	7	5	2	3	2	5
	22	1	9	2	3	1	4	8
	19	3	4	8	4	4	6	3
	4	4	12	4	5	5	3	4
	10	8	3	3	6	6	1	9
	6	2	8	1	7	8	5	2
	2	5	2	7	8	9	9	10
	17	6	17	9	9	7	8	7
	5	10	5	12	10	10	10	6
	11	23	21	21	11	14	13	12
	20	22	15	15	12	12	11	21
	13	13	11	11	13	11	19	16
	7	20	14	17	14	17	18	15
	3	18	16	14	15	15	15	19
	23	19	6	16	16	18	20	20
	1	16	20	19	17	13	22	14
	14	14	19	18	18	16	17	11
	12	17	18	20	19	22	14	18
	9	11	22	13	20	19	16	17
	16	15	13	22	21	20	23	23
	18	21	10	10	22	21	12	22
	8	12	23	23	23	23	21	13
Error	0.36	0.19	0.18	**0.10**		**0.06**	0.12	0.14

Figure 12. Computational and expert rankings

We calculated the error between each expert's ranking and their consensus gold standard of rankings. The first expert's error measure is 0.06, and the second and third expert's errors are 0.12 and 0.14 correspondingly as shown in the last row of the table in Figure 12. The average error of our experts against their gold standard ranking is 0.11.

2.4.6 Baseline - Sorting with Aggregate Score

The first baseline we used was constructed by sorting the organizations according to the number of different keywords observed in their corpus. While this provided a pattern similar to a Guttman Scale, and orderings of the organizations matched to a certain degree with the gold standard as shown in Figure 12, the *error* for this baseline was 0.19, which is higher than the average expert's performance.

2.4.7 Baseline - Principal Component Analysis

A stronger baseline was built by employing principal component analysis (Jolliffe 2002), and sorting the organizations according to their projections in the first principal component of the term-document matrix. Since experts selected the R/CR scale relevant keywords only, it was expected that the first principal component would reflect the corresponding scale. PCA proved to be performing better than the aggregate score sorting, with an *error* measure of 0.18. However, this error rate is still higher than the error rate of each expert.

2.4.8 Performance of the Rasch Model Ranking System

The Rasch models allow us to get a natural order of the organizations, according to their "abilities", i.e.: radicalism and counter-radicalism in this case. This system had an *error* measure of 0.10, which actually provided a higher ranking performance than the average performance of our experts' – performing better than the majority of our area experts.

2.4.9 Evaluations of the Intial Rasch Model Experiments

Our experiments showed that the hypothesized compatibility of the R/CR scale for the Indonesian corpus is valid. Not only the Rasch model was statistically fitting the response matrix, but also the generated ranking performance was better than the average expert performance. Among our computational baseline methods, the Rasch Model was the only method producing expert-level performance as shown in Figure 12. This preliminary analysis with the R/CR scale shows that when experts assist the system with keyword selection, the web corpus of organizations provides rich enough information and patterns to enable a computational method to rank them accurately.

2.5 Web Application Overview

A sample snapshot of the web application can be seen in Figure 13. It is composed of four main widgets for visualization and navigation. The top-left section which contains the **Search and Navigation widget** (1) that allows filtering of the document subset using parametric search queries and keyword based search criteria. The top-right section is the **Quadrant widget** (2) which displays the organizations active in the currently selected time frame on a two dimensional axis, using violence and radicalism scales. The bottom-left section consists of two **Treemap widgets** (3) which displays the demographics and the top keywords (markers) of the current selection. The bottom-right section has a **Timeline widget** (4) which provides a visualization of the keywords (markers) trends on a time line.

The navigation in the user interface starts with the Navigation widget (top-left) of

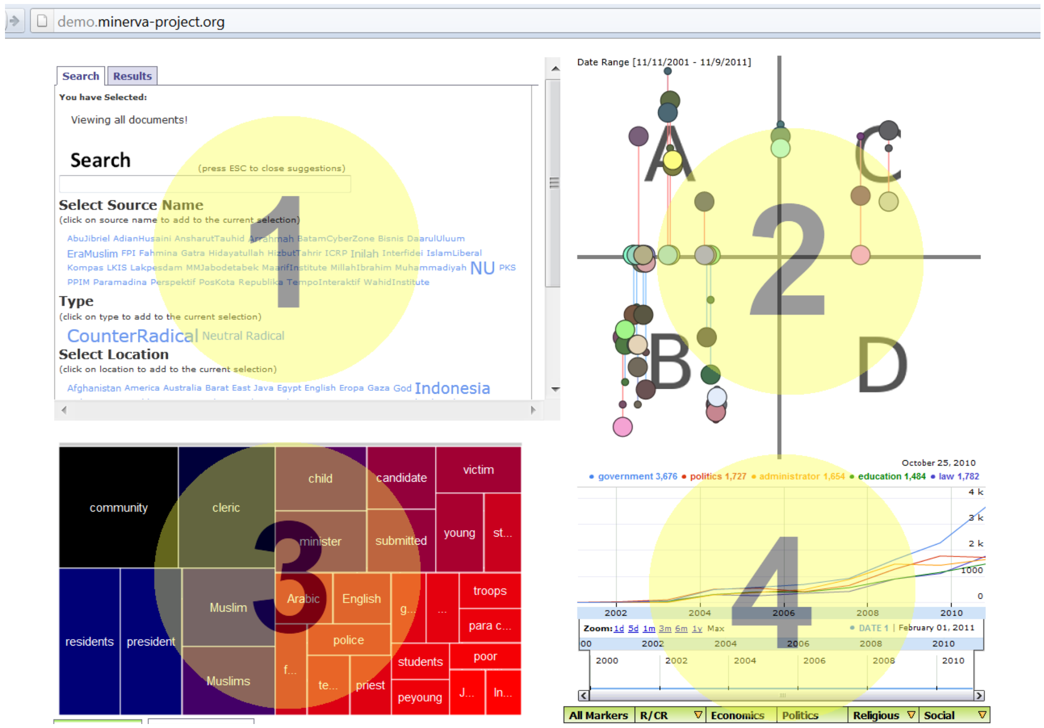

Figure 13. A sample snapshot of the web application.

the web application. Here the user is able to filter down the corpus utilizing full-text search queries, or faceting using keywords, locations, demographics, or choosing a subset of organizations. Any filtering done in this area, is then used by the rest of the application to focus the dataset on their respective widgets.

The Quadrant widget (top-right) provides a plot of the currently selected organizations on the two dimensional scale. The radical/counter-radical (R/CR) axis is dynamically calculated in real time, using the subset of organizations, and the time range of the current selection corpus to generate a response matrix, and use the previously calculated Rasch model to interpolate a scale value on that axis. The location change on the time range for each organization is shown as a color coded path, with three markers, a light circle corresponding to the position at the beginning of the period, a dark circle corresponding to the end of the period, and a dark-small circle for the middle. A red line between the circle denotes the rise of radical activities in the organization's behavior. A blue line denotes the opposite. The smaller circle is useful to see the overall movement of an organization. For example, between the range Aug 2005 and Aug 2007, EraMuslim's activities were radical (center of A quadrant), then became almost counter radical (the smaller circle denotes this mid point in the movement), and then jumped up again. The V/NV axis is retrieved from expert opinion in the current version, and dynamic calculation of this axis is left for a future version.

The Timeline widget (bottom-right) displays the trends of the most frequent markers on a time line. Initially the subset of markers presented defaults to all available, however it's possible to restrict the selection of markers to a more limited set among radical/counter-radical, economical, political, religious, or social domains. Timeline widget can also be used for selecting a date range of interest, which also is linked to the scale plot of the Quadrant widget.

The Treemap widgets (bottom-left) are used to display the relative frequencies of demographics and keywords (markers). The displayed marker category selection for this widget is synchronized with the Timeline widget.

In the following sections, we present some scenarios and findings to illustrate the capabilities of the web interface.

2.5.1 Scenario 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.

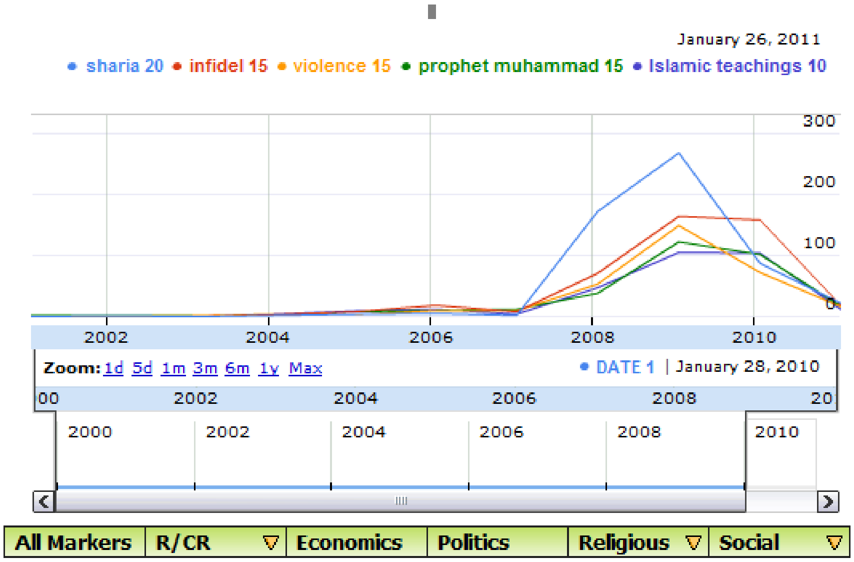

Figure 14. Trend of radical markers.

We can see[9] the most prominent markers associated with these radical organizations. Markers such as "infidel", "Sharia", and "violence" show an increasing trend between 2001 and 2011. A very strict interpretation of "Sharia" is used by radical organizations to justify their actions (Widhiarto 2010; Hasan 2009). "Sharia" peaks during this period as shown in Figure 14.

2.5.2 Scenario 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 (Frost, Rann, and Chin 2010; Rondonuwu and Creagh 2010) justified by a strict interpretation of

[9]Select the filter "Radical" from the search options and then in the Markers Menu select [Religious → Radical Markers]

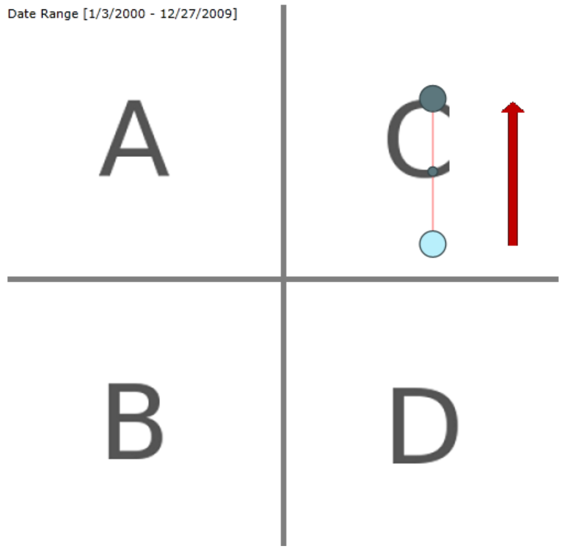

Figure 15. Consistent rise of FPI on the radical scale.

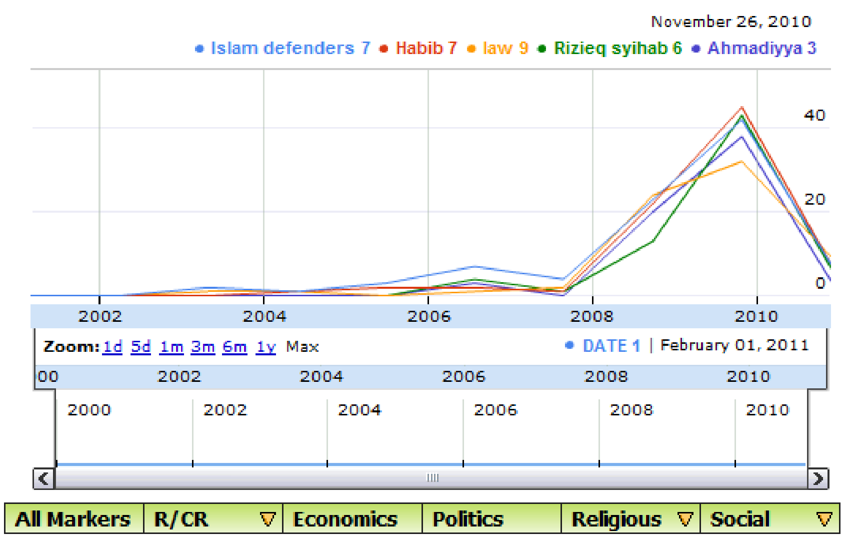

Figure 16. “Ahmadiyya” peaking during the period 2006 - 2010.

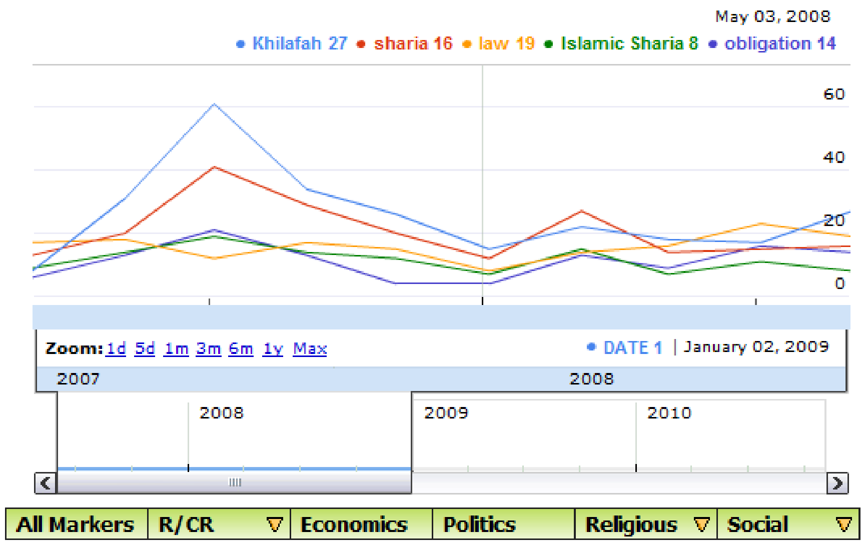

Figure 17. "Khilafah" ideology of Hizb ut- Tahrir.

Sharia(Study of Terrorism and Terrorism 2011). Our documents for FPI ranges between 2000 - 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 15. 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 radical markers such as "infidel", "Mujahedin", "pornography"[10]. Since 2006, we also saw a steep increase in the frequency of marker "Ahmadiyya", as shown in Figure 16, which indicates FPI's increased opposition to this heretical sect (Rahmat and Sihaloho 2011).

2.5.3 Scenario 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 (Ward 2009), which has been active in Indonesia since 1982 (Osman 2011). Between 2007 - 2009, our web application shows various radical and non-radical markers associated with this organization.

Radical	**Non-Radical**
"Sharia", "Infidel", "Caliph", "Violence"	"Politics", "Indonesian Islam", "Election", "Liberal", "Democracy"

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 (Hasan 2009). Hizb ut-Tahrir openly propagates the ideology of **Khilafah**, which believes in unification of all Muslim countries as a single Islamic State (Zakaria 2011; Mohamed Osman 2010). Figure 17 shows "Khilafah" as the most prominent marker[11] in Hizb ut- Tahrir's discourse.

By looking at the Quadrants widget (in Figure 18) we can infer that, HTI has been moderating its narrative.

[10] 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.

[11] 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]

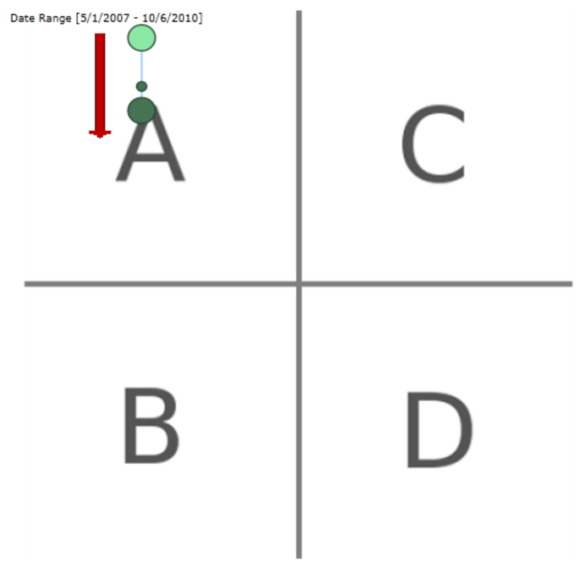

Figure 18. Decline of the HTI in the radical scale

2.5.4 Scenario 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 Bombing" using the keyword based Navigation widget.

The "counter radical" markers[12] 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".

On analyzing the occurrences of radical markers[13] in B-Quadrant, we find that

[12]Select CounterRadical filter in the search option, then from the Markers Menu select [R/CR → Counter Radical]

[13]In the Markers Menu select [R/CR → Radical]

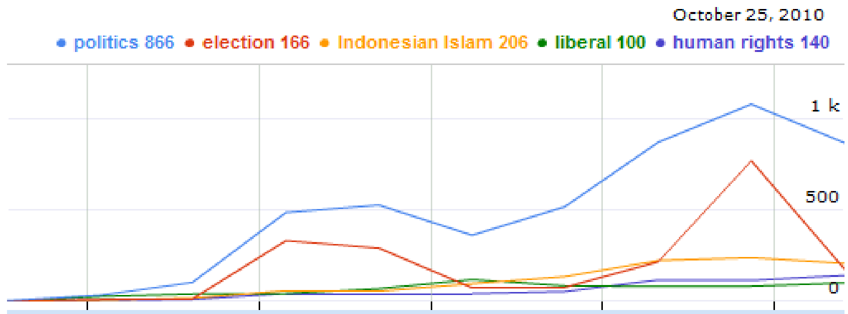

Figure 19. Counter Radical markers associated with CounterRadical organizations

Counter Radical organizations are very vocal against all of radical markers. One of the interesting radical markers is "Suicide Bombing". Most of the Counter Radical organizations are against suicide bombings.(Malang 2006). 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 Berubah?" (English translation: "Why I changed?")[14]. This article is by a reformed terrorist, debunking the misinterpretation of the jihad related verses used by violent groups.

[14]http://islamlib.com/id/artikel/mengapa-saya-berubah/

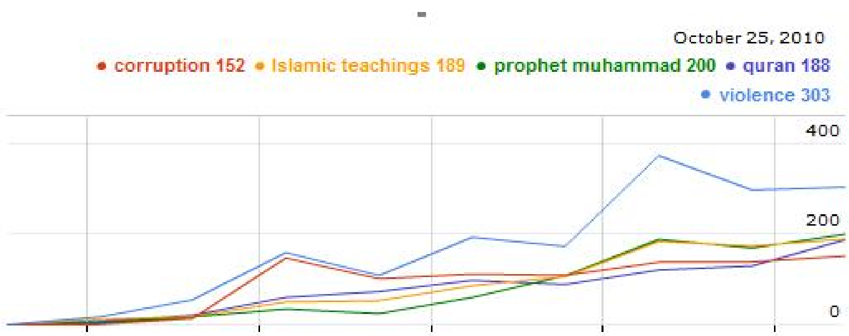

Figure 20. Radical markers associated with CounterRadical organizations

Chapter 3

PERSPECTIVE BASED SCALING

3.1 Debates and Perspective Analysis

Upon inspecting the keywords selected by our team of experts we observed that, some of these keywords correspond to differing perspectives on a set of topics that are debated within these web sites. Definition of **debate** is "a formal discussion on a particular topic in a public meeting or legislative assembly, in which opposing arguments are put forward."[15]. During a debate on a particular topic, like education, both radical and counter-radical organizations discuss different perspectives – such as "secular multi-cultural education" vs. "sharia based religious education".

During the design of an automated perspective detection algorithm, we made the following simplifying assumptions:

1. Organizations will *mostly* discuss their own perspective in a debate;
2. Organizations will *occasionally* mention others' perspectives, however, then relate them back to their own perspective.

In (Tikves et al. 2012), we published a mathematical formulation of the perspective keyword generation problem for a given topic, and provide an NP-Completeness proof of this problem, and design an exact solution through an ILP (integer linear programming) based solver.

[15] Oxford Online Dictionary

The input to this algorithm also takes the polarity suggestion from experts into consideration, for automatically identifying the discriminating perspectives of those organizations from opposite sides of a debate.

However, due to the algorithmic complexity and the strict constraints of the exact model, the ILP based solver was not always able to produce acceptable solutions. Namely, for larger debated topics, the run-time requirements[16] exceeded acceptable limits of the study, and for more intervened debates, none of the possible item sets could satisfy strict constraints of the ILP definition.

In order to resolve this, in our current version of the system, we have worked with a feature selection framework, SLEP. The discussion of the implementation of SLEP is discussed in the next section.

3.2 SLEP: A Sparse Learning Package

In order to address the scalability problem encountered in ILP we resorted to SLEP (Liu, Chen, and Ye 2009), again with the underlined motivation to select a subset of discriminating features that can (a) classify and (b) satisfy Guttman scale (McIver and Carmines 1981). The following steps describe our algorithm:

1. For each topic, calculate the frequency of the words occurring within a fixed size window of the topic keyword
2. Filter the term $\times$ document matrix to include only the most frequent 1000 words from each camp
3. Formulate the problem in a general sparse learning frame (Liu, Chen, and Ye

[16]Given data volume projections, we have estimated an upper bound of one hour run time restriction per topic. We have run the cplex ILP solver several hours for each topic before a timeout.

2009). Logistic formulation fits our application, since it is a dichotomous classification problem

$$\min_x \sum_{i=1}^{m} w_i \log(1 + \exp(-y_i(x^T a_i + c)) \tag{3.1}$$

$$+ \lambda |x|_1 \tag{3.2}$$

$$+ \frac{\rho}{2} ||x||_2^2 \tag{3.3}$$

where D_i is the *document* i and F_j is the feature $(word)$ j. A is the term $\times$ document matrix with all $A_{ij} \geq 0$, $y_i \in y$ is the class of each document D_i coded as +1 for *Radical* (R) and -1 for *Counter-Radical* (CR) and x_j is the weight for each feature F_j. Let us explain further the three terms involved in the convex optimization problem.

- $\sum_{i=1}^{m} w_i \log(1 + \exp(-y_i(x^T a_i + c))$, this first term is related to the logistic classification error. We set the weights w_i values to be all 1 so that all documents have the same weight.
- $\lambda |x|_1$, this term involving the L_1 norm deals with the sparsity of the solution vector x. We experienced with several lambda values which resulted with an x vector of various sparsity.
- $\frac{\rho}{2} ||x||_2^2$, since we were mainly driven by sparsity we do not use this last term, as it deals with the ridge regression, which is an extra level of shrinkage. We set the weight of $\rho = 0$.
- We used the MATLAB implementation of the SLEP package[17] which utilizes gradient descent approach to solve the aforementioned optimization problem. This package can handle matrices of 20M entries within a couple of seconds on a contemporary workstation.

[17] http://www.public.asu.edu/~jye02/Software/SLEP

- The features with non-zero values on the x vector are the candidate discriminants. Let $\mathcal{F}_R$, where $x_j > 0$ be the discriminant for the R class. Similarly, let $\mathcal{F}_{CR}$, where $x_j < 0$ be the discriminant for the CR class due to the coding schema in step 3. Given that the optimized formulation resulted with a sparse x vector, most of the words F_j had $x_j = 0$ and hence were not included in either $\mathcal{F}_R$ or $\mathcal{F}_{CR}$.

Note that the sets of features $\mathcal{F}_R$ and $\mathcal{F}_{CR}$ may not satisfy the Guttman pattern. These sets needed to be further filtered such that $\mathcal{F}'_R \subseteq \mathcal{F}_R$ and $\mathcal{F}'_{CR} \subseteq \mathcal{F}_{CR}$ for this purpose.

3.2.1 Post Processing of the Perspectives

While the exact solution of this formulation will provide us with features to distinguish between radical, and counter-radical perspectives, they will include items with very low support due to their exhaustive nature. This brings two issues, namely the model will overfit the data, and also the set of results might not always be relavant.

In order to overcome these limitations, we employed a simple frequency based filter, which will only include items that occur significantly in the respective radical or counter-radical subset of the corpus. This provided us with a much cleaner result set.

3.3 Feature Expansion Algorithm

We have observed that including all of the newly discovered features in the scale resulted a poor performance. This is because, they neither provided the desired

Guttman pattern, nor the resulting scale aligned with the expert opinion. However, exhaustively enumerating all possible subsets to find an optimal one would also be undesirable due to time complexity. Thus we have devised a greedy expansion based algorithm to select the items that make up the scale. It chooses a sufficiently optimal subset of these features by expanding an initial set, incrementally adding features that offer a higher performance.

One possible implementation is shown in the algorithm in Fig. 21. This greedy algorithm will start from an initial set of features I, and iteratively select the features that increase the performance of the solution. The performance of a solution is evaluated by the Solve function, which takes a candidate input, and returns the performance value according to expert agreement.

Each iteration of the loop (lines 3 – 15) tries to iteratively expand the current set of selected features (lines 12 – 14). First, it evaluates the performance of the currently selected subset (line 4), and then identifies each not yet selected feature that provides a performance increase (lines 7 – 8), and finally collects them into the selected feature set for the next iteration (lines 5, and 8 – 10). When it can no longer include any new features, the algorithm will terminate.

Another performance trade-off was done using the natural grouping of the features. Since the features in our problem are grouped by topic, we decided to keep these natural groupings, thus making each c in set C a collection of features.

Here the $\unrhd$ comparison function will assume greater-than-or-equal-to semantics. This is because, while we want to have the best possible scoring features as possible, we also want to be able to have a larger set of perspectives that can be used to explain the underlying latent scale.

In order to be able to handle the case of an empty initial feature set (I), we ex-

```
1:  procedure Greedy-Selection(I, C, ⊵)
    initial features I
    candidate features C
    comparison function ⊵

2:      S ← I
3:      repeat
4:          m ← Solve(S)
5:          N ← ∅
6:          for all c ∈ C \ S do
7:              p ← Solve(S ∪ {c})
8:              if p ⊵ m then
9:                  N ← N ∪ {c}
10:             end if
11:         end for
12:         if N ≠ ∅ then
13:             S ← S ∪ N
14:         end if
15:     until N = ∅
16:     return S
17: end procedure
```

Figure 21. Feature set expansion algorithm

panded the algorithm as shown in Fig. 22. This modification (to lines 6 – 11 in the original algorithm) will choose the best available features in the first iteration that are within a score difference of δ of each other. This also assumes Solve function will return a sensible upper limit value when an empty set is given as its input.

3.4 eRm Iterative Item Elimination Algorithm

While the eRm Rasch analysis package already does trivial eliminations in the model (for example, ignoring full/empty 1/0 responses), it also provides an al-

for all $c \in C$ **do**
  $p \leftarrow \text{Solve}(S \cup \{c\})$
  if $p \trianglerighteq m$ **then**
    if $S = \emptyset \wedge p - m < \delta$ **then**
      $N \leftarrow \{c\}$
      $m \leftarrow p$
    else
      $N \leftarrow N \cup \{c\}$
    end if
  end if
end for

Figure 22. Feature set expansion algorithm modification, enabling special handling of the empty initial set of features

gorithm to clean up a model from features that do not adhere to the Rasch model/Guttman pattern.

The overall idea of the algorithm is summarized in the algorithm in Fig. 23. While our greedy feature selection algorithm worked by expanding a set of features, this algorighm works by going the opposite direction, and reducing the feature set in each step. Here \$ is the R member access operator, where $r\$x$ is the feature set of rasch model object r, and $\setminus$ is set difference. Functions Eval, and LowestRankedFeature are references to eRm provided facilities to evaluate, and find the worst contributing item of Rasch models.

3.5 Baseline Performance

In our previous study (Tikves et al. 2011) we have automatically generated a Rasch model from the organizational corpus data, and the expert selected items. We have observed that, against several baseline algorithms, including score sorting, and princi-

```
procedure StepWiseIt(m, Eval)
Rasch Model (RM) m
Evaluation Function Eval

    r ← m
    repeat
        e ← Eval(r)
        if not Fits(e) then
            i ← LowestRankedFeature(e)
            x ← r$x \ {i}
            r ← RM(x)
        end if
    until Max # of Steps, or Fits(e)
    return r
end procedure
```

Figure 23. Feature elimination algorithm provided by the eRm package

pal component analysis, the Rasch model was able to demonstrate the best available performance, and was ranked at expert level.

In order to have a baseline for comparison of the automatically generated items, we have opted to use this scale also in the perspectives version of the study.

3.6 Candidate Perspectives

We have run both the ILP, and the SLEP based feature generators on all the 50 topics that has been identified. ILP was able to identify perspectives for 18 of the topics, while failed for the rest, due to either finding no viable exact solution, or timeouts. This resulted in a total of 2869 perspectives, with 159 average on each topic. Since these exact features also included items with very low support, we have filtered these results to include only the ones with higher frequency in the corpus. The final set contained a total of 227 perspectives on all 18 solved topics. On the other hand, SLEP was able to successfully generate candidate perspective on every 50 topic, totaling 1065 perspectives, with an average of 21 on each topic.

3.6.1 Interpretation of the Goodness of Fit statistics

The Rasch package provides an internal analysis of whether the data fits the theoretical model. In practice, we used the provided $LR - test$ as a "PASS/FAIL" metric, and removed the candidate models when they were refused by this test.

3.7 History of Our Work on Scale Generation

As desctibed in chapter refchap:scaling, our previous work (Tikves et al. 2011) depended on more direct interaction with experts' opinion to build a model that can capture the underlying dynamics of the scale. The experts both provided a set of target organizations, and also directly selected the items that would make up the scale, from a machine generated candidate list. The candidate list consisted of the union of top-100 n-grams from each organization's individual corpus, which were a total of 790 items. The resulting scale has utilized a total of 55 of keywords selected by experts.

Using this framework, we were able to build the scales, that would both demonstrate no lack of fit with the Rasch model, and also performed at expert level accuracy.

After demonstrating the feasibility of the Rasch model in ranking the organizations on R/NR scale using manually selected n-gram based items, we experimented with automatically extracted perspectives as the items. This not only produced better results, but also possible perspectives that could be utilized the explain the underlying discourse.

3.7.1 Response Table Extraction

Similar to the previous work (in Chapter 2.3.3), we utilized information extraction techniques to build response tables But utilizing automatically extracted perspectives, instead of expert selected keywords.

When perspectives were used, instead of the expert selected n-grams, the document corpus is replaced with the subset the contained the topic keyword, and also frequencies were calculated only for the perceptive words present in a short window of these topic keywords. After this change, the same method was utilizied to generate response tables for each and every topic in the study.

3.8 Aligning Perspectives with the Scales

In order to identify the perspectives that make up the theoretical scale we are working on (R/CR bi-polar scales on Indonesian Islamic religious organizations), we have devised a set of experiments that measure their relation to the Rasch model, and the expectation of field experts.

Initially, as a baseline, we have re-run the original scale with the expert selected features, with the new evaluation metrics. The mean displacement of the features was 0.1172, while the mean square displacement score was 0.0287. (The slight difference with the original paper is due to the handling of the missing items, discussed in Section 2.4.3).

In order to observe the effect of the StepWiseIt, we have run the elimination algorithm on the original set of features. The mean displacement was decreased to 0.1115, while the mean square displacement stayed the same. The algorithm has eliminated 15 features to reach this score. The summary of these experiments can be seen in Table 1

Table 1. Experimental results for the original expert selected feature based scales

	error	msd	run time
Original	0.1172	0.0287	14s
Original + StepWiseIt	0.1115	0.0287	53s

3.9 Evaluation Metrics

In addition to the measure forumated in Section 2.4.4, we have also opted to include another measurement to take stability of the items into consideration. Based on the $L^2 - Norm$ of the normalized displacement function, the msd measure can be defined as the following:

$$msd(G, R, O) = \sum_{o \in O} \frac{disp(G, R, O, o)^2}{|O|} \tag{3.4}$$

Since our initial work, we have also modified the evaluation of the missing items. Specifically, for empty/full response patterns, the Rasch model would not be able to make any inference. Since we experimented with dynamic features, and the missing items varied in each test, we have opted to position them in their neutral places. This change has introduced a slight difference from the experimental results of our original study.

3.10 The Initial Experiments with Feature Expansion

After establishing the baseline, we evaluated the perspective based features discovered by the ILP solver. First we built a model including all the candidate features proposed by the solver. This resulted in an mean displacement of 0.1323 and mean square displacement of 0.0284. While the performance was near the expert level, the hand selected features performed (13%) better than this initial run.

Then the features were refined with StepWiseIt, and our Greedy-Selection algorithms. The StepWiseIt failed to provide better results, and actually performed worse, with mean displacement of 0.1632, and mean square displacement of 0.0386, while failing the $LR - test$ for Rasch model fitness. The likely reason for this is that StepWiseIt performs item eliminated locally based on individual item fitness, but the sparse nature causes loss of global Guttman pattern.

When we built an optimum item set from scratch using the Greedy-Selection algorithm, we were able to identify 14 topics that contributed with better fitting perspectives. The expanding topic sets can be seen in Table 2. The final solution had a mean displacement of 0.1020, with a mean square displacement of 0.0189. An additional cleanup using the StepWiseIt algorithm over this existing solution did not produce better results.

The summary of these experiments can be seen in Table 3.

Table 2. The topics chosen be the Greedy-Selection algorithm from the candidate perspectives of the ILP solution.

Iteration	Topics
1	kufur disbelief
2	kdrt, kekafiran, kesetaraan, konstitusional, multikultural, sekularisme, tabligh, toleransi (*domestic violence, infidelity, equality, constitutional, multicultural, secularism, tabligh, tolerance*)
3	bunuh, gender, homoseksual, musyrikin, syirik (*suicide, gender, homosexuals, idolaters, paganism*)

Table 3. Scaling experiments with the ILP solver based data

	error	msd	run time
ILP	0.1323	0.0284	3m:38s
ILP + StepWiseIt	0.1632	0.0386	56m:47s
Greedy(ILP)	0.1020	0.0182	
Greedy(ILP) + StepWiseIt	0.1122	0.189	

3.10.1 SLEP Based Features

In addition to the ILP based exact features, we also ran separate experiments for the SLEP output. These yielded a total of 449 features on counter radical, and 616 features on the radical scales. The overall run time duration was 6 hours and 4 minutes. The resulting scales had a mean displacement of 0.1398 and mean square displacement of 0.0312. We opted not to run the StepWiseIt on this particular case, since the expected run time would be in the order of weeks, which would not be practical for the real life conditions of the project.

Like the ILP based candidates, we also ran the Greedy-Selection algorithm on the SLEP input (Table 4). Over two iterations, the algorithm was able to identify 15 top-

Table 4. Scaling experiments with the SLEP solver based data

	error	msd	run time
SLEP	0.1398	0.0312	6h:04m
Greedy(SLEP)	0.0982	0.0189	

Figure 24. Run time performance of the Rasch model fitting algorithm in the eRm package. The x axis corresponds to the number of items, while the y axis represents the run time length in seconds. Notice that the scatter plot shows fitness to the x^2 polynomial prediction line.

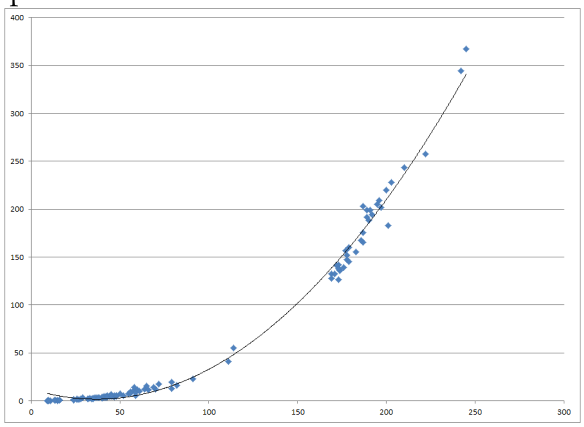

Table 5. The topics chosen be the Greedy-Selection algorithm from the candidate perspectives of the SLEP solution.

Iteration	Topics
1	manusia (*human*)
2	beragama, bunuh, dakwah, demokrasi, jihad, kafir, kristen, liberal, multikultural, pluralisme, politik, sipil, syariat, syirik (*religion, kill, propaganda, democracy, jihad, infidel, Christian, liberal, multicultural, pluralism, political, civil, Sharia, polytheism*)

ics, whose perspectives were closely related to the underlying scale. The expanding topic set can be seen in Table 5. The best mean displacement achieved was 0.0982, with a corresponding mean square displacement of 0.0189.

The main reason that this table does not share a significant amount of topics with

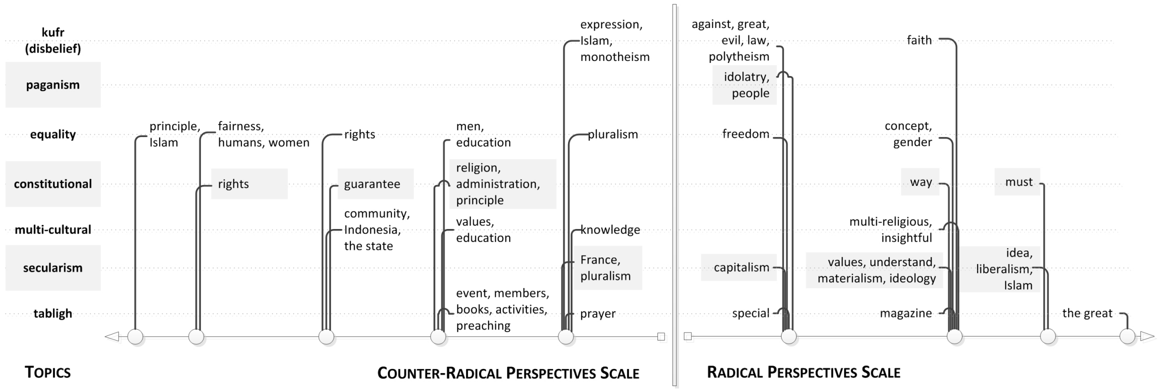

Figure 25. A sample set of perspectives generated by the ILP based solver. Here each row represents a debate topic, while the linear scales represent the locations of the perspectives. The left side items are the counter-radical, and the right side items are the radical perspectives in each of these topics.

the ILP based topic set, is that the ILP solver could not provide results for the great majority of the topics selected by SLEP. The common ones, like “multikultural”, “syirik” were selected in both, while similar topics (like “politik”/”konstitusional”) were chosen when available.

3.10.2 Sample Perspectives

A set of sample perspectives selected by the ILP solver are displayed in Figure 3.10.2. Here the columns represent individual topics, while two rows correspond to radical, and counter-radical perspectives on these topics. The items have been machine translated from Indonesian into English.

Chapter 4

RESULTS AND DISCUSSION

In our studies we developed an end-to-end system that semi-automatically ranks organization on social scales, based on their public web corpus, along with input from field experts. The field experts provide initial organization lists, and help with topic and scale assignments, while the automated system handles tasks beginning with crawling the web corpus, ending with the final ranking of these organizations on each scale.

Specifically, the experts provide:

1. Prepare the scale definitions (offline)
2. Enter organization lists, and their home URLs (web UI)
3. Provide binary categorizations of organizations (web scaling UI)
4. Assign topic to scales, after the initial analysis of the system (web UI)

While the system automates:

1. Crawling the web pages, extracting text, and generating the initial corpus (Section 2.3.1)
2. Discovering discourse topics
3. Discovering the possible perspectives of organizations on each topic by scale, based on the binary classifications, and the web corpus, after experts complete step 4
4. Generating "response tables", to assign the organizations to set of perspectives. These tables present the discovered perspectives of organizations on each topic.

5. Building ranking models based on Rasch model, and ordering the organizations on these scales

Our primary contribution was enabling social scaling methods for the web corpus, with minimal expert interaction. We were able to demonstrate the efficacy of our techniques based on the early "Quadrants" model which is based on *Radicalism*, and *Violence* bi-polar scales with the Indonesia case (presented in (Tikves et al. 2011)).

The benefit of Rasch model in ranking is being able to asses both the ranked objects, and the keywords in the same dimension. It also allows researchers to make deductions on where each ranked object stands wrt. to features, and which features are more likely to be observed in individual objects (organizations in this study). This also true in reverse, thus enabling calculating the probabilities of observing a feature in different ranked objects (organizations).

Possible future extension of this work can include a working prediction model, for organizations that are not part of the initial study, incorporating other sources, especially social media outlets, and finally reducing the required expert input, possibly replacing some steps with "crowdsourcing", thus enabling to work on different geographical regions where experts many not be readily available.

REFERENCES

Andrich, D. 1988. *Rasch models for measurement.* Sage.

Bayat, A. 2007. *Making Islam Democratic: Social Movements and the Post-Islamist Turn.* Stanford University Press.

Crelinsten, R.D. 2002. "Analysing Terrorism and Counter-terrorism: A Communication Model." *Terrorism and Political Violence* 14:77–122.

Dahlberg, L. n.d. "The Internet and Democratic Discourse: Exploring The Prospects of Online Deliberative Forums Extending the Public Sphere" (December): 615–633. http://www.ingentaconnect.com/content/routledg/rics/2001/00000004/00000004/art00007.

Drehmer, D.E., J.A. Belohlav, and R.W. Coye. 2000. "An exploration of employee participation using a scaling approach." *Group & Organization Management* 25 (4): 397.

Durkheim, E. 2004. "The cultural logic of collective representations." *Social theory the multicultural and classic readings, Wesleyan University: Westview Press.*

Frost, Frank, Ann Rann, and Andrew Chin. 2010. *Terrorism in Southeast Asia.* http://www.aph.gov.au/library/intguide/FAD/sea.htm. [Online; accessed 21-November-2011], November.

Gelgi, Fatih, Hasan Davulcu, and Srinivas Vadrevu. 2007. "Term Ranking for Clustering Web Search Results." In *WebDB.*

Guttman, L. 1950. "The basis for scalogram analysis." *Measurement and prediction* 4:60–90.

Hartigan, J.A., and M.A. Wong. 1979. "Algorithm AS 136: A k-means clustering algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1): 100–108.

Hasan, N. 2009. "Islamic militancy, Sharia, and democratic consolidation in post-Suharto Indonesia." *RSIS Working Papers ; 143/07.*

Himelboim, Itai. 2010. "Civil Society and Online Political Discourse: The Network Structure of Unrestricted Discussions." *Communication Research* (October). doi:10.1177/0093650210384853. http://dx.doi.org/10.1177/0093650210384853.

Hunter, D.R., and K. Lange. 2004. "A tutorial on MM algorithms." *The American Statistician* 58 (1): 30–37.

Jolliffe, I.T. 2002. *Principal Component Analysis.* Springer Series in Statistics.

Likert, Rensis. 1932. "A Technique for the Measurement of Attitudes." *Archives of Psychology* 140:1–55.

Liu, J., J. Chen, and J. Ye. 2009. "Large-scale sparse logistic regression." In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining,* 547–556. ACM.

Mair, P., and R. Hatzinger. 2007. "Extended Rasch modeling: The eRm package for the application of IRT models in R."

Malang. 2006. *NU chairman deplores suicide bombing attempt.* http://www.nu.or.id/page/en/dinamic_detil/15/28282/News/NU_chairman_deplores_suicide_bombing_attempt.html. [Online; accessed 22-November-2011], November.

McIver, J.P., and E.G. Carmines. 1981. *Unidimensional Scaling.* Vol. 24. Sage Publications, Inc.

Mohamed Osman, Mohamed Nawab. 2010. "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. doi:10.1080/09546553.2010.496317. eprint: http://www.tandfonline.com/doi/pdf/10.1080/09546553.2010.496317.

Osman, Mohamed Nawab Mohamed. 2011. "Preparing for the caliphate." *The Asian Studies Association of Australia's E-Bulletin* (September): 14–16.

Papacharissi, Zizi. 2002. "The virtual sphere: the internet as a public sphere." *New Media Society* 4, no. 1 (March): 9–27. http://nms.sagepub.com/cgi/content/abstract/4/1/9.

Pawitan, Y. 2001. *In all likelihood: statistical modelling and inference using likelihood.* Oxford University Press, USA.

Rahmat and Markus Sihaloho. 2011. *FPI Vows to Disband Ahmadiyah 'Whatever It Takes'.* http://www.thejakartaglobe.com/home/fpi-vows-to-disband-ahmadiyah-whatever-it-takes/423477. [Online; accessed 21-November-2011], February.

Rasch, G. 1961. "On general laws and the meaning of measurement in psychology." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Psychology, 4,* 332.

Rondonuwu, Olivia, and Sunanda Creagh. 2010. *Opposition grows to Indonesia's hardline FPI Islamists.* http://in.reuters.com/article/2010/06/30/idINIndia-49777620100630. [Online; accessed 21-November-2011], June.

Salton, Gerard, and Christopher Buckley. 1988. "Term-weighting approaches in automatic text retrieval." In *Information Processing and Management,* 513–523.

Simmel, G. 2008. *Sociological Theory.* New York: McGraw-Hill.

Study of Terrorism, National Consortium for the, and Responses to Terrorism. 2011. *Terrorist Organization Profile: Front for Defenders of Islam.* http://www.start.umd.edu/start/data_collections/tops/terrorist_organization_profile.asp?id=4026. [Online; accessed 21-November-2011].

Thurstone, L. L. 1928. "Attitudes can be measured." *American Journal of Sociology* 33:529–554.

Tikves, Sukru, Sujogya Banerjee, Hamy Temkit, Sedat Gokalp, Hasan Davulcu, Arunaba Sen, Steven Corman, et al. Accepted for publication. "A system for ranking organizations using social scale analysis." *Social Network Analysis and Mining (SNAM).* http://link.springer.com/article/10.1007%2Fs13278-012-0072-x.

Tikves, Sukru, Sujogya Banerjee, Hamy Temkit, Sedat Gokalp, Hasan Davulcu, Arunabha Sen, Steven Corman, Mark Woodward, Inayah Rochmaniyah, and Ali Amin. 2011. "A System for Ranking Organizations Using Social Scale Analysis." In *EISIC,* 308–313. IEEE. http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6059524.

Tikves, Sukru, Sedat Gokalp, M'hamed H. Temkit, Sujogya Banerjee, Jieping Ye, and Hasan Davulcu. 2012. "Perspective Analysis for Online Debates." In *ASONAM,* 898–905.

Tilly, C. 2004. *Social Movements.* Boulder, CO, USA: Paradigm Publishers.

Tunkelang, Daniel. 2009. *Faceted Search.* Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers. http://dx.doi.org/10.2200/S00190ED1V01Y200904ICR005;%20http://dx.doi.org/10.2200/S00190ED1V01Y200904ICR005.

Wallace, A. 1956. "Revitalization Movements." *American Anthropologist* 58:264–281.

Ward, Ken. 2009. "Non-violent extremists? Hizbut Tahrir Indonesia." *Australian Journal of International Affairs* 63 (2): 149–164. doi:10.1080/10357710902895103. eprint: http://www.tandfonline.com/doi/pdf/10.1080/10357710902895103.

W.B. Michael, J. Kogan. 2010. *Text Mining: Applications and Theory.* Wiley.

Widhiarto, Hasyim. 2010. *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. [Online; accessed 21-November-2011], June.

Woodward, Mark, Inayah Rohmaniyah, Ali Amin, and Diana Coleman. 2010. "Muslim Education, Celebrating Islam and Having Fun As Counter-Radicalization Strategies in Indonesia." *Perspectives on Terrorism* 4 (4).

Zakaria, Yamin. 2011. *A Global Caliphate: Reality or Fantasy?* http://usa.mediamonitors.net/content/view/full/91207. [Online; accessed 21-November-2011], November.