

**Said-Huntington Discourse Analyzer:
A machine-learning tool for classifying and analyzing discourse**
Muhammad Ali Hashmi
MSJ, Northwestern University, 2012

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

© Massachusetts Institute of Technology 2015. All rights reserved

Author.....
Muhammad Ali Hashmi
Program in Media Arts and Sciences
May 8, 2015

Certified by.....
Ethan Zuckerman
Principal Research Scientist
MIT Media Lab
Thesis Supervisor

Accepted by.....
Pattie Maes
Academic Head, Program in Media Arts and Sciences

**Said-Huntington Discourse Analyzer:
A machine-learning tool for classifying and analyzing discourse**

Muhammad Ali Hashmi

Submitted to the
Program in Media Arts and Sciences,
School of Architecture and Planning, on May 8, 2015,
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences at the
Massachusetts Institute of Technology

Abstract

Critical discourse analysis (CDA) aims to understand the link “between language and the social” (Mautner and Baker, 2009), and attempts to demystify social construction and power relations (Gramsci, 1999). On the other hand, corpus linguistics deals with principles and practice of understanding the language produced within large amounts of textual data (Oostdijk, 1991). In my thesis, I have aimed to combine, using machine learning, the CDA approach with corpus linguistics with the intention of deconstructing dominant discourses that create, maintain and deepen fault lines between social groups and classes. As an instance of this technological framework, I have developed a tool for understanding and defining the discourse on Islam in the global mainstream media sources. My hypothesis is that the media coverage in several mainstream news sources tends to contextualize Muslims largely as a group embroiled in conflict at a disproportionately large level. My hypothesis is based on the assumption that discourse on Islam in mainstream global media tends to lean toward the dangerous “clash of civilizations” frame. To test this hypothesis, I have developed a prototype tool “Said-Huntington Discourse Analyzer” that machine classifies news articles on a normative scale— a scale that measures “clash of civilization” polarization in an article on the basis of conflict. The tool also extracts semantically meaningful conversations for a media source using Latent Dirichlet Allocation (LDA) topic modeling, allowing the users to discover frames of conversations on the basis of Said-Huntington index classification. I evaluated the classifier on human-classified articles and found that the accuracy of the classifier was very high (99.03%). Generally, text analysis tools uncover patterns and trends in the data without delineating the ‘ideology’ that permeates the text. The machine learning tool presented here classifies media discourse on Islam in terms of conflict and non-conflict, and attempts to put light on the ‘ideology’ that permeates the text. In addition, the tool provides textual analysis of news articles based on the CDA methodologies.

Thesis supervisor: Ethan Zuckerman
Title: Principal Research Scientist
Director, Center for Civic Media

**Said-Huntington Discourse Analyzer:
A machine-learning tool for classifying and analyzing discourse**

Muhammad Ali Hashmi

Thesis Advisor:

Ethan Zuckerman
Principal Research Scientist
MIT Media Laboratory

Thesis Reader:

Edward Schiappa
John E. Burchard Professor of the Humanities
MIT

Thesis Reader:

Catherine Havasi
CEO and Co-Founder at Luminoso
Research Affiliate, MIT Media Laboratory

Acknowledgements

To the members of my thesis committee, Ethan Zuckerman, Dr. Catherine Havasi, Dr. Edward Schiappa, I extend my deepest gratitude.

I want to thank my supervisor, Ethan Zuckerman, for his extraordinary support and assistance in the preparation of this thesis. Imbued with the qualities of head and heart, Ethan is an institution in himself. It was a privilege to work with one of the leading media theorists of our time, from whom I have learnt so much.

I want to thank Dr. Catherine Havasi, my thesis reader, who played an instrumental role in the completion of this project. Her expertise in machine learning and her knowledge of computational linguistics were invaluable for my thesis.

I want to thank Professor Edward Schiappa, my thesis reader, for helping me frame this work in terms of its structure and overarching aims; his seminal work “Defining Reality” had a formative impact on the critical aspects of this thesis.

I would like to extend my warmest thanks to Lorrie Lejeune for her constant support during the course of my thesis project.

I would like to thank my MIT colleagues— Erhardt Graeff, Matt Carroll, Nathan Matias, Catherine D’Ignazio, Edward Platt, Patsy Baudoin, Heather Craig, Dalia Othman, Yu Wang, Jude Mwenda, Alexis Hope, Rahul Bhargava, Matt Stempeck, Sands Fish, Chris Peterson, Chelsea Barabas and Emilie Reiser—for being an inspiration to me.

I would be remiss if I didn’t thank Zeba Khan and Farah Chaudhry— subject matter experts on media and Islam—for sharing their insights on my research topic.

I would like to dedicate this thesis to the memory of my father, Ali Akbar Hashmi, who passed away in 2005. He was a humble, simple and generous man. His kindness left an indelible mark on me.

Finally, I would like to thank my dear friends and family: my uncles—Farrukh Seer, Nadeem Haq, Muhammad Shahrukh, Saleem-ul-Haq, Fahim-ul-Haq, Naved Haqqi and Syed Muhammad Khalid—for their love and care; Omar Anwaar, Sohaib Anwaar and Wasif Yaqeen for being *jigars par excellence*; Naeem Raza, Amer Inayat, Waqar Cheema, Shahbaz Noor and Irfan Hussain for music, transcendence and madness; Numan Sheikh for his friendship, and for being part of my formative ‘beat’ years; Jawad Shakeel, my best friend since grade one, for supporting me in all my ventures; Mosharraf Zaidi, Saima Zaidi, Saud Naqvi and Shahzad Khan for believing in me; Javed Mian, genius among geniuses, for believing in ‘me-we’; Farhan Khokhar, my brother-in-law, for his support and encouragement; Hafeez Jamali for being a true comrade and friend, and for offering sound personal and theoretical advice; my dearest mother, Nasreen Hashmi, who has fought brain tumor for decades, for being an extraordinary example of courage, love, and strength; Hammad Hashmi, my brother, for being my anchor and pillar of support, and

for being there through the thick and thin; my brother, Saad Hashmi, who signifies lust-for-life in my life, for providing constant support and resources; Aiza Hashmi, my sister, who has always been my unshakable companion, for her unrelenting moral support; Manaal Hashmi, my daughter, for being the meaning of the words ‘blessing, happiness and joy’ in my life; and, most importantly, my wife, Javeria Ali Hashmi—the love of my life, the muse for my verses, my shelter from the scorching sun, my blanket of cool wind— for making everything possible for me.

Table of Contents

Abstract	3
Acknowledgements	7
Table of Contents	9
Table of Figures	11
Abbreviations	12
Problem	13
Proposed Work	18
Said-Huntington Discourse Analyzer	19
Related Work	21
Implementation & Methods	24
Media Cloud	24
Query for Retrieving Media Articles on Islam	25
Rolling-30-Day Corpus	26
Defining Conflict	27
Pre-2001 Corpus	28
Conflict Terms Word List	29
Classification Methods	29
Feature Engineering	32
Classifier Evaluation	33
Said-Huntington Index	35
Color-codes for Conflict & Nation-State Terms	37
Scorecard	37
Conflict Map	38
Non-Conflict Map	39
Importance Visualization	40
Conversations	41
Dashboard View	44
Tool Evaluation Tasks and Results	45
Task 1	45
Task 1 Results	46
Task 2	49
Task 2 Results	49
Task 3	51
Task 3 Results	51
Task 4	54
Results Task 4	54
Task 5	56
Results Task 5	56
Task 6	58
Task 6 Results	58

Insights & Findings	59
Coverage of Islam	59
The Times of India Example	61
Analysis of Conversation 1	62
Analysis of Conversation 2	63
Analysis of Conversation 3	64
Analysis of Conversation 4	65
Analysis of Conversation 5	66
Discussion	68
Limitations & Future Directions	70
Bibliography	71
Appendix	75
User Evaluations Survey/Questionnaire	75
Topics/Conversations Questions	76
Hypothesis Questions	100
User Interface Questions	101
User Classification of Articles Form	106

Table of Figures

Figure 1 Tool Framework.....	18
Figure 2 Hypothesis Framework.....	20
Figure 3 Classifier Selection Model.....	34
Figure 4 Sample Run of the Classifier Selection Model.....	35
Figure 5 Conflict Scorecard.....	37
Figure 6 Conflict Map.....	38
Figure 7 Non-Conflict Map	39
Figure 8 ‘Importance’ Visualization (Guardian)	40
Figure 9 ‘Importance’ Visualization (Washington Post)	41
Figure 10 “Probabilistic Topic Models” (Blei, 2012, p. 78)	42
Figure 11 Conversation Example	43
Figure 12 Dashboard View of Conversations.....	44
Figure 13 User Classification of Articles.....	45
Figure 14 User Classification Results	47
Figure 15 Kappa Interpretation.....	47
Figure 16 Classifier Performance.....	48
Figure 17 Pairwise User Agreement on Conflict/Non-Conflict Maps Classification.....	49
Figure 18 Marginal Distribution of Conflict/Non-Conflict Maps Classification	50
Figure 19 Topic Coherence Results.....	51
Figure 20 Marginal Distribution of Topic Coherence	52
Figure 21 Topic Classification Ratings.....	52
Figure 22 Marginal Distribution of Conversations Agreement.....	53
Figure 23 Feedback on Tool’s Practical Usability in Different Settings.....	58
Figure 24 Conflict Scorecard Results.....	59
Figure 25 Times of India Conversations.....	61
Figure 26 Times of India Conversation 1	62
Figure 27 Times of India Conversation 2	63
Figure 28 Times of India Conversation 3	64
Figure 29 Times of India Conversation 4.....	66
Figure 30 Times of India Conversation 5.....	67

Abbreviations

SVM	Support Vector Machine
LDA	Latent Dirichlet Allocation
NLP	Natural Language Processing
SiNLP	Simple NLP
NLTK	Natural Language Tool Kit
TF-IDF	Term Frequency–Inverse Document Frequency
CDA	Critical Discourse Analysis
CL	Corpus Linguistics
ML	Machine Learning
OED	Oxford English Dictionary

Problem

Since the advent of the Internet, the rate and magnitude of text production by humans has massively increased, as measured in the form of documents drawn from the Web counterparts of newspapers, RSS feeds, social media sites, blogs and other online media. As the scale of text production has grown it has become increasingly difficult to make sense of this mass of data without employing computational data mining methods. Alongside the issue of massive quantities of available text data, the nature of insights required from text has also changed. Simple results, such as counts of query hits or other superficial descriptive summaries of text data, are no longer sufficient for understanding complex large-scale text datasets. Instead, the demand for more semantically rich insights from text analysis has dramatically grown.

An important area in text analysis is discourse analysis. Discourse is loosely defined as a way of speaking about an issue. Discourse analysis is defined in a variety of different ways. For instance, it has been defined as “the study of language in use” (Grenoble, 2006, p. 1). Another definition describes it as “a variety of procedures for examining chunks of language, whether spoken or written (Allwright and Bailey, 1991, p. 61).”

Foucault described discourse as “the general domain of all statements sometimes as an individualizable group of statements, and sometimes as a regulated practice that accounts for a number of statements” (Foucault, 1972, p. 90). He treats discourses as “practices that systematically form the objects of which they speak” (Foucault, 1972, p. 49). What this means is that how we speak about things is linked with our societal power relations. More broadly,

discourse shapes how we interpret our world. It constructs objects that we use to understand and define our world. This is why understanding of discourse is important because discourse results in the formation of stereotypes that contribute to systemic inequality, discrimination and injustice. More importantly, discourse drives policy measures and maintains power relations through policies and legislation (Campbell, 2002; Elmore, 2009; Foucault and Bouchard, 1980).

At the core of discourse analysis lie two questions: 1) how are topics, ideas and issues represented and organized in these texts? 2) And how do we ‘speak’ about certain issues within the context of representation? The first question deals with how texts are organized in topical structures, while the second question is linked with language associated with these structures.

For social science and journalism, discourse analysis is typically employed for the purpose of understanding the context and content of discussion around an issue or a set of issues. As computational methods, topic modeling and text classification come closest to enabling a quantitative approach toward discourse analysis of texts on a scale previously unimaginable in social science and in journalism.

In social science and journalism the structure of text collections can often grow and change as a matter of course. Consider the example of news content, where novel events are covered on a continual basis. It is a continually shifting landscape of events and language related to these events. Though a broad organization of topics may exist (e.g. distinction between health and sports news), at a more granular level the array of topics covered continuously change and adapt to the news cycle. For this reason, the development of tools and methods to facilitate agile, on-

demand understanding of text structure in news text corpora are of immense importance. More important, though, is understanding the ideological content that resides in the text. For this reason, a ‘critical’ approach is needed in conducting text analysis. Typically, text analysis tools uncover patterns and trends in the data without uncovering the ‘ideology’¹ that is embedded in the text. In doing so, they conceal the function of the relation of ‘what is being said’ to its social, and, more importantly, political context.

Critical Discourse Analysis is “fundamentally concerned with analyzing opaque as well as transparent structural relationships of dominance, discrimination, power and control as manifested in language” (Wodak, 2001, p. 2). The field of CDA has been criticized for over-reliance on small amounts of data, whose results are not generalizable (Baker, 2012; Stubbs, 1997). However, Mautner (2009) argues that corpus linguistics techniques—which deal with how language is generated in large amounts of data (Oostdijk, 1991)—could be harnessed to aid CDA. He cites the advantages of using corpus linguistics in CDA as:

helping to give analysts an initial focus in their data, lending credence to generalizations, and counteracting the criticism that researchers within CDA supposedly ‘cherry-pick’ small amounts of data that support pre-conceived ideologies. (Baker et al., 2013a, p. 258).

Furthermore, Fairclough (2013, p. 45) notes that:

the hidden power of media discourse ... [depends] on systematic tendencies in news reporting and other media activities. A single text on its own is quite insignificant: the

¹ Terry Eagleton defines ideology as “the ways in which what we say and believe connects with the power-structure

effects of media power are cumulative, working through the repetition of particular ways of handling causality and agency, particular ways of positioning the reader, and so forth.

Hence, corpus mining becomes essential for understanding the ‘cumulative effects’ of the media discourse. This is why there is considerable opportunity to build a computational² tool that utilizes CDA and corpus linguistics to interpret discourse on a continual basis.

The method of using computational tool for analyzing discourse that is presented in this thesis draws inspiration from the idea of discourse as a system of representation that is revealed through interaction over time— which is particularly relevant for news and media texts. Structurally, an important aspect of discourse is the iteration and accumulation of lexical content in the text i.e. how certain lexical representations are repeated across large collections of texts. For instance, by iteratively associating African-American identity with crime instead of socioeconomic conditions, the public discourse (in news and media) creates a negative image of the African-American identity. Understanding the iterative accumulation of lexical content requires a corpus-based approach.

The emergence of new data mining approaches has now made it possible to extract discourse structures in media contexts at scales that would have been inconceivable to earlier scholars. Along with basic corpus-based methodologies, like keyness,³ collocation and concordance, we can use sophisticated big data approaches to analyze media corpora. These approaches— such as topic modeling, network analysis and sentiment classification—can help us identify, demystify

² Computational linguistics, which deals with the computational models of language, overlaps significantly with corpus linguistics; however, the main objective of corpus linguistics is to investigate the nature of language in large amounts of text. Hence, I am subsuming computational linguistics under corpus linguistics.

³ Relative importance of a word in document context

and frame dominant discourses and ideologies that do not simply represent reality but bring into being stereotypes that negatively effect social subjectivities and relations.

Proposed Work

This thesis aims to combine CDA approaches with corpus linguistics (Figure 1) using machine learning and natural language processing (NLP) techniques. The objective of CDA approaches is to make more visible the hidden aspects of discourse by looking at the latent social ideologies that permeate social texts. Hence, CDA approaches are hypothesis-driven analyses. In contrast, corpus linguistics is an agnostic way of studying language patterns in large amounts of text. My thesis combines both approaches, i.e. CDA and corpus linguistics, by using analytical framework of CDA to create hypothesis-driven features, metrics and design requirements for a corpus analysis tool. The tool, which could be used in both newsroom and academic settings, uses corpus linguistics techniques to analyze and machine-classify patterns of discourse expressed in the media text corpora.

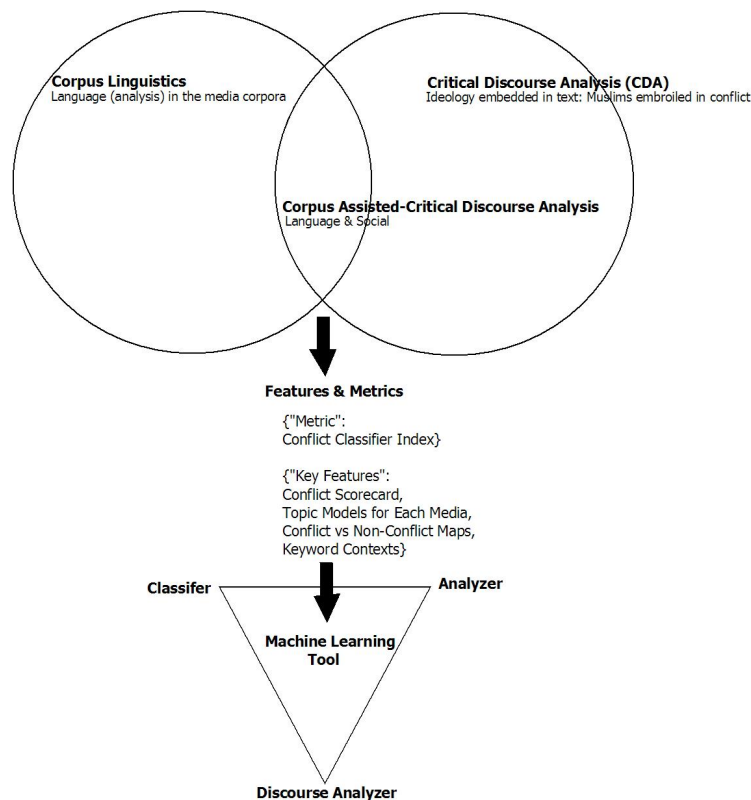


Figure 1 Tool Framework

Said-Huntington Discourse Analyzer

As an instance of this technological framework, I have developed a tool for understanding and defining the discourse on Islam in the global media. My motivation for developing this technological framework comes from the work of Edward Said and Samuel P. Huntington. Said was a notable theoretician in the fields of post-colonialism and critical theory. Huntington was an influential political scientist who gained worldwide prominence through his “clash of civilizations” thesis. One of the main arguments of Huntington was that the world of Islam is in civilizational conflict with the West (Huntington, 1993).

In his response to Huntington’s clash of civilization, Said argues that “unedifying labels like Islam and the West” are misleading, and ignorance and conflict are the crops of these labels (Said, 2001, p. 192). Said (2008, p. xi) further argues that the relationship between Islam and American media is characterized by “highly exaggerated stereotyping and belligerent hostility.” Similarly, Abrahamian (2003) argues that the American mainstream media framed 9/11 under the trope of Samuel Huntington’s “clash of civilizations” (Huntington, 1993), which has contributed significantly to the development and deepening of fault lines between Muslim cultures in Middle East and North Africa and European and North American societies.

How is Islam represented linguistically in the global media sources? Are there any significant differences in the ways Islam is represented in different media sources? These are some of the questions that I have used to develop a framework for critical discourse analysis.

The following four hypotheses inform our critical framework:

- Hypothesis 1: The 9/11 has influenced/and influences the framing of discourse on Islam in America (Powell, 2011).
- Hypothesis 2: The mainstream Western media discourse largely portrays Muslims as “a homogeneous population embroiled in conflict, either as aggressors or victims—most frequently the former” (Baker et al., 2013b, p. 262).
- Hypothesis 3: Muslims are portrayed as the “Other” through the binary lens of “us and them,” framed within Huntington’s “clash of civilization” paradigm (Abrahamian, 2003).
- Hypothesis 4: Muslims are connected with terrorism more strongly than other religious communities (Morey and Yaqin, 2011).

In particular, our critical framework (Figure 2) is based on the second hypothesis, which sees media representations of Muslims contextualized largely around conflict. These media representations are ‘ideological’ constructs because these representations see events related to the Muslim world through the lens of conflict. The machine learning tool “Said-Huntington Discourse Analyzer”— which I have developed as part of this framework— classifies media discourse on Islam in terms of conflict and non-conflict, and attempts to put light on the ‘ideology’ that permeates the text. As part of the user evaluation, we⁴ test whether our hypothesis conforms to the insights extracted from the tool.

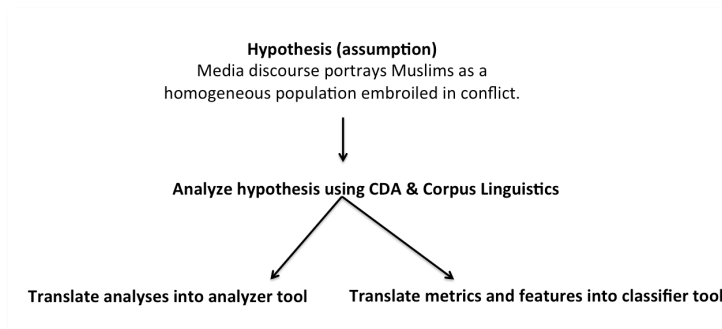


Figure 2 Hypothesis Framework

⁴ In this thesis ‘we’ refers to I (the author) when referring to tasks/actions related to the thesis.

Related Work

There has been an explosion of activity in the area of sentiment analysis (Pang and Lee, 2008) and knowledge discovery (Awad and Khanna, 2015) using machine learning in recent years.

Currently, there are many tools available that employ natural language processing (NLP) to extract text patterns in machine-readable data sources. Simple NLP (SiNLP) is a general text analysis tool that delineates linguistic features for analyzing text (Crossley et al., 2014). Coh-Metrix tool allows users to analyze texts on the basis of comprehension (Graesser et al., 2014). Sketch Engine (Kilgarriff et al., 2004) is a corpus query engine that allows users to study word patterns in text corpora.

Researchers use tools like Sketch Engine to study and analyze media texts. Baker et al. (2013b) use methods from Critical Discourse Analysis and corpus linguistics to analyze representations around the word ‘Muslim’ in a corpus of British newspaper articles published between 1998 and 2009. The study reports that Muslims are largely seen as a homogenous group associated with conflict. This particular work builds its analysis on the basis of collocates, which are words that occur near a given word. It uses Sketch Engine to analyze noun collocates of the word ‘Muslim’ and groups them under categories like conflict, culture and religion. However, one problem with this approach is that it does not take into account broader patterns within the content since it focuses on the locus of collocates rather than the entire text.

More sophisticated data mining approaches have been used to classify discourse in some studies. For instance, Montiel et al. (2014) investigated domestic media's production of nationalism during the China-Philippines Maritime Dispute. Their findings show how international conflict is linguistically created in the media ecosystem. In addition, they demonstrate how mathematical models could be used to classify sets of words belonging to a particular newspaper involved in reporting of the dispute.

Researchers use Media Cloud—a large-scale content analysis system developed by the Berkman Center at Harvard and the MIT Center for Civic Media—to study media discourse. Goncalves (2014) uses cluster analysis on a corpus of news articles—that are investigated using Media Cloud—to analyze frames in the controversy surrounding rolezinhos flashmobs organized by low-income youth in Brazilian shopping malls. Graeff et al. (2014) curates stories using Media Cloud to build a media controversy map of the killing of Trayvon Martin in 2012.

One problem in classification of complex discourses pertains to the availability of suitable corpora. In classification of non-standard categories—such as sarcasm, humor or a category that one defines in a specific manner (in our case conflict)—annotated categorized corpora are not readily available; hence, one has to build the corpus first for applying machine learning techniques. Usually it is difficult to achieve high inter-user agreement on classification of categories that are not straightforward to interpret. For example, González-Ibáñez et al. (2011) use machine learning techniques and human judges to detect sarcasm in a Twitter corpus. It reports that neither human judges nor machine learning classifiers perform well on

detecting sarcasm. More importantly, they reported that an agreement of 50% was achieved among three users on the task of sarcasm identification.

There are multiple studies that focus on the issue of conflict from the point of view of information retrieval and machine learning. King and Lowe (2003) reported that in the extraction of conflict data from Reuters news reports, the computer program performed equally to trained Harvard undergraduates. Media analysis has also been used to ‘forecast’ and ‘predict’ conflict. Hunt (1996) provides a methodology for using mass media indicators to predict international conflict. Similarly, Leetaru (2013) uses advanced machine learning techniques to forecast conflict using media text corpora.

Implementation & Methods

Media Cloud

Our tool builds on the work of Media Cloud. Developed by the Berkman Center at Harvard and the MIT Center for Civic Media, Media Cloud is “is an open source, open data platform that allows researchers to answer complex quantitative and qualitative questions about the content of online media”⁵. It curates more than 50,000 global news sources, ranging from mainstream media to alternative media blogs. Although our tool could be used on any kind of machine-readable text source, the primary reason for choosing Media Cloud is to utilize the availability of diverse open-source news corpora that are available in the platform.

Apart from Media Cloud, the tool uses the following toolkits/libraries as part for its implementation: NLTK (Natural Language Tool Kit) for natural language processing (Loper and Bird, 2002); Scikit-learn (Pedregosa et al., 2011) and Gensim (Řehůřek and Sojka, 2010) for machine learning; and D3 for visualizations (Zhu, 2013).

The tool extracts news stories using Media Cloud API from the following news sources:

- Al Jazeera
- Times of India
- The New York Times
- Daily Mail Online
- USA Today
- Washington Post
- Dawn
- Reuters
- Fox News

⁵ <http://mediacloud.org/about/>

- The News International
- MSNBC

The aforementioned news sources are diverse and cover news sources from both majority Muslim countries and countries where Muslims are a minority. Our news sources comprise three sources from Islamic countries, namely Al Jazeera, The News and Dawn; one source from India, namely Times of India; three sources from England, namely Daily Mail, Guardian and Reuters; and four news sources from the United States, namely New York Times, Washington Post, USA today and Fox News.

Query for Retrieving Media Articles on Islam

The topic of Islam is represented in the media sources by a variety of keywords. I modified the query developed by Baker (2010) to extract a wide range of Islam-related topics. Baker (2010) compared British broadsheet and tabloid newspapers to study representations of Islam for the 1999-2005 time period. I crafted the following solr⁶ query for Media Cloud API to extract articles that directly or indirectly deal with Islam:

Query = '("Allah" or "Allahu" or "ayatollah" or "burka*" or "burqa*" or "fatwa*" or "hejab*" or "hijab" or "imam*" or "islam*" or "Koran*" or "Mecca" or "Medina" or "Mohammedan*" or "Moslem*" or "Muslim*" or "mosque*" or "mufti*" or "jihad*" or "jehad" or "mullah*" or "Quran" or "Koran" or "sharia*" or "shia*" or "shi-ite*" or "Shiite" or "sunni*" or "wahabi" or "yashmak*") and not ("shiatsu") and not ("sunnily")'*

The mark “*” acts as the wildcard, and retrieves all keywords that have the keyword stem (or root). For instance, the term “Koran*” retrieves documents containing the words Koran, Koranic and Korans.

⁶ Open source enterprise search platform

Rolling-30-Day Corpus

The tool pulls the data from the Media Cloud using its API on a daily basis. Using “published date” filter, the tool restricts the articles to the latest thirty-day period. Afterward, stop words— or “words of little intrinsic meaning” (Eck et al., 2004, p. 330) like ‘a’, ‘an’ and ‘the’— are removed as part of pre-processing. Using stemming and lemmatization, we conflate different inflected forms of a word so they could be analyzed as a single entity. For instance, the term ‘churches’ is reduced to the root ‘church’ and the term ‘killing’ is reduced to the form ‘kill’. The extracted text of the articles is tokenized and tagged using a part-of-speech tagger, which assigns parts of speech (such as noun, verb or adjective) to each word. The purpose of tagging the data is to extract grammar-based text that is appended to the document tuple:

$$Document_n = \{full\text{-}text, grammar\text{-}based\ text\}$$

The tool’s classifier-building heuristic uses both types of text in its implementation. It uses noun-phrase grammar to extract our grammar-based text because noun phrases are good linguistic markers for: feature engineering (Scott and Matwin, 1999); information retrieval (Smeaton, 1992); and representations of the content (Fagan, 1989).

The tool uses the following list of tags to extract the noun phrases from the articles.

Tags list = ['NN', 'NNS', 'NNP', 'NNPS', 'JJ', 'JJR', 'JJS']⁷

⁷	JJ	Adjective
	JJR	Adjective, comparative
	JJS	Adjective, superlative
	NN	Noun, singular or mass
	NNS	Noun, plural
	NNP	Proper noun, singular

In addition, the tool uses the following rules to generate the grammar for noun phrases:

(Rule1) NBAR = (NN*|JJ*)(NN*)

(Rule2) NBAR IN NBAR

These rules are based on key-phrase extraction proposed by Kim et al. (2010, p. 576).

Defining Conflict

Our hypothesis requires classification of articles on the basis of conflict. Articles are classified as ‘conflict articles’ if they are about groups or people or individuals embroiled in a conflict, either as victims or aggressors. Specifically, conflict in our context means:

- a. An encounter with arms; a fight, battle (OED)⁸.
- b. Fighting, contending with arms, martial strife (OED).
- c. A struggle marked by terrorism or extremism or force or militancy.
- d. Militant struggle defined as opposition among social entities.
- e. Armed conflict

On the other hand, articles are classified as ‘non-conflict articles’ if they are about subjects other than conflict. These subjects could include spirituality, religion, health, art, economy, sports and local politics (in which groups are engaged in non-militant political issues).

⁸ Oxford English Dictionary

Pre-2001 Corpus

Since the advent of 9/11, Muslims have been portrayed in many mainstream media as “a homogeneous population embroiled in conflict, either as aggressors or victims—most frequently the former” (Baker et al., 2013b, p. 262). This implies that the coverage on Islam is largely about conflict issues. In order to machine classify articles on the basis of conflict and non-conflict, we need a good-sized sample of articles about non-conflict topics. One issue with retrieving articles that are not about conflict is that since 9/11 the media coverage has represented Islam largely in conflict terms. In order to overcome this issue, I extracted content from Lexis Nexis using articles published before 2001. I used the following query to retrieve the articles:

```
[(((("Islam" and "health" and not ("war" or "terror" or "extrem!")) or  
("Islam" and "sport!" and not ("war" or "terror" or "extrem!")) or  
("Islam" and "econom!" and not ("war" or "terror" or "extrem!")) or  
("Islam" and "spiritual!" and not ("war" or "terror" or "extrem!")) or  
("Islam" and "business" and not ("war" or "terror" or "extrem!")) or  
("Islam" and "science" and not ("war" or "terror" or "extrem!")) or  
("Islam" and "arts" and not ("war" or "terror" or "extrem!")) or  
("Islam" and "entertainment" and not ("war" or "terror" or "extrem!")) or  
("Islam" and "community" and not ("war" or "terror" or "extrem!")) or  
("Islam" and "religion" and not ("war" or "terror" or "extrem!")) or  
("Islam" and "music" and not ("war" or "terror" or "extrem!")) or  
("Islam" and "women" and not ("war" or "terror" or "extrem!"))))  
and Date(geq(01/01/1999) and leq(03/31/1999)))] (1000)  
Source                      Source Information [Newspaper Stories, Combined Papers]
```

The query construct (“war” or “terror” or “extrem!”) filters all the articles that contain the three aforementioned terms. The exclamation mark after “extrem” truncates the term as a stem and removes all inflections of the term ‘extrem’, such as extremism, extremist etc. I used this filter

construct in conjunction with topics that are not about conflict— such as ‘Islam’ and ‘sport’ or ‘Islam’ and ‘music’—to create a pre-2001 corpus of about 3000⁹ non-conflict articles. I intentionally excluded ‘politics’ and ‘international relations’ as query constructs because conflict, in our context, is primarily associated with the topic of politics.

Conflict Terms Word List

Conflict terms such as ‘force’, ‘military’ and ‘terror’ are good candidates for features representing conflict in a given article or document. A list of conflict terms provides us with a rudimentary criterion for defining content on the basis of conflict. The tool utilizes open source lexical database WordNet to build the conflict terms word list. The basic element in WordNet is a synonym set, or synset, which represents a lexical concept schema, or a set of semantically related words grouped together. The tool uses sysnets for the terms ‘force’, ‘military’, ‘terror’, ‘military’ and ‘extremism’ to create the semantic network of the conflict-related vocabulary for our list.

Classification Methods

In machine learning, support vector machines (SVMs) are supervised machine learning methods developed on the basis of statistical learning theory (Cortes and Vapnik, 1995; Vapnik and Kotz, 1982). In the case of binary classification, the aim of the SVM algorithm is to find the hyperplane that divides the data points belonging to the two classes in a way that maximizes the distance between the decision boundaries. SVMs are suitable for text classification (Joachims,

⁹ The size of pre-2001 corpus is comparable to our daily extract for rolling-30-day corpus

2002), achieve “good classification performance” (Joachims, 2001, p. 136), and are quick to train and evaluate (Dumais et al., 1998). SVMs have been shown to surpass other classification algorithms on classification measures in a number of comparative evaluation studies (Joachims, 1998; Pang et al., 2002; Yang and Liu, 1999).

Apart from SVMs, Naïve Bayes classifiers are a family of probabilistic classifiers that are also popular in text classification. These supervised machine learning algorithms employ Bayes’ theorem with ‘naïve’ assumption of independence between the features. Bayes’ theorem allows us to calculate posterior probabilities for prediction using given prior probabilities. Naïve Bayes classifiers, cited historically for effective use in text classification, are “known for their inherent robustness to noise and their fast learning time” (Maxion and Townsend, 2002, p. 221); and are, therefore, frequently included as a core methodology in comparative evaluation of text classification techniques (Joachims, 1998; Pang et al., 2002; Yang and Liu, 1999).

There are different ways we can set up a Naïve Bayes classifier. In text classification the two commonly employed implementations of Naïve Bayes classifier are multinomial model and multivariate Bernoulli model (Yu, 2008). In the Bernoulli model, one feature represents each word in the document; while in the multinomial model, one feature represents each word position in the document (Meretakis et al., 2000). More specifically, multinomial model takes into account both the order of words and the number of occurrences of the words in a document. Regardless of these differences, both multinomial and Bernoulli models are suitable for classification with discrete features (in our case, word counts for text classification). Previous studies on topic classification tasks have shown that the multivariate Bernoulli model is more

suitable for data sets with small vocabularies, while the multinomial model performs better on larger vocabularies (Lewis, 1998; McCallum and Nigam, 1998).

In the thesis implementation I developed a heuristic that combines Naïve Bayes and SVM algorithms with different feature options in order to find the best performing classifier in terms of classification accuracy. Feature engineering is an integral part of machine learning and “there is ... no replacement for the smarts you put into feature engineering” (Domingos, 2012, p. 84). In our case, we produce the classifier by employing a variety of feature options.

The purpose of tagging the data is to extract grammar-based text that is appended to the document tuple:

$$Document_n = \{full\text{-}text, grammar\text{-}based\ text\}$$

Our classifier-building heuristic uses both types of text in its implementation.

Our goal is to build the classifier through a heuristic implementation. We use the pre-2001 corpus to extract text documents that do not contain any conflict term that we have compiled in our conflict terms word list. Assuming that most of the current media articles largely cover conflict issues, we use pre-2001 corpus for extracting non-conflict corpus in order to have sufficient number of non-conflict training samples for machine learning. We use the current rolling-30-day corpus to find articles that exclude all terms from the conflict term word list and combine them with the pre-2001 corpus text documents that also exclude these terms. Hence, this document list becomes our master list of non-conflict training samples¹⁰. We define the

¹⁰ The ratio of pre-2001 (old) non-conflict articles to rolling-30-day (current) non-conflict articles is typically 1:10. So if we have 100 articles in the current corpus, we will have 1000 articles from the old corpus. I found that there is

conflict articles as articles that contain at least one or more conflict terms. Our goal is to find best classification score for a classifier with the lowest number of conflict terms as part of its feature space. Our intuition behind using the lowest number of conflict terms is that we want to retain a diverse and large feature space for classifying articles that have fewer conflict terms. A higher number of conflict terms will reduce our feature space by restricting the number of documents that are classified as conflict articles. Hence, once the classifier is optimized by cross validation, we stop the classification heuristic. In order to find the best classification score with fewest number of conflict terms, we iterate through the rolling-30-day corpus, increasing the count of conflict terms by one, until we find the maximum classification score. From the labeled text documents (conflict, non-conflict), we extract both grammar-extracted text and full text of the articles. Using natural language processing all text documents are lemmatized, tokenized and stemmed after the removal of stop words.

Feature Engineering

Feature engineering is an integral part of machine learning. Domingos (2012) notes that feature engineering is the most important element in the success or failure of machine learning projects.

Feature engineering is defined as “the process of manipulating and combining features to arrive at more informative ones” (Bowles, 2015, p. 17). In our implementation we use both full-text and grammar-based texts to build our model. The raw text is transformed into numerical feature vectors through the process of vectorization. A vectorizer is a sparse matrix representation of the

a dearth of non-conflict articles in the current corpus for many mainstream media; hence, there is a need to augment our training set using the old data.

document features. Two vectorizer variations are widely used in text classification, namely count vectorizer and term frequency–inverse document frequency (TF-IDF) vectorizer.

The TF-IDF model reflects the relevance of an n -gram in a text document in a collection of documents (Manning et al., 2008; Salton and McGill, 1983). This method informs about the discriminative word feature space for a given document. The TF-IDF model was originally a term weighting scheme developed for information retrieval but has found good use in document classification and clustering.

The count vectorizer tokenizes the text and counts the word occurrences of a given corpus of text documents. We use the vectorizers on two different n -gram¹¹ lexical models. One is the 1-gram (individual words); the other is the 2-gram model of words. The 2-gram model preserves some of the local ordering information in the text documents. The vocabulary extracted by this 2-gram vectorizer is hence much bigger and usually resolves ambiguities encoded in local positioning configurations. Both vectorizers produce a sparse label-feature matrix, where each row stands for a given $k = 2$ labels (conflict, non-conflict) and n columns for each sorted feature vector.

Classifier Evaluation

We use two cross validation metrics to evaluate our classifier. First, we randomly split the documents into random and training sets. Then we sample the training set while holding out 30%

¹¹ In computational linguistics, a word “ n -gram is a contiguous sequence of n items [words] from a given sequence of text or speech” (Navarro-Galindo et al., 2012, p. 41). For example, for the text phrase “thus spake zarathustra”, we have the following 1-gram and 2-gram representations:

1-gram = ['thus', 'spake', 'zarathustra']

2-gram = ['thus', 'spake', 'zarathustra', 'thus spake', 'spake zarathustra']

of the data for testing (evaluating) our classifier. We compute the F1 score for the test data, which becomes our cross validation (70:30) score. Secondly, we use the k-fold cross-validation score by splitting the data, fitting a model and computing the score five consecutive times (with different splits each time). We calculate the mean score from the five k-fold iterations, along with the 95% confidence interval of the score estimate from the k-fold cross validation data.

Our classifier selection method is as follows (Figure 3):

```
For "number of conflicts terms" (i) in range (0 and Length of Conflict Terms Word List):  
  Create non-conflict documents corpus with 0 conflict terms from the pre-2001 dataset  
  Create non-conflict documents corpus with 0 conflict terms from the current rolling-30-day corpus  
  Combine the two lists as final master non-conflict corpus  
  Create master conflict documents corpus with conflict terms greater than "number of conflicts terms" (i)  
  Combine conflict and non-conflict corpora into master corpus (data)  
  For text type in (grammar-based text, full text):  
    For vectorizers in (count vectorizer, TF-IDF vectorizer):  
      For ngrams in (1-gram, 2-gram):  
        For classifiers in (SVM, Multinomial, Bernoulli):  
          Vectorize the data  
          Break the list into test and train set  
          Calculate score for 70/30 split  
          Calculate k-fold cross validation score  
Pick the classifier with highest 70/30 or k-fold validation score with lowest "number of conflicts terms"
```

Figure 3 Classifier Selection Model

A sample run of the classifier selection model is shown in the following table (Figure 4):

<i>Conflict Terms</i>	<i>Vectorizer</i>	<i>Classifier</i>	<i>70/30 Score</i>	<i>fold mean score (95%)</i>
9	<i>TFIDF</i>	<i>SVM</i>	99.25	"99.33 (+/- 0.30)"
9	<i>TFIDF</i>	<i>MultinomialNB</i>	90.05	"87.47 (+/- 6.00)"
9	<i>TFIDF</i>	<i>BernoulliNB</i>	99.25	"98.21 (+/- 0.78)"
9	<i>Count</i>	<i>SVM</i>	99.00	"99.55 (+/- 0.18)"
9	<i>Count</i>	<i>MultinomialNB</i>	94.53	"93.44 (+/- 2.31)"
9	<i>Count</i>	<i>BernoulliNB</i>	99.25	"98.21 (+/- 0.78)"
10	<i>TFIDF</i>	<i>SVM</i>	99.73	"99.35 (+/- 0.31)"
10	<i>TFIDF</i>	<i>MultinomialNB</i>	90.22	"86.94 (+/- 6.29)"
10	<i>TFIDF</i>	<i>BernoulliNB</i>	99.46	"99.02 (+/- 0.61)"
10	<i>Count</i>	<i>SVM</i>	100.00	"99.84 (+/- 0.10)"
10	<i>Count</i>	<i>MultinomialNB</i>	95.11	"93.47 (+/- 2.20)"
10	<i>Count</i>	<i>BernoulliNB</i>	99.46	"99.02 (+/- 0.61)"
11	<i>TFIDF</i>	<i>SVM</i>	99.70	"99.73 (+/- 0.18)"

Figure 4 Sample Run of the Classifier Selection Model

The table lists classification scores for the three classifiers with different feature sets. The sample data is generated from 1-gram, grammar-based text. The best classifier is highlighted in boldface. As the results suggest, the best classifier is SVM (grammar-based text, 1-gram, conflict terms = 10, Count vectorizer). The best performing classifier is serialized and is used to classify (conflict, non-conflict) articles in the rolling-30-day corpus.

Said-Huntington Index

Our index is a composite number comprising three different metrics. The probability estimate of our classifier is the first metric; the relative weight of this metric in our index is 50 percent. The second metric, with a relative weight of 30 percent, is the number of conflict terms in an article. The third metric, with a relative weight of 20 percent, is the number of terms denoting the nation-states in a given article. The goal of Said-Huntington index is to measure the degree of conflict in an article. The machine learning classifier has the highest weight (50%) because

it is the classifier that a priori determines whether a document has conflict or not. Since we define conflict as one that is characterized by force and militancy, we have assigned 30% weight to the conflict terms that typify force and militancy; in doing so, we are giving more weight to the conflict characterized by these terms in the index. The presence of terms that denote a nation-state may indicate the likelihood of a conflict because nation-states are at the root of modern conflict (Wimmer and Min, 2006). Hence, we have assigned 20% weight to terms that denote nation-state.

The Said-Huntington index is calculated as follows:

$$\text{Said Huntington Index} = 50 \times P + \text{Minimum} \left(30, 30 \times \frac{x}{10} \right) + \text{Minimum} \left(20, 20 \times \frac{y}{10} \right)$$

P = Probability score of conflict classifier

x = number of conflict terms in an article

y = number of terms denoting nation/state in an article

A score of greater than or equal to 50 classifies the article as ‘Huntington’, whereas a score of less than 50 classifies the article as ‘Said’. For example, an article is classified as 90% conflict by our supervised machine learning classifier. It contains four conflict terms and two terms that denote nation-states. The Said-Huntington index for this article is calculated as follows:

$$50 \times 0.90 + 30 \times 4/10 + 20 \times 2/10 = 45 + 12 + 4 = 61$$

An index value of 100 indicates extreme conflict in the article, while a value of 0 indicates the absence of conflict in the article.

Color-codes for Conflict & Nation-State Terms

In any visual representation in our tool, if the keyword is a conflict term, i.e. it is part of the conflict terms word list, we color it as red; on the other hand, if the keyword denotes a nation-state, we color it as orange. This makes it easier for user to comprehend discourse structure.

Scorecard

Using the Said-Huntington index classification, we aggregate the results for all the media sources in our corpus. We use color-coded liquid fill gauge visualization (Figure 5) to indicate the percentage of articles that are about conflict in a given source.

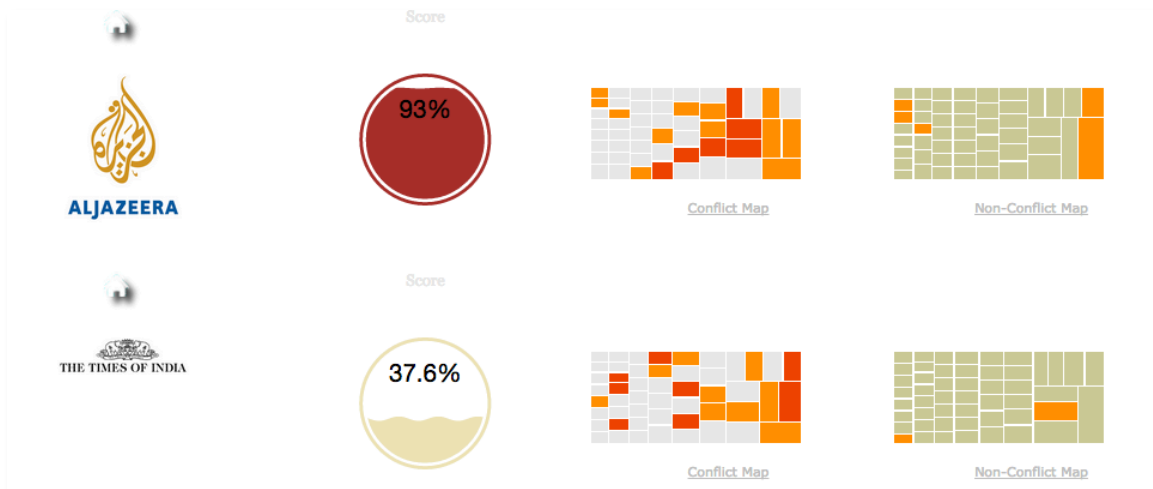


Figure 5 Conflict Scorecard

Conflict Map

The 'Conflict Map' visualization page (Figure 6) shows the treemap¹² of 50 most frequent terms from the articles that are classified in the 'conflict' category using the Said-Huntington index. In addition, it also provides a sample of articles that are classified under the conflict category.

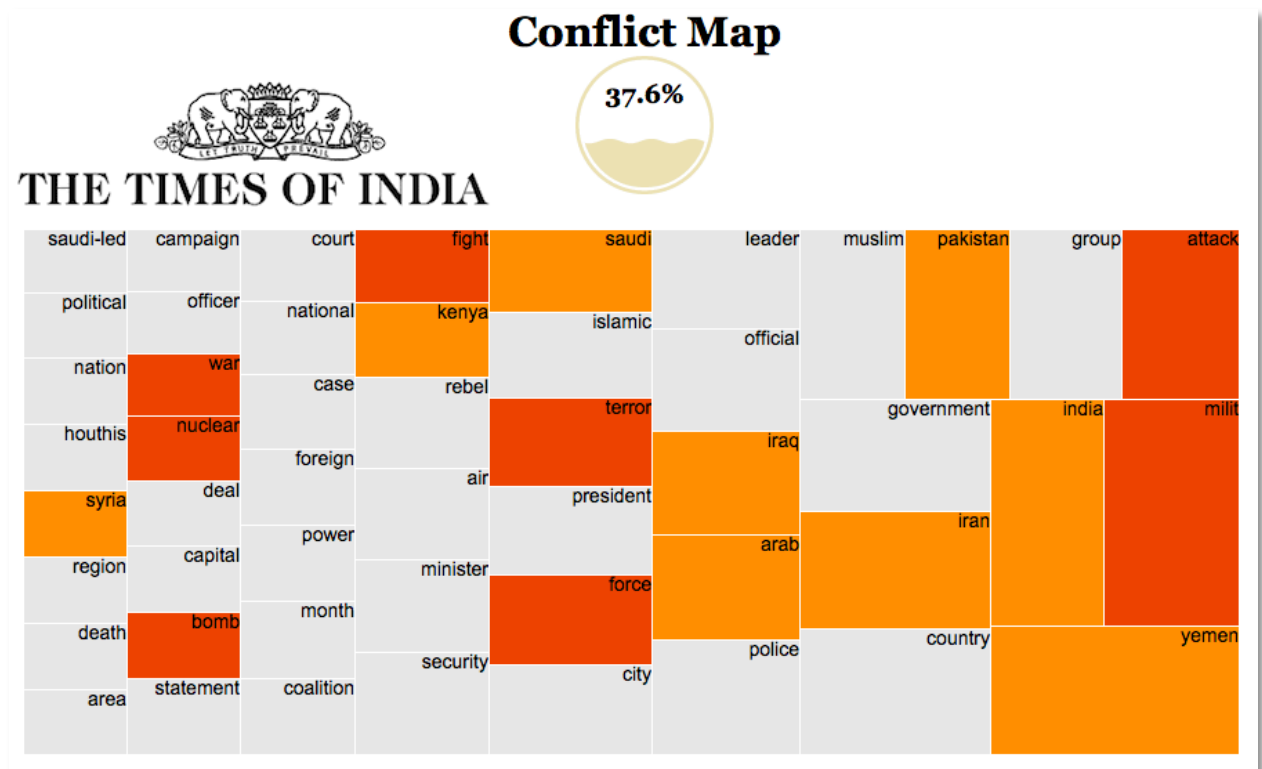


Figure 6 Conflict Map

¹² Treemapping is a method for displaying hierarchical data by using nested rectangles.

Non-Conflict Map

The 'Non Conflict-Map' visualization page (Figure 7) shows the treemap of 50 most frequent terms from the articles that are classified in the 'non-conflict' category using the Said-Huntington index. In addition, it samples articles that are classified under the non-conflict category.

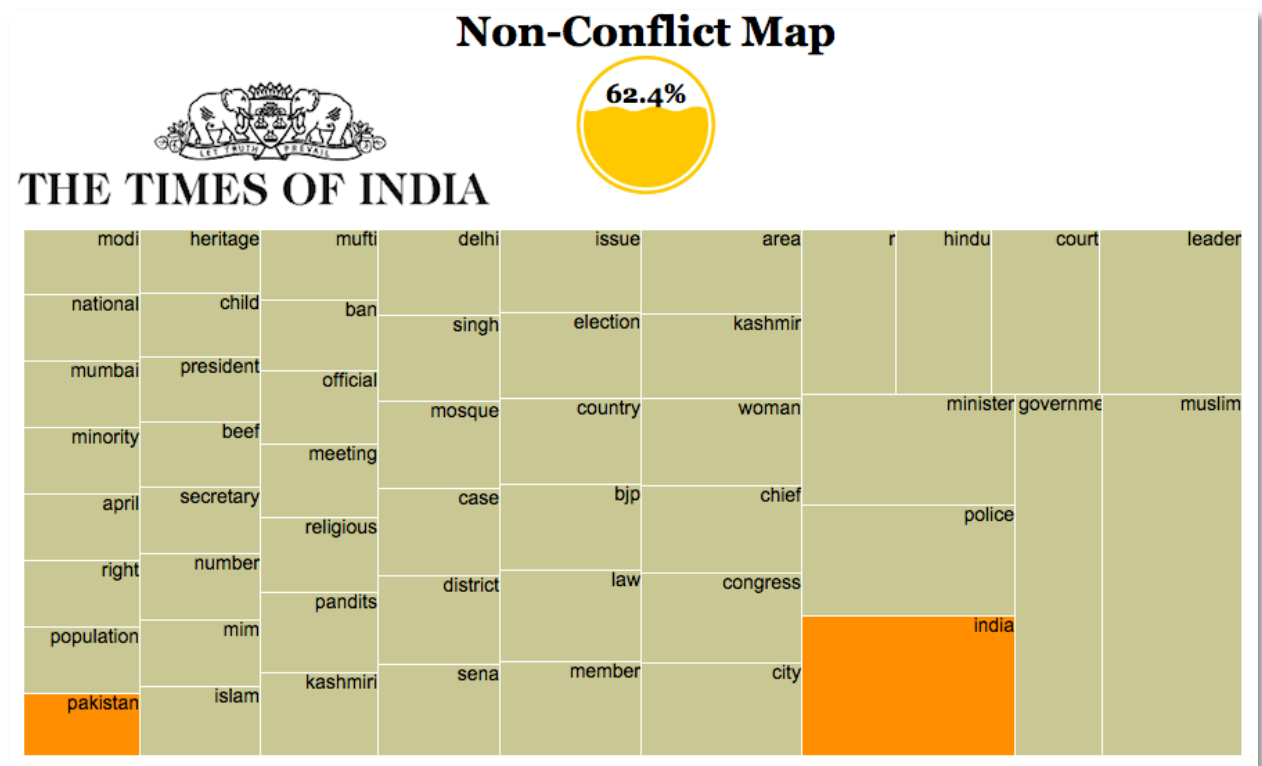


Figure 7 Non-Conflict Map

Importance Visualization

For a given media source, we also provide the user with the ‘importance’ visualization, which shows which terms are more rare and important in the given source compared with other sources. This allows the user to differentiate one media source from other sources in terms of journalistic coverage priorities. This is implemented by using a TF-IDF weighting on the 50-most-frequent terms within a media corpus. For instance, the Guardian ‘importance’ map (Figure 8) shows that—in its corpus of articles on Islam—the terms ‘britain’, ‘scotland’, ‘labour’ and ‘london’ were more important to the Guardian than to other media sources. In contrast, the Washington Post ‘importance’ map (Figure 9) shows that the terms ‘sanction’, ‘israel’, ‘usa’ and ‘framework’ were more important to the Washington Post than to other media sources. In conjunction with ‘importance’ visualization, we also provide a treemap of the most frequent terms for a given media source.

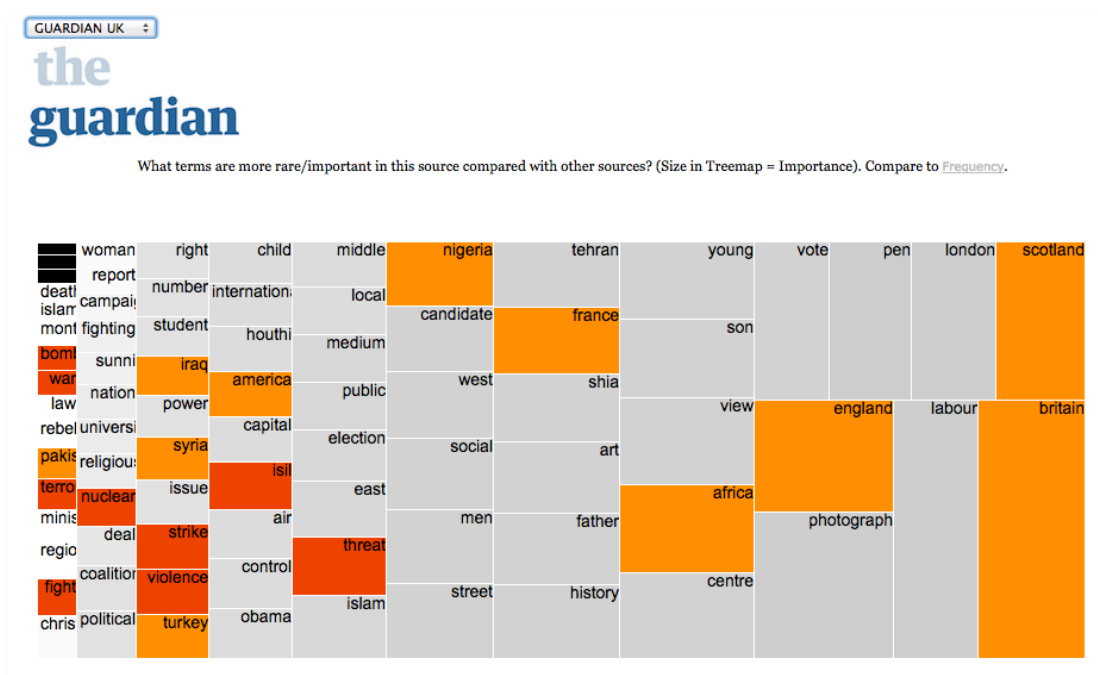


Figure 8 ‘Importance’ Visualization (Guardian)

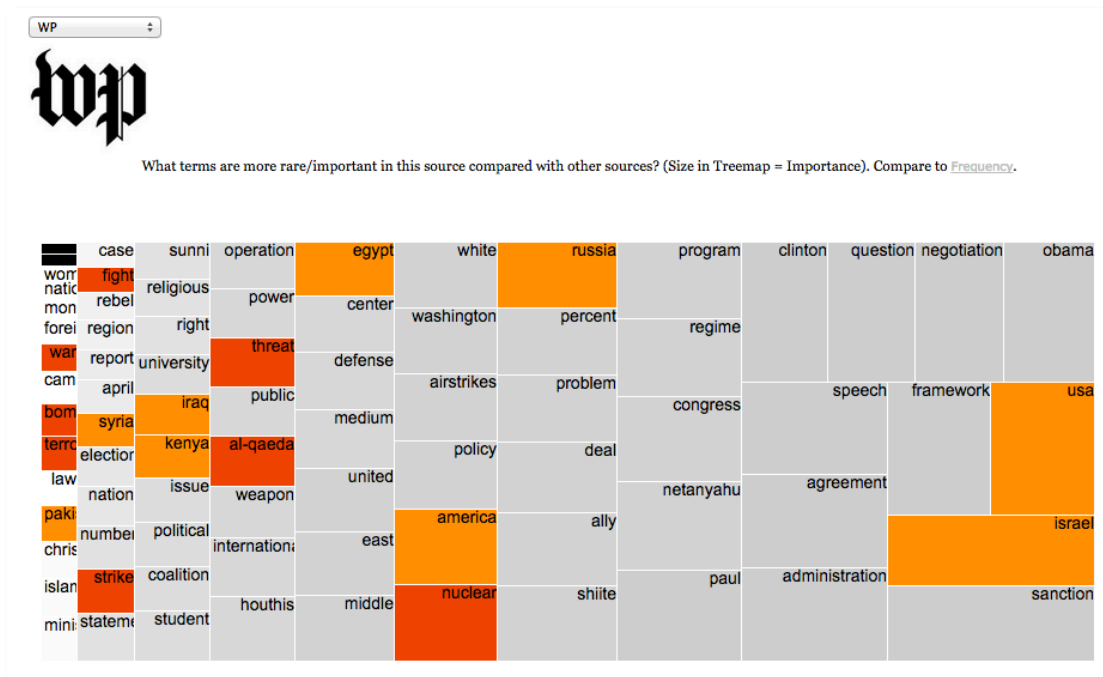


Figure 9 ‘Importance’ Visualization (Washington Post)

Conversations

In order to understand the latent structure of the discourse on Islam, we employ unsupervised machine learning technique based on probabilistic topic modeling. Past studies indicate that probabilistic topic modeling is useful “in exploring and predicting the underlying semantic structure of documents” (Du et al., 2010, p. 148). We extract five topics and corresponding keywords using the Latent Dirichlet Allocation (LDA) model for all the media sources.

LDA is a generative probabilistic model that is used to extract the latent topics present in a text corpus (Blei et al., 2003). It is “useful for finding related documents as well as visualizing the topical content of each document”(Asuncion et al., 2010, p. 98). The intuition behind LDA

algorithm (see Figure 10) is that documents exhibit multiple topics. For instance, words like ‘gene’, ‘dna’ and ‘genetic’ will cluster together because they are part of the same topic¹³ (Blei, 2012). On the other hand, words like ‘data’, ‘number’ and ‘computer’ will tend to cluster together since words about data analysis are part of the same subject.

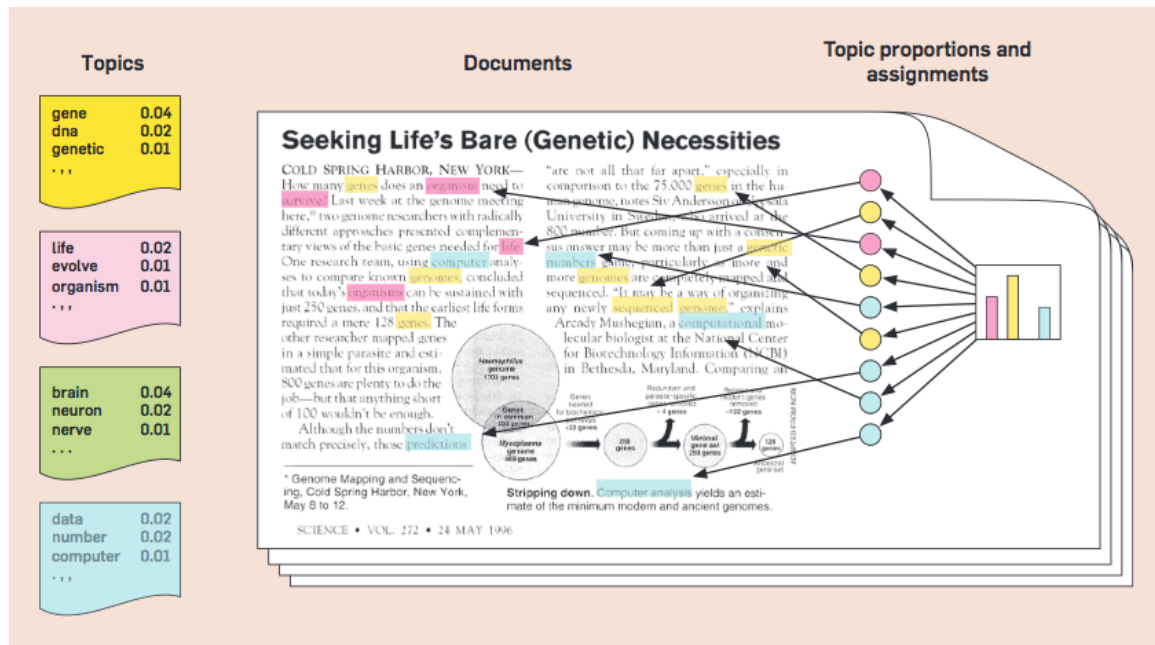


Figure 10 “Probabilistic Topic Models” (Blei, 2012, p. 78)

Our methodology aims to extract semantically rich insights that are organized in a topical manner. This is particularly relevant for understanding text as a discourse or as a form of conversation. At the core of this discourse analysis lie two questions: 1) how are different conversations represented in these texts? And 2) what is the nature of discussion about a certain topic? The extraction of topics shows how the general topic of “Islam” is represented in a given media corpus. It shows how the conversation on “Islam” takes place in that media and in what

¹³ A topic is defined as a “distribution over a fixed vocabulary” (Blei, 2012, p. 78).

context. The advantage of using this approach is that it brings into front the underlying representation of the discourse structure by delineating different topical spaces within a given media corpus.

A conversation is a cluster of statistically/semantically related core terms grouped by LDA. Each conversation is represented using a bubble frequency chart (See Figure 11).

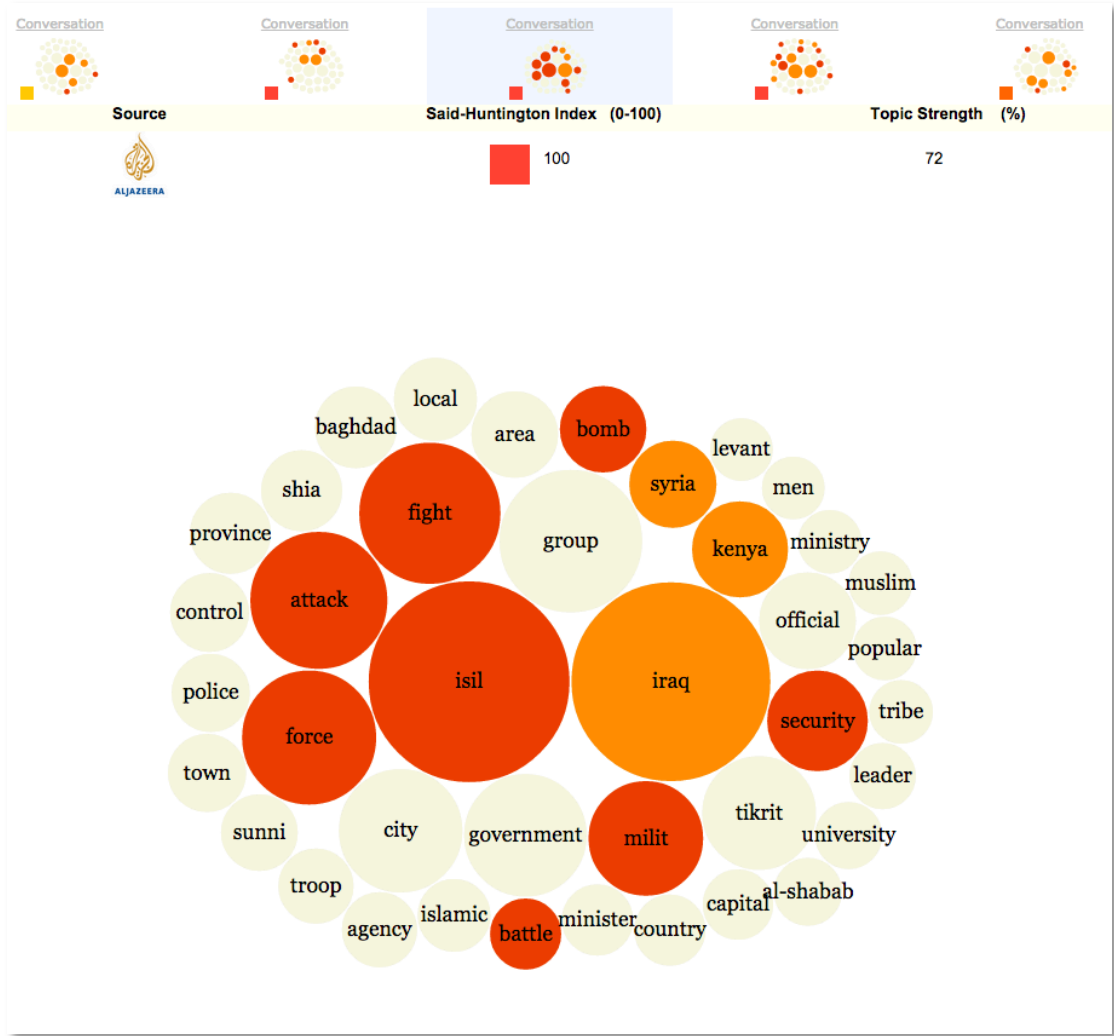


Figure 11 Conversation Example

The higher the frequency of a keyword in a topical cluster, the bigger its size in the bubble frequency map. In order to determine whether the conversation is related to conflict, we classify the conversation using the Said-Huntington index—which is represented by a red-green (spectrum) color-coded square box and the index value. The red color indicates extreme conflict, whereas green color indicates absence of conflict. The color-coded visualization makes it easier for the user to discern conversations based on classification index. We also provide the relative ‘topic strength’ of the conversation as compared with other conversations. The relative strength of each topic is extracted using the following equation:

$$\text{Topic Strength} = \frac{\text{frequency of most frequent term in the conversation for a given media}}{\text{frequency of most frequent term in all the conversations for a given media}}$$

Dashboard View

In our dashboard view, we provide a comparative visual summary of all the conversations using thumbnail versions of bubble frequency charts (Figure 12).

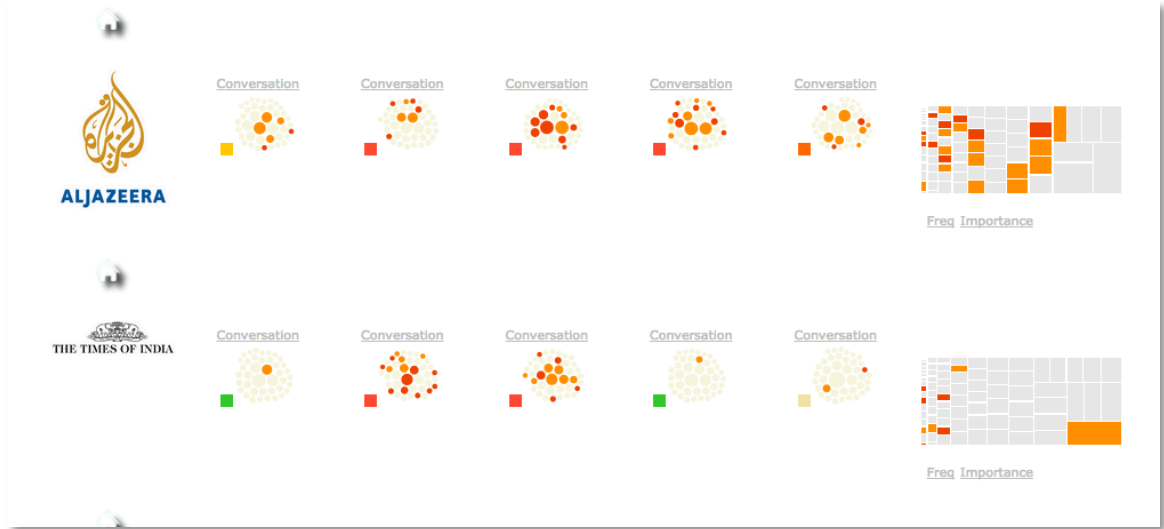


Figure 12 Dashboard View of Conversations

Tool Evaluation Tasks and Results

We¹⁴ designed a series of tasks¹⁵ to evaluate the tool using quantitative, qualitative and statistical methods. For our evaluation experiment we recruited 5 users, two of them were Muslims, while three were non-Muslims.

Task 1

We retrieved 103 articles (see Appendix) on Islam using our tool's Media Cloud query. We applied the published date range filter to select articles from a date range that was from recent past but did not overlap the latest 30-day-rolling period when the experiment was conducted¹⁶. We curated the articles on a web form and asked the users to classify the articles based on our definition of the term conflict. We provided three classification options to the user, namely 'Conflict', 'Non-Conflict' and 'Not-sure' (Figure 13).

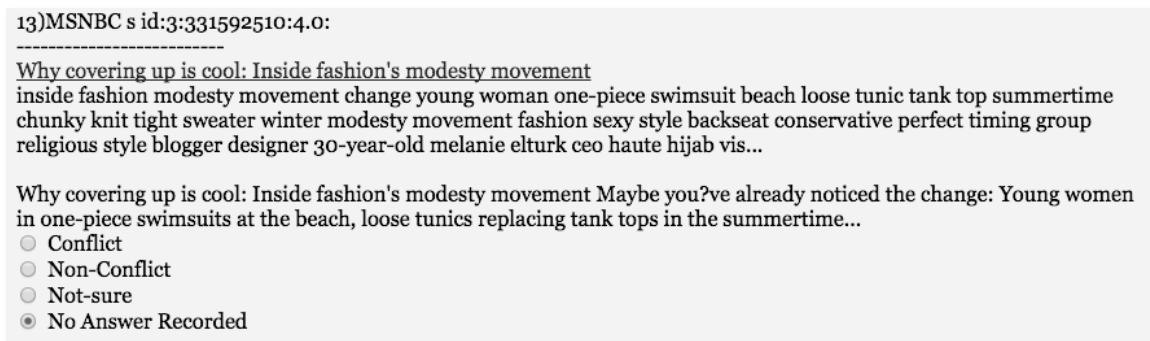


Figure 13 User Classification of Articles

¹⁴ In this thesis 'we' refers to I (the author) when referring to tasks/actions related to the thesis.

¹⁵ See Appendix for details on user evaluation.

¹⁶ The user evaluations are based on the daily run of the tool executed on 04/28/2015.

After the users completed their task, we ran our machine learning classifier to classify this corpus of 103 documents. The goal of this task was to calculate the degree of inter-rater agreement among users on our classification criteria—i.e. what we mean by conflict. Additionally, we wanted to measure the performance of our supervised machine learning classifier—in terms of precision, recall, specificity and accuracy— using the classification labels provided by the human users.

The experimental evaluation of the task was based on the following questions:

- a) Did the users agree on our definition of the term ‘conflict’?
- b) And whether there is agreement between machine classification and human classification results?

Task 1 Results

We calculated¹⁷ Fleiss’s kappa, Krippendorff’s alpha and pairwise average as part of assessing the reliability of agreement among users when assigning categorical ratings to classifying articles. These statistical measures are used for assessing the reliability of agreement between. In addition, we calculated the marginal distributions of raters, giving insight into how often each rater used a particular category (‘Conflict’, ‘Non-Conflict’ and ‘Not-sure’). The results are shown in the summary in Figure 14. As per Landis and Koch (1977) interpretation of κ values (Figure 15), our Kappa statistical measure indicates almost perfect agreement among the users.

¹⁷ Geertzen, J. (2012). Inter-Rater Agreement with multiple raters and variables. Retrieved May 2, 2015, from <https://mlnl.net/jg/software/ira/>

The average pairwise¹⁸ user agreement was also very high (95.8%). This implies that the users were in agreement on the definition of conflict.

Data
 5 raters and 103 cases
 1 variable with 515 decisions in total
 no missing data

1: frame_u1_u5

<i>Fleiss</i>		<i>Krippendorff</i>		<i>Pairwise avg.</i>	
A_obs	=	0.958	D_obs	=	0.042
A_exp	=	0.487	D_exp	=	0.514 % agr = 95.8
Kappa = 0.919		Alpha = 0.919		Kappa = 0.919	

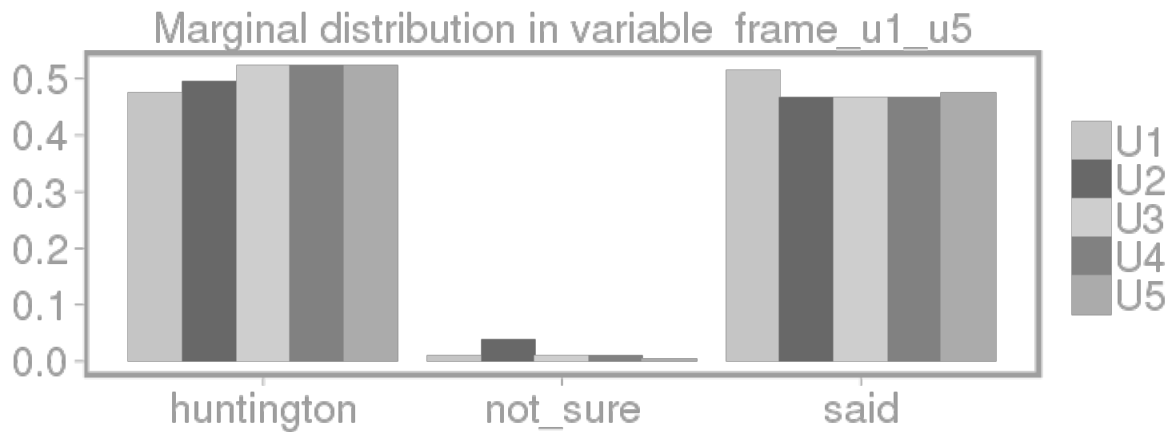


Figure 14 User Classification Results

<i>Kappa (κ)</i>	<i>Interpretation</i>
<i>0.41 – 0.60</i>	<i>Moderate agreement</i>
<i>0.61 – 0.80</i>	<i>Substantial agreement</i>
<i>0.81 – 1.00</i>	<i>Almost perfect agreement</i>

Figure 15 Kappa Interpretation

¹⁸ Pairwise agreement is the agreement between a pair of coders; the average pairwise agreement is the average of all pairwise agreements using all possible coder pairs.

Using the majority vote from our user-classified results, we evaluated our classifier performance by calculating sensitivity, specificity, precision and accuracy measures (Figure 16). Sensitivity of a system relates to how well the test predicts a class correctly; specificity relates to how well the test excludes a class correctly; the precision relates to reproducibility and repeatability; and the accuracy relates to how well the binary classifier identifies a class. Our results suggest high performance of the classifier on all measures. The majority vote resolves the disagreements among human coders and provides a set of truth-labels against which we can evaluate our classifier. Our classification accuracy was very high (99.03%), which implies that our machine classifier accurately predicted human classified articles. An accuracy of 100% implies that the predicted values are exactly the same as the actual values.

Sensitivity	$TP / (TP + FN)$	100.00%
Specificity	$TN / (TN + FP)$	97.92%
Precision	$TP / (TP + FP)$	98.21%
Accuracy	$(TP + TN) / (TP + FP + FN + FP)$	99.03%

Figure 16 Classifier Performance

Task 2

The non-conflict and conflict map pages show the treemap of 50 most frequent terms for each category. It also provides sample articles that are classified in the respective category using the Said-Huntington index. We asked the users to go through conflict and non-conflict map pages and describe whether the sample articles were classified correctly. The goal of this task was to ascertain whether the tool could classify discourse based on the conflict/non-conflict criterion. The experimental evaluation of the task was based on the following question:

Did the users agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

Task 2 Results

Our results indicate that there was 90.9% average pairwise user agreement on the classification of our results. And in addition, the users “agreed” with our classification results in 59 out of 60 cases (Figure 17, 18).

(A = Agree; NAD = Neither Agree Nor Disagree)

	<i>User1</i>	<i>User2</i>	<i>User3</i>	<i>User4</i>	<i>User5</i>
<i>Al Jazeera</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
<i>Times of India</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
<i>The New York Times</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
<i>Daily Mail Online</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
<i>USA Today</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
<i>Washington Post</i>	<i>A</i>	<i>NAD</i>	<i>A</i>	<i>A</i>	<i>A</i>
<i>Dawn</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
<i>Reuters</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
<i>Fox News</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
<i>The News International</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
<i>MSNBC</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>

Figure 17 Pairwise User Agreement on Conflict/Non-Conflict Maps Classification

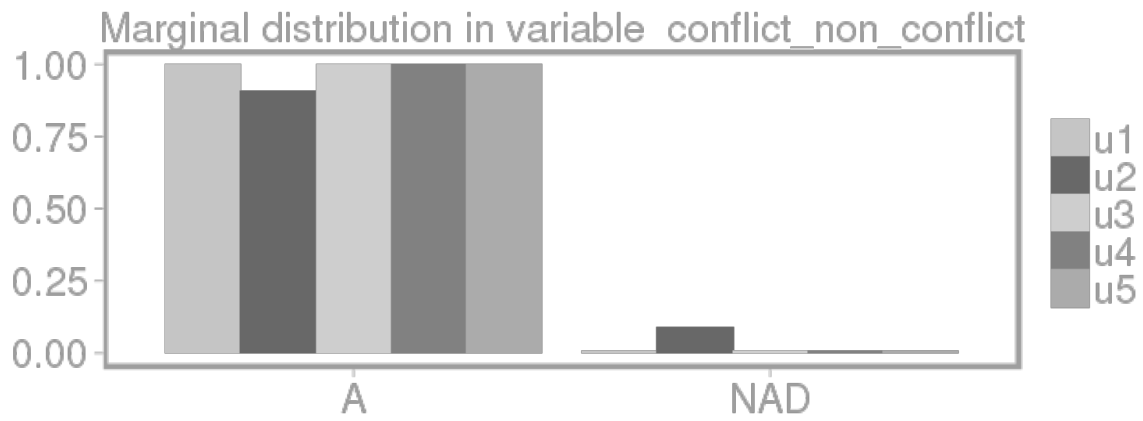


Figure 18 Marginal Distribution of Conflict/Non-Conflict Maps Classification

Task 3

Our tool delineates conversations for a media source (such as Al Jazeera or Reuters) by extracting five topics and corresponding keywords using the LDA model. In task 3, we asked the users to go through each LDA-generated conversation. We asked them to describe: a) whether they agreed with the Said-Huntington index classification for the conversation; and b) whether the topic or conversation, as a whole, was coherent.

The experimental evaluation of the task was based on the following questions:

- a) How effective was the LDA model in discovering the latent topics in each media news corpus? Did the users classify the topics as semantically coherent?
- b) How effective was Said-Huntington index in classifying these conversations in terms of conflict and non-conflict?

Task 3 Results

The users found 86% (Figure 19) of conversations as semantically coherent, implying that our algorithm was effective in extracting coherent topics.

<i>Topic Coherent?</i>	<i>User1</i>	<i>User2</i>	<i>User3</i>	<i>User4</i>	<i>User5</i>		
<i>(YES)</i>	48	39	57	55	57	256	85.33%
<i>(Somewhat)</i>	8	13	3	4	3	31	10.33%
<i>(NO)</i>	4	8	0	1	0	13	4.33%
						300	100.00%

Figure 19 Topic Coherence Results

There was 77.5% pairwise user agreement on whether our topics were coherent or not (Figure 20).

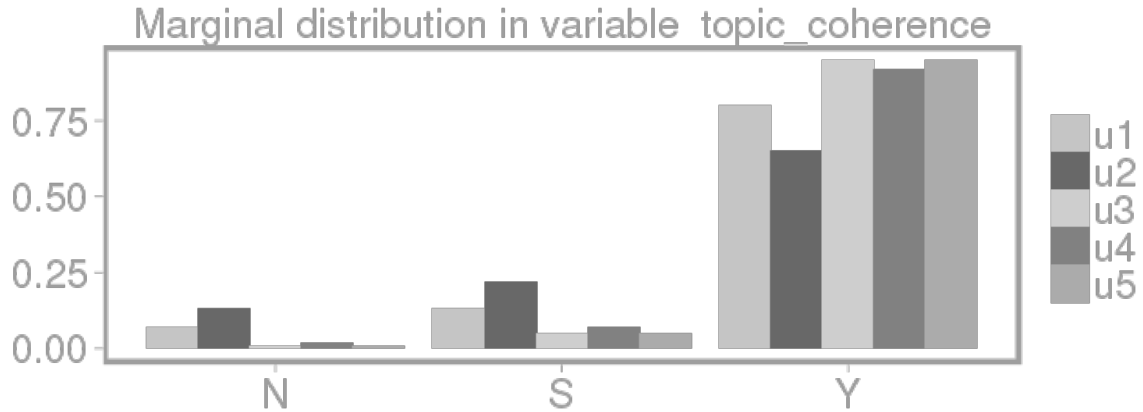


Figure 20 Marginal Distribution of Topic Coherence

The users rated 95% (Figure 21) of conversations as classified correctly, implying that our algorithm was effective in the classification of these conversations.

Total conversations = 60

	<i>User1</i>	<i>User2</i>	<i>User3</i>	<i>User4</i>	<i>User5</i>	
<i>Agree</i>	58	52	58	58	59	95.00%
<i>Disagree</i>	2	4	2	2	1	3.67%
<i>Neither Agree nor Disagree</i>	0	4	0	0	0	1.33%
	60	60	60	60	60	

Figure 21 Topic Classification Ratings

There was 92.5% average pairwise user agreement on whether the user agreed with our classification of the conversation or topic (Figure 22).

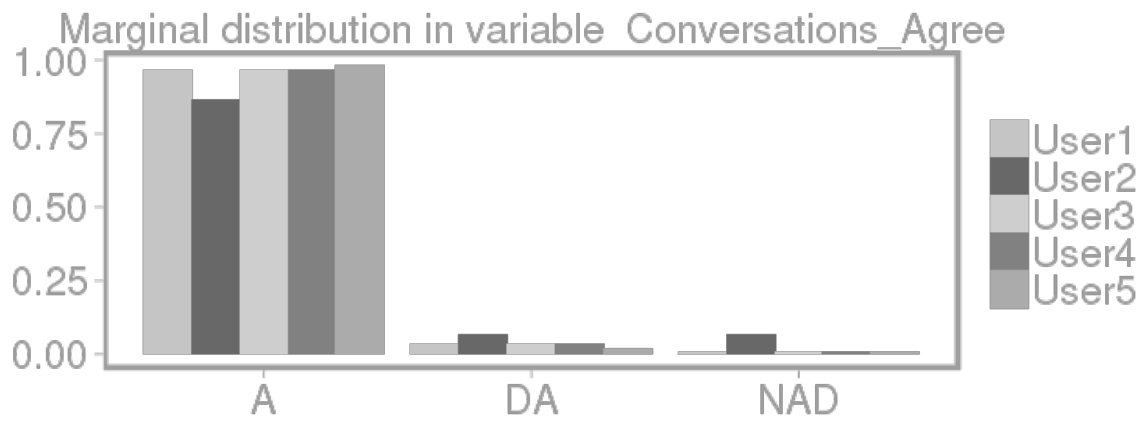


Figure 22 Marginal Distribution of Conversations Agreement

Task 4

As part of tool's teleology and functionality our goal was to ascertain whether the tool can delineate media discourse based on our conflict hypothesis and Said-Huntington classification index. We wanted to determine whether the tool was effective in presenting and demarcating the latent discourse patterns to the user. The experimental evaluation of tool's functionality was based on the following questions:

- a) After using the tool, did the users agree with the thesis that global media largely covers Islam in the context of conflict?
- b) Did the users find the tool helpful in understanding/deciphering the global media discourse on Islam?
- c) Did the users agree with the assertion that Said-Huntington index effectively classifies the conversations and the articles?

Results Task 4

The tool is based on the hypothesis that mainstream global media tends to lean toward dangerous "clash of civilizations" frame by focusing primarily on conflict. Our results from the task 4 suggest that the implementation of tool on the basis of CDA framework allowed users to interpret the discourse in the critical/ideological sense. The users confirmed our hypothesis by interpreting the insights from our tool. The results of task 4 are as follows:

- All the users "strongly agreed" with our hypothesis that global media largely covers Islam in the context of conflict.

- Four out of five users found the tool “very helpful” in understanding/deciphering the global media discourse on Islam; one user found it “somewhat helpful”.
- Three users “strongly agreed” with the assertion that the Said-Huntington index effectively classifies the conversations and the articles; while two just “agreed” with the assertion.

Task 5

The user interface is an integral part of the tool since it determines how well it translates its design goals. The experimental evaluation of the tool interface was based on the following questions:

- a) Did the users agree with the assertion that the visual representations and color layouts in the tool were effective in unraveling the underlying discourse structure of the media coverage?
- b) Did the users find the tool navigation uncomplicated and straightforward?
- c) Did the users agree with the assertion that the scorecard visual interface effectively communicates the conflict coverage for the media sources?
- d) Did the users find conflict- and non-conflict map visualizations helpful in understanding the conflict content?
- e) Did the users find the color highlighting of conflict- and state-related terms helpful in visualizing the conflict coverage?
- f) Did the users find color-coded bubble charts/frames effective in delineating different conversations/topics within a media source?

Results Task 5

The feedback from task 5 suggests that the tool met its interface design goals. The results of task 5 are as follows:

- Four out of five users “agreed” with the assertion that the tool navigation is uncomplicated and straightforward, while one user “disagreed” with the assertion.

- All the users “agreed” with the assertion that the scorecard visual interface effectively communicates the conflict coverage for the media sources.
- Three users found the conflict- and non-conflict map visualizations “very helpful” in understanding the conflict content, while one found it “helpful” and one found it “somewhat helpful”.
- Three users found the color highlighting of conflict- and state-related terms “very helpful” in visualizing the conflict coverage, while two found it “helpful”.
- Four out of five users found the color-coded bubble charts/frames “effective” in delineating different conversations/topics within a media source. One user found it “somewhat effective”.
- Three out of five users “strongly agreed” with the assertion that the visual representations and color layouts in the tool were effective in unraveling the underlying discourse structure of the media coverage. Two users just “agreed” with the assertion.

Task 6

Finally, we wanted to evaluate the tool on its practical usability in different settings. As part of evaluation we asked the users to provide feedback on the applications of the tool.

Task 6 Results

The results (Figure 23) suggest that the users see the application of the tool in diverse settings. All users suggested that the tool could be used in newsrooms, media organizations and academic research.

A summary of the results is given below:

	<i>This tool could be used in:</i>				
	<i>Newsrooms</i>	<i>Watchdog organizations</i>	<i>Academic Research</i>	<i>Media Organization</i>	<i>Marketing</i>
<i>U1</i>	<i>X</i>		<i>X</i>	<i>X</i>	
<i>U2</i>	<i>X</i>	<i>X</i>	<i>X</i>	<i>X</i>	
<i>U3</i>	<i>X</i>	<i>X</i>	<i>X</i>	<i>X</i>	<i>X</i>
<i>U4</i>	<i>X</i>	<i>X</i>	<i>X</i>	<i>X</i>	<i>X</i>
<i>U5</i>	<i>X</i>	<i>X</i>	<i>X</i>	<i>X</i>	

Figure 23 Feedback on Tool's Practical Usability in Different Settings

Insights & Findings

Coverage of Islam

Our scorecard data confirms our hypothesis that the mainstream media largely covers Islam in terms of conflict. We found that nine out of twelve media sources covered Islam in the context of conflict¹⁹ (Figure 24).

<i>Source</i>	<i>Conflict %</i>	<i>Country of Origin</i>	<i>Religion</i>	<i>Conflict % > 50?</i>
<i>Al Jazeera</i>	<i>93%</i>	<i>Qatar</i>	<i>Muslim</i>	<i>Yes</i>
<i>Times of India</i>	<i>37.60%</i>	<i>India</i>	<i>Non-Muslim</i>	<i>No</i>
<i>The New York Times</i>	<i>82.70%</i>	<i>USA</i>	<i>Non-Muslim</i>	<i>Yes</i>
<i>Daily Mail Online</i>	<i>64.50%</i>	<i>UK</i>	<i>Non-Muslim</i>	<i>Yes</i>
<i>USA Today</i>	<i>86%</i>	<i>USA</i>	<i>Non-Muslim</i>	<i>Yes</i>
<i>Washington Post</i>	<i>83.60%</i>	<i>USA</i>	<i>Non-Muslim</i>	<i>Yes</i>
<i>Dawn</i>	<i>64.20%</i>	<i>PK</i>	<i>Muslim</i>	<i>Yes</i>
<i>Reuters</i>	<i>84.60%</i>	<i>UK</i>	<i>Non-Muslim</i>	<i>Yes</i>
<i>Fox News</i>	<i>84.90%</i>	<i>USA</i>	<i>Non-Muslim</i>	<i>Yes</i>
<i>The News</i>	<i>37.10%</i>	<i>PK</i>	<i>Muslim</i>	<i>No</i>
<i>MSNBC</i>	<i>75.80%</i>	<i>USA</i>	<i>Non-Muslim</i>	<i>Yes</i>

Figure 24 Conflict Scorecard Results

It is surprising that Al Jazeera—whose vision is “to introduce to the Arab world free reporting that is distant from propaganda, and at the same time to give the Arab world the opportunity to express opinions” (Barker, 2007)—largely covers Islam within the context of conflict. About 93% of its articles represented coverage related to conflict.

¹⁹ Our user evaluation of conflict and non-conflict article samples supports the validity of our findings. The users “agreed” with our classification of conflict and non-conflict articles in 59 out of 60 cases.

Similarly, Dawn News, which is the second most widely-read newspaper (Shah, 2010) in Pakistan (where 97% of Pakistanis are Muslims), also covered Islam essentially in the context of conflict (64%). In contrast, The News International, the largest newspaper (Shah, 2010) in Pakistan, reported about 37% of its articles on conflict related issues.

On the other hand, less than 38% of articles in the Times of India— which is the largest English-language daily newspaper in the world (Thussu, 2013)—were related to conflict. For a country like India, which has been actively engaged in cross-border conflict with Pakistan, it is notable to see its most popular newspaper reporting nearly 72% of articles on non-conflict issues. In the next part, we will analyze the example of the Times of India in detail to demonstrate some discourse-related findings and insights from our tool.

The Times of India Example

The Times of India has one of the lowest percentages of conflict articles in its corpus. An appraisal of its conversations also supports this aspect. Three out of five of its conversations are classified as non-conflict on the Said-Huntington index (Figure 25). The users agree with the Said-Huntington index classification in 24 out of 25 cases (96%) of conversations within Times of India; there are no conversations that the user found not coherent, implying that the conversations were semantically coherent topics. In the next sections, we will examine each conversation individually.

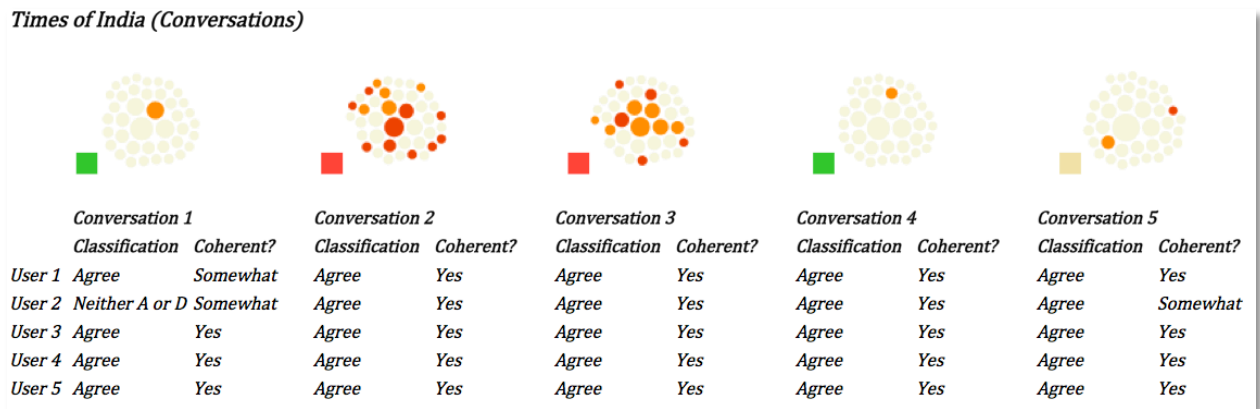


Figure 25 Times of India Conversations

Analysis of Conversation 1

The first conversation (Figure 26) in Times of India loosely exhibits the discourse surrounding India's domestic politics. It is a strong topic, as the topic strength of 85% suggests. Its Said-Huntington index is 13, which classifies it as a non-conflict conversation. The conversation exhibits heterogeneity of voices from its diverse communities. Terms like 'minority', 'religion', 'modi', 'government', 'sikh', 'law', 'bill' and 'population' suggest a permeation of state interest in its sub-national groups.

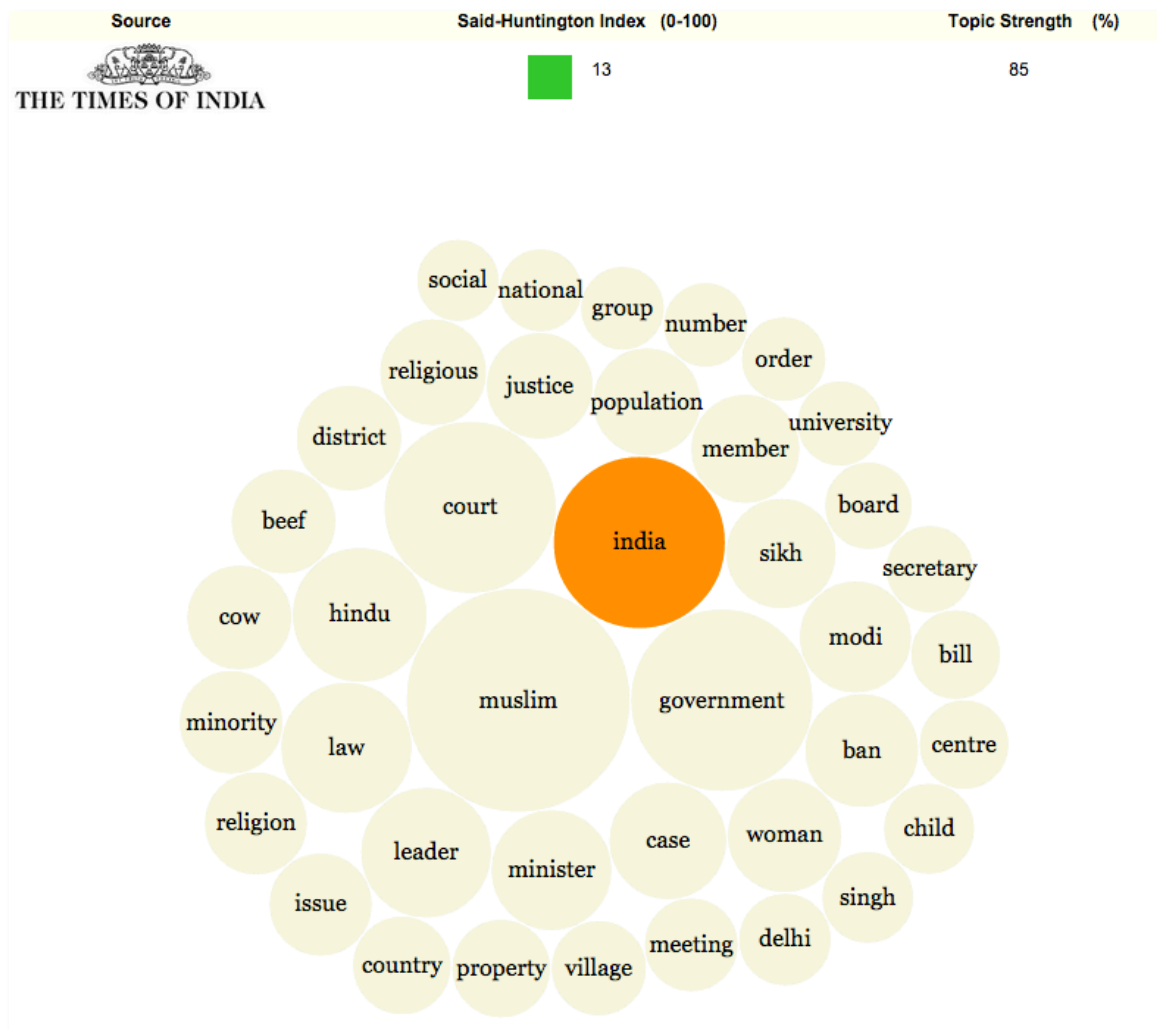


Figure 26 Times of India Conversation 1

Analysis of Conversation 2

The second conversation (Figure 27) is a prototypical conflict conversation. Terms like ‘attack’, ‘military’, ‘terror’, ‘violence’, ‘isil’ and ‘syria’ suggest that this conversation represents ongoing global conflict in the Middle East. Its Said-Huntington index of 100 classifies it as a conversation typified by extreme conflict. It is plagued with conflict-related terms like ‘attack’, ‘military’, ‘terror’, ‘violence’, ‘isil’ and ‘syria’.

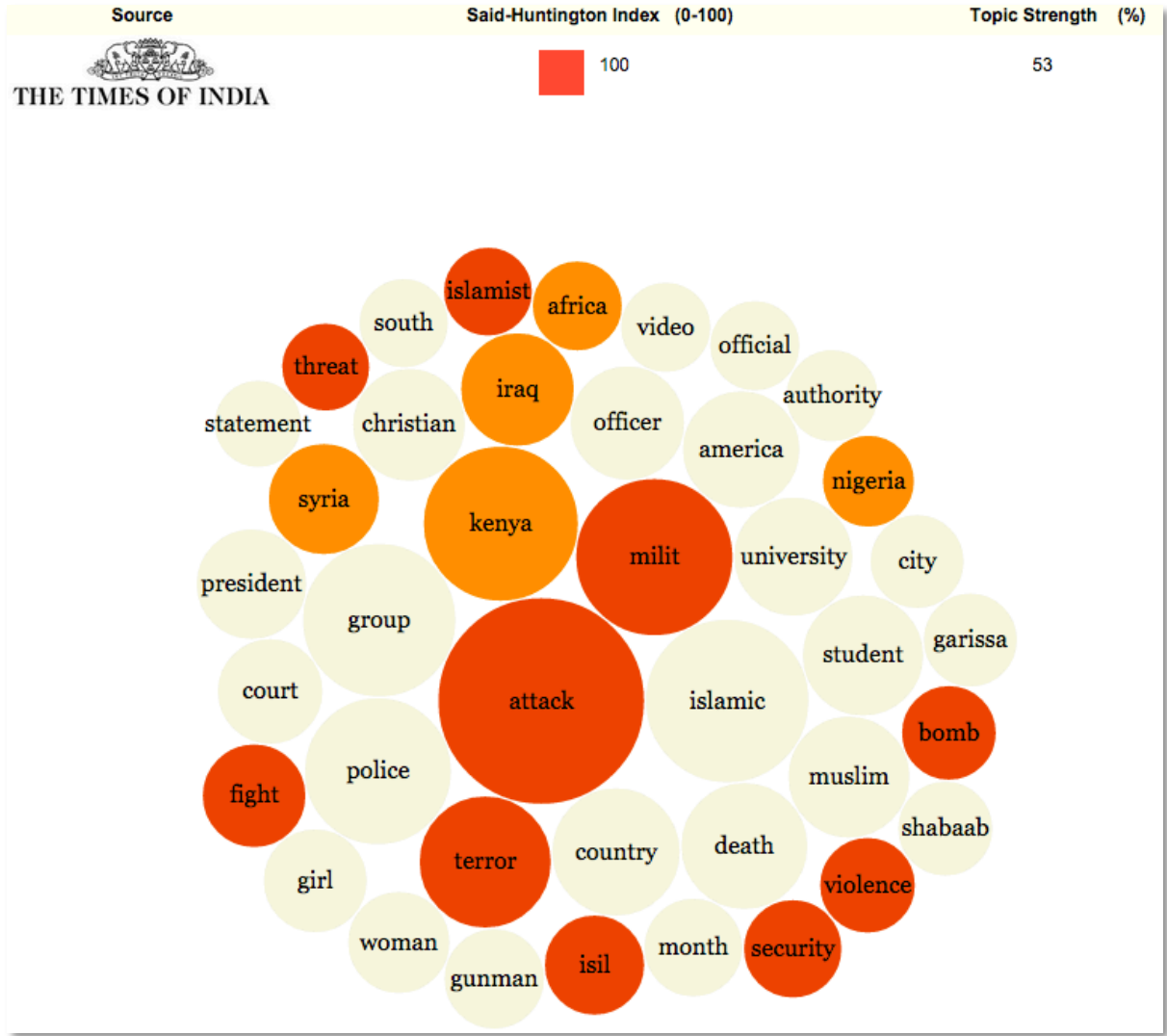


Figure 27 Times of India Conversation 2

Analysis of Conversation 3

The third conversation (Figure 28) could represent recent ongoing conflict in Yemen that has created major fault lines in the region. Like the previous conversation, its Said-Huntington index is 100, which classifies it as a conversation of extreme conflict. It is the strongest topic in the conversation cluster, as the topic score of 100% suggests. Terms like ‘yemen’, ‘iran’, ‘pakistan’ and ‘saudi’ describe the key actors in the conflict, whereas terms like ‘security’, ‘milit’, ‘force’ and ‘strike’ denote conflict engagements.

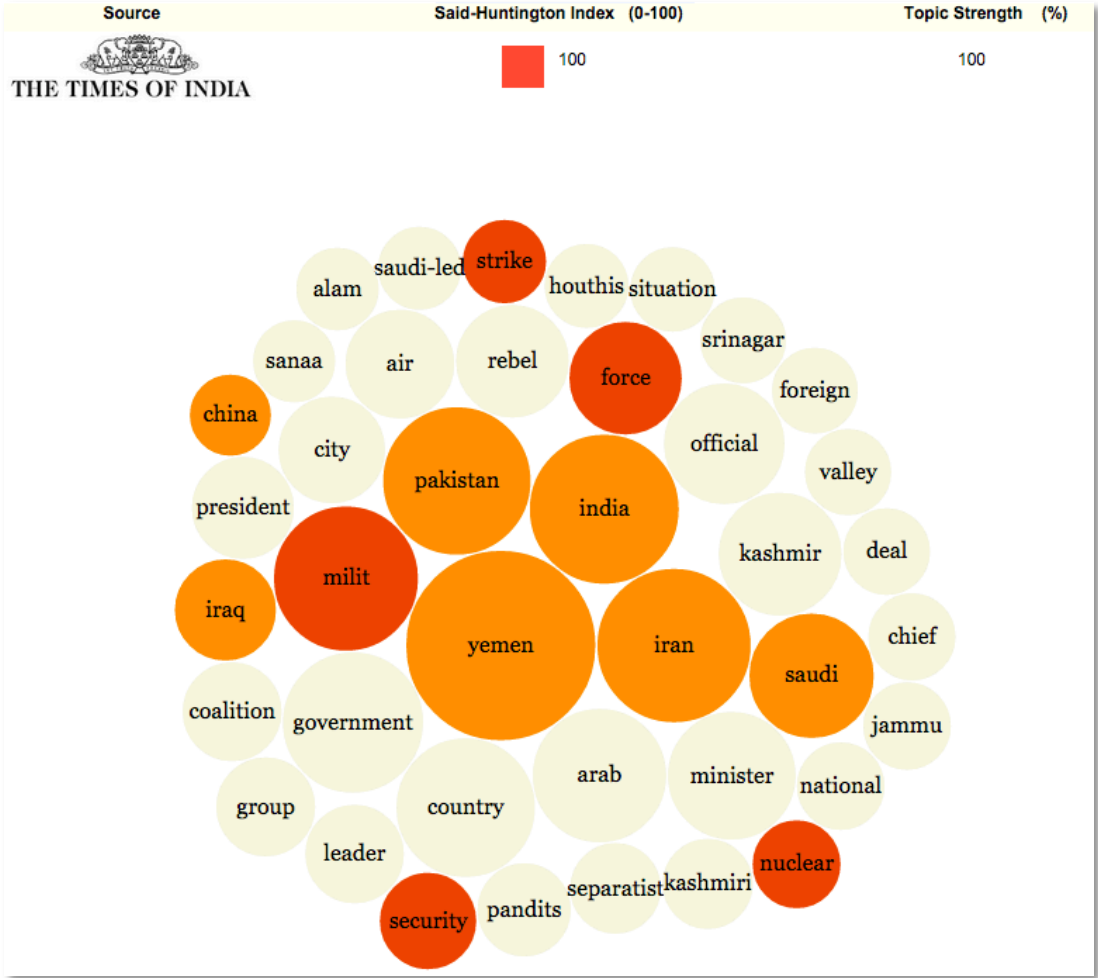


Figure 28 Times of India Conversation 3

Analysis of Conversation 4

The fourth conversation (Figure 29) could represent political dialectics surrounding the Muslim community in India. Terms like ‘muslim’, ‘urdu’ and ‘islam’ typify Muslim community. The topic strength of 61% suggests that it is a moderately strong topic. Its Said-Huntington index is 7, which classifies it as a non-conflict conversation. Terms like ‘congress’, ‘mim’, and ‘election’ could be explained by recent by-elections in Bandra²⁰, India, where Congress and MIM²¹, two Indian parties, competed for Muslim vote in the region. The presence of the term ‘bjp’ could be described in the context of Bharatiya Janata Party’s relationship with Muslims. BJP is a well-known exponent of Hindu nationalism in India and the Muslims in India see BJP as an opposition against Muslim communal concerns.

²⁰ <http://timesofindia.indiatimes.com/city/mumbai/Congress-and-MIM-compete-for-Muslim-vote-in-Bandra-bypoll/articleshow/46752713.cms>

²¹ All India Majlis-e-Ittehadul Muslimeen

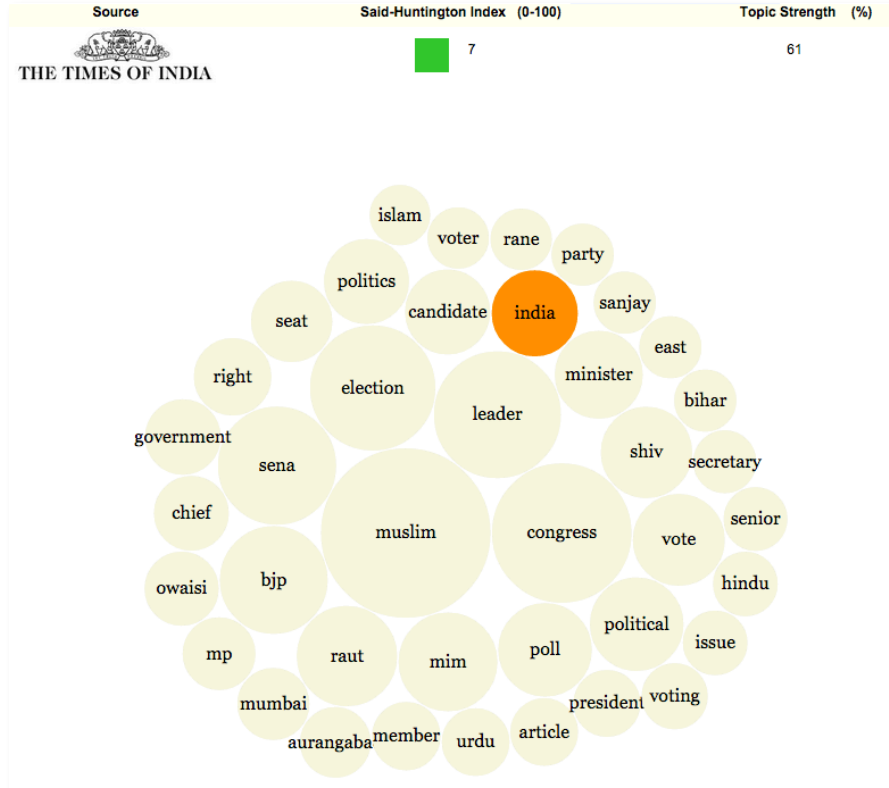


Figure 29 Times of India Conversation 4

Analysis of Conversation 5

The fifth conversation (Figure 30) is relatively hard to interpret, even though four out of five users categorized it as a semantically coherent topic. It has moderately strong topic strength of 65%. Its Said-Huntington index is 24, which classifies it as a non-conflict conversation, yet it contains the term ‘terror’. But overall the topic exhibits city/municipal politics with terms like ‘police’, ‘city’, ‘area’, ‘resident’, ‘house’, ‘local’ and ‘district’. Presence of terms like ‘mosque’, ‘temple’ and ‘church’ could denote the sacred public spaces, which are part of day-to-day civic life in India.

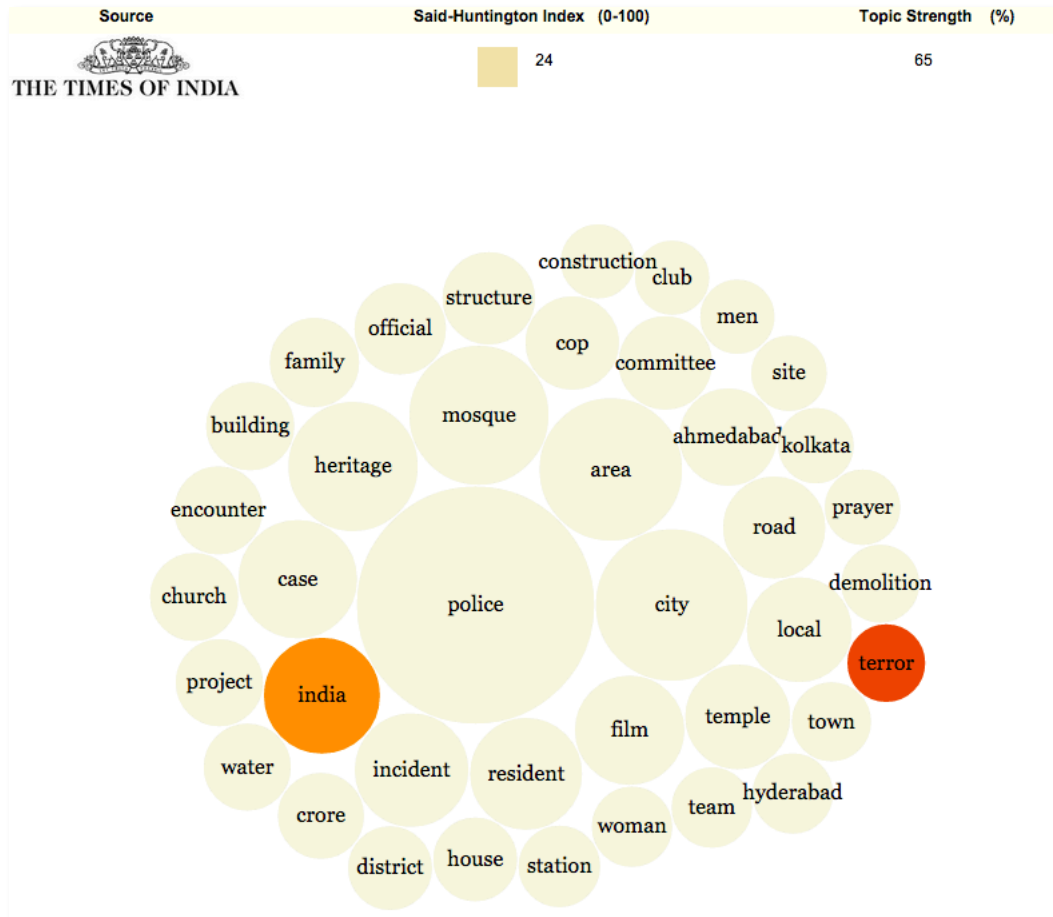


Figure 30 Times of India Conversation 5

Our preceding analysis suggests that the discourse in Times of India is not exclusively centered on conflict. When Islam or Muslims are discussed in Times of India, they are discussed in a diverse range of topics.

Discussion

In accordance with the objectives of my research, I have developed a tool that uses data driven approaches for classifying news articles to identify polarization, bias and the dominant discourses in mainstream news media. This tool detected that coverage in several news sources tends to disproportionately portray Muslims as a group embroiled in conflict, hence confirming my hypothesis that discourse on Islam in mainstream global media has privileged the “clash of civilizations” frame.

As a practical application, my tool attempts to deconstruct the stereotypical discourse on Islam in the mainstream media to address a bias problem in global journalism that affects 1.6 billion Muslims worldwide and their relationship with the rest of the world.

The tool presented here has been developed for use in both newsrooms and academic settings. Individuals, civic society members and ethnic and religious organizations can use this tool to highlight cultural stereotyping and bias in the media. The tool empowers user-creators to substantiate their concerns around bias and agendas with qualitative, data-supported evidence. The tool facilitates the agency of user-creators to make media and public organizations more accountable, bias-free and transparent.

As part of the tool implementation, I have also developed Said-Huntington Index, a composite indicator for measuring the degree of conflict in news documents. In the user evaluation of the tool, I tested whether the thematic patterns extracted from the tool lend support to our hypothesis. The tool was evaluated using quantitative, qualitative and statistical methods. The

evaluation results suggest that the tool met its design and implementation objectives. In the machine learning implementation component of the tool, I developed a heuristic for producing the best performing classifier for predicting conflict in an article. I evaluated the classifier on human-classified articles and found that the accuracy of the classifier was very high (99.03%). This implies that the heuristic produces high-performance classifiers.

Our tool delineates semantically meaningful conversations for a media source using LDA topic modeling, allowing the users to discover frames of conversations on the basis of Said-Huntington index classification. In our user evaluation, the users rated 95% of conversations as classified correctly, implying that our algorithm was effective in the classification of these conversations.

This thesis is an attempt to create a hypothesis-driven framework using Critical Discourse Analysis to produce a computational tool that could be used for discourse analysis in applied settings. Generally, text analysis tools uncover patterns and trends in the data without delineating the ‘ideology’ that is embedded in the text. In contrast, our tool delineates the ideology that permeates the text. Hence, this tool is a novel computational application of critical theory. It is also important to bear in mind that critical theory has been criticized for elitist theorizing (Mack, 2010) and lack of “extended empirical studies” (Alvesson and Deetz, 2000, p. 110). This tool not only has the capacity to test empirical evidence for hypotheses/claims/conclusions from critical theory through practical application of Critical Discourse Analysis, it also provides a practical hermeneutics for media narratives.

Limitations & Future Directions

One of the main limitations with the tool is that at this point in time it does not allow the users to ask hypothesis-driven questions of their own. A generalized version of the tool that extracts conversations and topic maps based on standard topics is relatively easy to implement but my goal is to retain the CDA functionality in the generalized version. As part of the next iteration, I aim to implement another instance of this technological framework that will attempt to delineate American media discourse on the African-American community from the point of view of crime. This implementation will allow me to develop a generalized set of implementation parameters that will enable the users to utilize this tool for querying on open questions.

Another limitation with the tool, which I aim to overcome in future work, is that the users cannot use an LDA-generated conversation to retrieve news stories that could be classified within the context of the conversation. We can easily provide this functionality by converting lexical features of the conversation into a classifier.

Finally, the tool currently provides cross-sectional data insights, which do not take into account the time dimension. As part of future work, I intend to incorporate in the tool the longitudinal data perspective, which will allow the users to see the changes in discourse over time.

Bibliography

- Abrahamian, E., 2003. The US Media, Huntington and September 11. *Third World Quarterly* 24, 529–544.
- Allwright, D., Bailey, K.M., 1991. *Focus on the language classroom: An introduction to classroom research for language teachers*. Cambridge University Press.
- Alvesson, M., Deetz, S., 2000. *Doing critical management research*. Sage.
- Asuncion, H.U., Asuncion, A.U., Taylor, R.N., 2010. Software traceability with topic modeling. Presented at the Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 1, ACM, pp. 95–104.
- Awad, M., Khanna, R., 2015. Machine Learning and Knowledge Discovery, in: *Efficient Learning Machines*. Springer, pp. 19–38.
- Baker, P., 2012. *Contemporary corpus linguistics*. A&C Black.
- Baker, P., 2010. Representations of Islam in British broadsheet and tabloid newspapers 1999–2005. *Journal of Language and Politics* 9, 310–338.
- Baker, P., Gabrielatos, C., McEnery, T., 2013a. Sketching Muslims: A corpus driven analysis of representations around the word “Muslim” in the British press 1998–2009. *Applied linguistics* 34, 255–278.
- Baker, P., Gabrielatos, C., McEnery, T., 2013b. *Discourse analysis and media attitudes: the representation of Islam in the British press*. Cambridge University Press.
- Barker, G., 2007. Interview With Wadah Khanfar, Director General, Al Jazeera [WWW Document]. PBS. URL http://www.pbs.org/frontlineworld/stories/newswar/war_interviews.html (accessed 5.9.15).
- Blei, D.M., 2012. Probabilistic topic models. *Communications of the ACM* 55, 77–84.
- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3, 993–1022.
- Bowles, M., 2015. *Machine Learning in Python: Essential Techniques for Predictive Analysis*. John Wiley & Sons.
- Campbell, J.L., 2002. Ideas, politics, and public policy. *Annual review of sociology* 21–38.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine learning* 20, 273–297.
- Crossley, S.A., Allen, L.K., Kyle, K., McNamara, D.S., 2014. Analyzing Discourse Processing Using a Simple Natural Language Processing Tool. *Discourse Processes* 51, 511–534.
- Domingos, P., 2012. A few useful things to know about machine learning. *Communications of the ACM* 55, 78–87.
- Du, L., Buntine, W.L., Jin, H., 2010. Sequential latent dirichlet allocation: Discover underlying topic structures within a document. Presented at the Data Mining (ICDM), 2010 IEEE 10th International Conference on, IEEE, pp. 148–157.
- Dumais, S., Platt, J., Heckerman, D., Sahami, M., 1998. Inductive learning algorithms and representations for text categorization. Presented at the Proceedings of the seventh international conference on Information and knowledge management, ACM, pp. 148–155.
- Eagleton, T., 1996. *Literary theory: An introduction*. U of Minnesota Press.

- Eck, M., Vogel, S., Waibel, A., 2004. Language Model Adaptation for Statistical Machine Translation Based on Information Retrieval. Presented at the LREC.
- Elmore, R.F., 2009. School Improvement and the Reduction of Poverty. *Poverty and Poverty Alleviation Strategies in North America* 26, 165.
- Fagan, J.L., 1989. The effectiveness of a nonsyntactic approach to automatic phrase indexing for document retrieval. *Journal of the American Society for Information Science* 40, 115–132.
- Fairclough, N., 2013. *Language and power*. Routledge.
- Foucault, M., 1972. *The Archaeology of Knowledge: Translated from the French by AM Sheridan Smith*. Pantheon Books.
- Foucault, M., Bouchard, D.F., 1980. *Language, counter-memory, practice: Selected essays and interviews*. Cornell University Press.
- Goncalves, A.A., 2014. *Conflicting Frames: the dispute over the meaning of rolezinhos in Brazilian media*.
- Graeff, E., Stempeck, M., Zuckerman, E., 2014. The battle for “Trayvon Martin”: Mapping a media controversy online and off-line. *First Monday* 19.
- Graesser, A.C., McNamara, D.S., Cai, Z., Conley, M., Li, H., Pennebaker, J., 2014. Coh-Metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal* 115, 210–229.
- Grenoble, L.A., 2006. *Discourse analysis*. Glossos 8.
- Huntington, S.P., 1993. The clash of civilizations? *Foreign affairs* 22–49.
- Hunt, W.B., 1996. *Getting to war: Predicting international conflict with mass media indicators*. University of Michigan Press.
- Joachims, T., 2002. *Learning to classify text using support vector machines: Methods, theory and algorithms*. Kluwer Academic Publishers.
- Joachims, T., 2001. A statistical learning model of text classification for support vector machines. Presented at the Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, pp. 128–136.
- Joachims, T., 1998. *Text categorization with support vector machines: Learning with many relevant features*. Springer.
- Kilgarriff, A., Rychly, P., Smrz, P., Tugwell, D., 2004. Itri-04-08 the sketch engine. *Information Technology* 105, 116.
- Kim, S.N., Baldwin, T., Kan, M.-Y., 2010. Evaluating N-gram based evaluation metrics for automatic keyphrase extraction. Presented at the Proceedings of the 23rd international conference on computational linguistics, Association for Computational Linguistics, pp. 572–580.
- Leetaru, K., 2013. *Can we forecast conflict? A framework for forecasting global human societal behavior using latent narrative indicators*.
- Loper, E., Bird, S., 2002. NLTK: The natural language toolkit. Presented at the Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1, Association for Computational Linguistics, pp. 63–70.
- Mack, L., 2010. The philosophical underpinnings of educational research. *Polyglossia* 19, 5–11.

- Manning, C.D., Raghavan, P., Schütze, H., 2008. Introduction to information retrieval. Cambridge university press Cambridge.
- Mautner, G., 2009. Corpora and critical discourse analysis, in: Contemporary Corpus Linguistics. pp. 32–46.
- Maxion, R.A., Townsend, T.N., 2002. Masquerade detection using truncated command lines. Presented at the Dependable Systems and Networks, 2002. DSN 2002. Proceedings. International Conference on, IEEE, pp. 219–228.
- Meretakís, D., Fragoudís, D., Lu, H., Likothanassis, S., 2000. Scalable association-based text classification. Presented at the Proceedings of the ninth international conference on Information and knowledge management, ACM, pp. 5–11.
- Montiel, C.J., Salvador, A.M.O., See, D.C., De Leon, M.M., 2014. Nationalism in Local Media During International Conflict Text Mining Domestic News Reports of the China–Philippines Maritime Dispute. *Journal of Language and Social Psychology* 0261927X14542435.
- Morey, P., Yaqin, A., 2011. Framing Muslims. Harvard University Press.
- Navarro-Galindo, J.L., Samos, J., Muñoz-Alfárez, M.J., 2012. A Hybrid Approach to Text Categorization Applied to Semantic Annotation. Presented at the Database and Expert Systems Applications, Springer, pp. 39–47.
- Oostdijk, N., 1991. Corpus linguistics and the automatic analysis of English. *Rodopi*.
- Pang, B., Lee, L., 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval* 2, 1–135.
- Pang, B., Lee, L., Vaithyanathan, S., 2002. Thumbs up?: sentiment classification using machine learning techniques. Presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, Association for Computational Linguistics, pp. 79–86.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* 12, 2825–2830.
- Powell, K.A., 2011. Framing Islam: An analysis of US media coverage of terrorism since 9/11. *Communication Studies* 62, 90–112.
- Řehůřek, R., Sojka, P., 2010. Software framework for topic modelling with large corpora.
- Said, E.W., 2008. *Covering Islam: How the Media and the Experts Determine How We See the Rest of the World (Fully Revised Edition)*. Random House.
- Said, E.W., 2001. *The clash of ignorance*. October.
- Salton, G., McGill, M.J., 1983. Introduction to modern information retrieval.
- Scott, S., Matwin, S., 1999. Feature engineering for text classification. Presented at the ICML, pp. 379–388.
- Shah, H., 2010. Inside Pages: An Analysis of the Pakistani Press. *South Asia Monitor* 148, 1–5.
- Smeaton, A.F., 1992. Progress in the application of natural language processing to information retrieval tasks. *The computer journal* 35, 268–278.
- Stubbs, M., 1997. Whorf’s children: Critical comments on critical discourse analysis (CDA). *British studies in applied linguistics* 12, 100–116.
- Thussu, D.K., 2013. *Communicating India’s soft power: Buddha to Bollywood*. Palgrave Macmillan.

- Vapnik, V.N., Kotz, S., 1982. Estimation of dependences based on empirical data. Springer-Verlag New York.
- Wimmer, A., Min, B., 2006. From empire to nation-state: explaining wars in the modern world, 1816–2001. *American Sociological Review* 71, 867–897.
- Wodak, R., 2001. The discourse-historical approach. *Methods of critical discourse analysis* 63–94.
- Yang, Y., Liu, X., 1999. A re-examination of text categorization methods. Presented at the Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, ACM, pp. 42–49.
- Yu, B., 2008. An evaluation of text classification methods for literary study. *Literary and Linguistic Computing* 23, 327–343.
- Zhu, N.Q., 2013. *Data Visualization with D3.js Cookbook*. Packt Publishing Ltd.

Appendix

User Evaluations Survey/Questionnaire

Said-Huntington Discourse Analyzer

Survey

NAME: _____

DATE: _____

Topics/Conversations Questions

MEDIA SOURCE:



ALJAZEERA

1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree

- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

6) Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

MEDIA SOURCE:

The New York Times

1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

6) Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

MEDIA SOURCE:



THE TIMES OF INDIA

1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

MEDIA SOURCE:

Daily Mail

1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

6) Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

MEDIA SOURCE:



1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

6) Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

MEDIA SOURCE:



1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

6) Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable



1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

6) Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
 - Disagree
 - Neither agree nor disagree
 - Agree
 - Strongly Agree
- Not sure/not applicable

MEDIA SOURCE:



1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

6) Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

MEDIA SOURCE:



1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

6) Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

MEDIA SOURCE:



1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

6) Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

MEDIA SOURCE:



1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

6) Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
 - Disagree
 - Neither agree nor disagree
 - Agree
 - Strongly Agree
- Not sure/not applicable

MEDIA SOURCE:



1) Do you agree with system's Said-Huntington index classification for conversation 1?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

2) Do you agree with system's Said-Huntington index classification for conversation 2?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

3) Do you agree with system's Said-Huntington index classification for conversation 3?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

4) Do you agree with system's Said-Huntington index classification for conversation 4?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

5) Do you agree with system's Said-Huntington index classification for conversation 5?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Is it a coherent topic? Yes _____ No _____ Somewhat _____

6) Do you agree with the assertion that the sample articles in the conflict and non-conflict maps are classified appropriately?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Hypothesis Questions

Said-Huntington Discourse Analyzer

After using the tool, do you agree with the thesis that global media largely covers Islam in the context of conflict?

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

Did you find the tool helpful in understanding/deciphering the global media discourse on Islam?

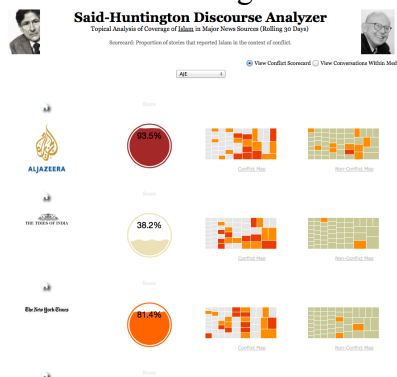
- Very helpful
- Somewhat helpful
- Neither helpful nor unhelpful
- Not helpful at all

User Interface Questions

1) Do you agree with the assertion that the tool navigation is uncomplicated and straightforward?

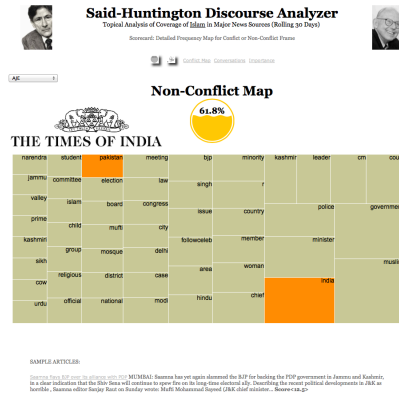
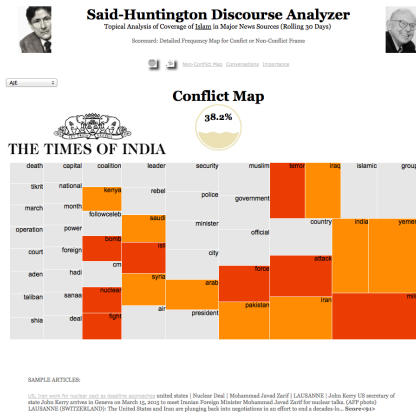
- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

2) Do you agree with the assertion that the scorecard visual interface effectively communicates the conflict coverage for the media sources?



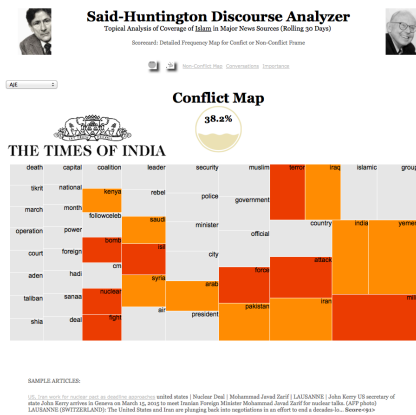
- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

3) Did you find the conflict- and non-conflict map visualizations helpful in understanding the conflict content?



- Very helpful
- Helpful
- Somewhat helpful
- Neither helpful nor unhelpful
- Not helpful at all





4) Did you find the color highlighting of conflict- and state-related terms helpful in visualizing the conflict coverage?



- Very helpful
- Helpful
- Somewhat helpful
- Neither helpful nor unhelpful
- Not helpful at all

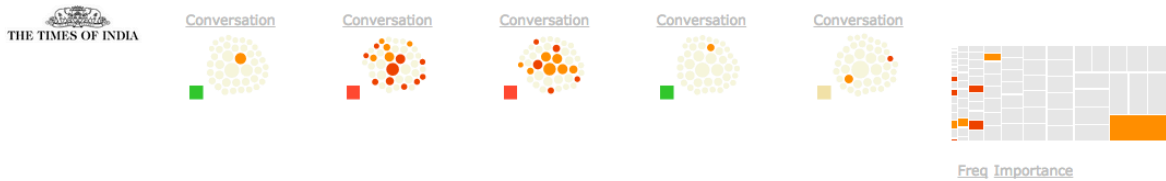
5) Did you find color-coded bubble charts/frames effective in delineating different conversations/topics within a media source?

6) Do you agree with the assertion that Said-Huntington index effectively classifies the conversations and the articles?

Source	Said-Huntington Index (0-100)	Topic Strength (%)
 ALJAZEERA	 100	89
Source	Said-Huntington Index (0-100)	Topic Strength (%)
 THE TIMES OF INDIA	 7	61

- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

7) Do you agree with the assertion that visual representations and color layouts in the tool were effective in unraveling the underlying discourse structure of the media coverage?



- Strongly Disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly Agree
- Not sure/not applicable

8) This tool could be used in:
(Check all that apply)

- Newsrooms
- Watchdog organizations
- Academic research
- Media organizations
- Marketing

Ethnographic studies

Not sure

User Classification of Articles Form

Name:

1)AJE h id:1:326770614:100.0:

[The fallen leaves of the Arab Spring](#)

hope self-searching middle east north africa reel conflict effect million syria grinding conflict month march practical significance reminder life livelihood innocence historical treasure tie violence account goal citizen arab spring country lofty choosing tyranny freedom dictatorship democracy pol...

High hopes of 2011 replaced by self-searching as Middle East and North Africa reel from conflicts' devastating effects. However, for millions of Syrians trapped in a grinding conflict, the month of M...

- Conflict
 Non-Conflict
 Not-sure
 No Answer Recorded

--

2)AJE h id:2:326780690:100.0:

[Syria is dead, long live the Syrians](#)

approach politics conflict diplomacy mar gmt war conflict human right politics east syria crisil jamesdenselow syria population country equivalent population croatia population dependent syria percent light syria national poverty report syria expectancy conflict medecins sans frontieres msf doctor ...

The approach of prioritising the 'hard politics' of conflict and diplomacy has failed. 15 Mar 2015 06:00 GMT | War & Conflict , Human Rights , Politics , Middle East , Syrian crisis @JamesDenselow Sy...

- Conflict
 Non-Conflict
 Not-sure
 No Answer Recorded

--

3)AJE h id:3:326801213:100.0:

[Analysis: Syria's chessboard politics](#)

lesson learnt syria crisis regional event statement regional implication israel war lebanon
basic geopolitical rule politics region cold war regional event local event implication
necessity regional international egypt syria epicentre arab homeland unfolding event
middle eastern chessboard politics...

The lesson learnt from the Syrian crisis four years on is that all regional events are interconnected. This statement specifically addresses the regional implications of Israel's 2006 war on Lebanon,...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

4)AJE h id:4:326801219:100.0:

[To understand Syria today, we must look to history](#)

middle east reality organic local identity strong ideological attachment mar gmt war
conflict politics human right middle east isil neopolitiks fourth anniversary syria revolution
civil war hundred thousand dead destruction country conflict closer resolution war
complex political dynamic region tra...

The Middle East is in reality an organic place where local identities are as strong as ideological attachments. 15 Mar 2015 09:20 GMT | War & Conflict , Politics , Human Rights , Middle East , ISIL @...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

5)AJE h id:5:326820264:100.0:

[Syria and the battered minorities of the Middle East](#)

syria minority middle east attack murder abduction humiliation minority middle east level
mar gmt war conflict politics human right middle east syria nation harryhagopian
neighbourhood east jerusalem christian muslim family armenia arab christian muslim
neighbour mum sensitive background armenia pa...

Syria and the battered minorities of the Middle East Attacks, murders, abductions and the

humiliation of minorities in the Middle East have now reached frightening levels. 15 Mar 2015 10:52 GMT | War...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

6)AJE s id:1:327303347:4.0:

[Myanmar court jails three for insulting Buddhism](#)

trio zealand national facebook depiction buddha headphone trio posting flyer social medium psychedelic depiction buddha headphone phillip blackwood tun thurein htut ko ko lwin insulting religion month disobeying order public servant trial gastro bar manager blackwood bar owner tun thurein employee ...

Trio, including a New Zealand national, convicted over Facebook post showing depiction of Buddha wearing headphones. The trio was convicted for posting a flyer on social media that showed a psychedel...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

7)MSNBC h id:6:326807483:100.0:

[Why Europe Is Pushing for Its Own Army](#)

france soldier infantry regiment ri bison rocco combat operation finger village dwakoleh surobi district sept ramstein germany conflict nato border case pan-europeanarmcommander multinational coalition force america long-term plan creation joint european union force backer efficient defense spendi...

French soldiers from The 126 Infantry Regiment (RI) The Bisons based at Rocco Combat Out Post take part in Operation 'Glued Finger 2' in the village of Dwakoleh in Surobi District, Sept. 21, 2010. RA...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

8)MSNBC h id:7:327087998:100.0:

[Meet the Press Transcript - March 15, 2015](#)

meet press march letter ayatollah republican supreme leader hillary clinton convenience personal email account controversy email presidential bid democrat plan divide america middle lindsey graham meet press email washington politician bridge century chuck todd insight analysis bai yahoo hillary cl...

MEET THE PRESS -- SUNDAY, MARCH 15, 2015 This Sunday, letter to the Ayatollah. Why did 47 Republicans write to Iran's supreme leader? And why some of them now regret it. Plus, is Hillary Clinton too ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

9)MSNBC h id:8:327532014:100.0:

[Meet the American Volunteers Fighting ISIS](#)

america jeremy woodward pose front isil emblem wall northern iraq kurdish troop fight isil coalition airstrikes isil position village white smoke target kurdish troop dozen volunteer america behalf milit volunteer milit northern iraq taste fight kurd burning village bulldozer fresh earth defensive ...

American Jeremy Woodward poses in front of an ISIS emblem painted on a wall in northern Iraq where he's joined Kurdish troops in their fight against ISIS. It began with coalition airstrikes that poun...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

10)MSNBC h id:9:328095509:100.0:

[Tunisia Attack: Is Arab Spring's Success Story Over?](#)

bystander guard bardo museum tunis tunisia tunisia troop major city massacre showpiece museum country implication attack status arab spring success death toll siege tunisia museum foreigner unity defiance display north africa nation democracy president beji essebsi thousand street capital solidarit...

Bystanders watch as policemen guard the Bardo museum in Tunis, Tunisia, Thursday.

Tunisia deployed troops to protect its major cities Thursday following a massacre in one of its showpiece museums, as...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

11)MSNBC h id:10:328371949:100.0:

[ISIS Is Rich, But Spending Money Fast](#)

isil analysis expert isil richest terror group iraq bank hundred million dollar official initial report theft estimate cost caliphate ground war intelligence analyst cash crunch group windfall oil revenue donation christian hostage executive chairman combating terror center west isil territory expe...

ISIS A new analysis by experts shows that ISIS is now the richest terror group in the world, thanks to a 2014 raid on an Iraqi bank that may have netted hundreds of millions of dollars. Officials who...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

12)MSNBC s id:2:330137683:6.0:

[Underground City Might Have Housed More Than 20,000](#)

turkey team biggest underground city underground city builder turk derinkuyu complex turkey region refuge carved-out cavern site bigger cave complex beneath byzantine-era hilltop castle nearby nevsehir foot meter cappadocia soft volcanic rock well-suited strange formation fairy chimney earth pyrami...

Turkish Team May Have Found Biggest Underground City Yet When it comes to hidden underground cities, few people were better builders than the Turks: The Derinkuyu complex in Turkey's Cappadocia regio...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

13)MSNBC s id:3:331592510:4.0:

[Why covering up is cool: Inside fashion's modesty movement](#)

inside fashion modesty movement change young woman one-piece swimsuit beach loose tunic tank top summertime chunky knit tight sweater winter modesty movement fashion sexy style backseat conservative perfect timing group religious style blogger designer 30-year-old melanie elturk ceo haute hijab vis...

Why covering up is cool: Inside fashion's modesty movement Maybe you've already noticed the change: Young women in one-piece swimsuits at the beach, loose tunics replacing tank tops in the summertime...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

14)MSNBC s id:4:332346832:1.0:

[From SingleSaints.com to JSwipe, how faith and dating now meet online](#)

religion young single faith relationship singlesaints faith faith app apps website christian mingle jdate ishqr cater specific religious population tinder broad jswipe fast access jewish single website relationship expert coach rachel dealto option daters faith mormon singlesaints ldsingles ldsmin...

Religion and online dating: How young singles are finding love within their faiths Relationships From SingleSaints.com to JSwipe, how faith and dating now meet online Looking to date someone of the s...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

15)MSNBC s id:5:334805411:1.0:

[This Singing Imam Is Not Your Typical Irish Music Star](#)

imam muhammad al-hussaini find ireland sean scene london ireland foreign concept london star scene eyebrow stereotype sean traditional unaccompanied singing style emerald isle sheikh muhammad al-hussaini imam scholar researching extremis coy talent beginner imam gaelic voice festival ireland learni...

Singing Imam: Muhammad Al-Hussaini Finds a Home in Irish Sean Nos Scene LONDON ?

While Irish song isn't a foreign concept in London, one new star on the scene has been raising eyebrows while shatteri...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

16)MSNBC s id:6:334938705:0.0:

[New Web Series Pokes Fun at Stereotypes Muslims Face](#)

web series halal stereotype muslim web series making typical america sitcom upbeat opening smiling face trouble lip chagrin teenage child theme halal parody stereotype edgy matter ordinary living town monster truck football stanza tune open declares six-minute episode qu osby sends message tone isl...

New Web Series 'Halal in the Family' Pokes Fun at Stereotypes Muslims Face A new web series has all the makings of a typical American sitcom: An upbeat opening song paired with smiling faces of a fam...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

17)USA_TODAY h id:11:328499130:100.0:

[Yemen bombings part of broader sectarian struggle](#)

yemen bomb broader sectarian struggle member houthi milit inspect scene suicide attack al-hashahush mosque sana yemen march photo yahya arhab european press agency washington horrific attack shiite mosque yemen highlight lethal sectarian struggle middle east iran shiite ally rival sunni milit organ...

Yemen bombings part of broader sectarian struggle Members of the Houthi militia inspect the scene of a suicide attack targeting the al-Hashahush mosque in Sana'a, Yemen, on March 20, 2015. (Photo: Ya...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

18)USA_TODAY s id:7:328622644:6.0:

[Measure to legalize gays' murder may move forward](#)

spring californian petition murder gay sacramento bee report lawyer matt mclaughlin proper paperwork fee initiative sodomite suppression bee san francisco chronicle report attorney general kamala choice signature-gathering phase harris legal previous case free-speech ground petition phase measure q...

At some point this spring, it appears that Californians will be asked to sign a petition legalizing the murder of gay people. And it will be perfectly legit. As the Sacramento Bee reports, a lawyer n...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

19)USA_TODAY h id:12:329123781:100.0:

[Yemen is latest in string of victories for Iran](#)

yemen latest string victory iran yemen protester slogan alight tire clash protest houthi takeover facility central city taiz yemen march photo ani epa washington violence drove adviser yemen key counterterror partner latest string regional victory iran rebel iran grip portion yemen advanced south -...

Yemen is latest in string of victories for Iran Yemeni protesters shout slogans and set alight tires during clashes after a protest against the Houthi takeover of several state facilities in the cent...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

20)USA_TODAY h id:13:330115263:100.0:

[Yemen fighting grows into regional conflict](#)

yemen grows regional conflict condition airstrikes iraq government offensive tikrit removal shiite milit fighting general sectarian yemen unfolding civil war regional conflict sunni-led saudi arab shiite power iran saudi-led coalition arab country airstrikes shiite rebel -backed regime president ha...

Yemen fighting grows into regional conflict The United States said that one of the

conditions for conducting airstrikes to support the Iraqi government in its offensive in Tikrit was the removal of S...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

21)USA_TODAY h id:14:330310524:100.0:

[Saudi airstrikes in Yemen target rebel stronghold](#)

saudi yemen target rebel stronghold shiite rebel houthis weapon saudi-led airstrikes rally sanaa yemen march photo hani ap sanaa yemen ap saudi arab northern stronghold yemen rebel key milit installation coalition gulf kingdom airstrikes strike province eastern area rich oil gas milit action chaoti...

Saudi airstrikes in Yemen target rebel stronghold Shiite rebels, known as Houthis, hold up their weapons to protest against Saudi-led airstrikes, during a rally in Sanaa, Yemen, on March 26, 2015. (P...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

22)USA_TODAY h id:15:331264875:100.0:

[Saudi-led air campaign blunts rebel advance in Yemen](#)

saudi-led air campaign blunts advance yemen airstrikes rebel group aden washington saudi-led milit coalition advance iran-backed rebel yemen regional coalition naval force country port rebel saudi brig gen asiri airstrikes houthi rebel aden stronghold -backed government power analyst saudi-led air ...

Saudi-led air campaign blunts rebel advance in Yemen Saudi airstrikes prevent rebel groups from taking Aden WASHINGTON ? A Saudi-led military coalition blunted the advance of Iranian-backed rebels in...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

23)NYTIMES h id:16:327004959:100.0:

[Op-Ed Contributors: The Kurds? Heroic Stand Against ISIS](#)

erbil iraq islamic control huge syria iraq advance shiite milit iran ally fight islamic kurd
640-mile front islamic advance steadfastness america alliance interest region relationship
kurd largest stateless nation sunni town tikrit saddam hussein force iraq iran offensive
mosul iraq largest city ar...

ERBIL, Iraq ? THE Islamic State continues to control a huge section of Syria. But in Iraq , its
advance has stalled. While Shiite militias and their Iranian allies fight the Islamic State
ferociously...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

24)NYTIMES h id:17:327257837:100.0:

[Iran Sent Arms to Iraq to Fight ISIS, U.S. Says](#)

iran arm isil intelligence official tehran credit atta kenare agence france-presse getty iran
advanced rocket missile fight islamic tikrit escalation firepower sign iran influence iraq
intelligence agency deployment week iraq force troop two-thirds milit iran america official
official condition ano...

Iran Sent Arms to Iraq to Fight ISIS, U.S. Says United States intelligence officials say that
Fateh-110 missiles, like this one displayed in 2010 in Tehran, were sent to Iraq. Credit Atta
Kenare/Agen...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

25)NYTIMES s id:8:327272271:0.0:

[Lens Blog: Moved by Mountains and the Sea in Indonesia](#)

mountain sea indonesia lens photography video visual journalism reason project men sea
trip yogyakarta city java indonesia photography idea project nearby mountain curiosity
simple feature connect environment indonesia muslim belief paganism faith personal
connection mountain sea zakaria interested...

Moved by Mountains and the Sea in Indonesia Lens: Photography, Video and Visual

Journalism And that is one of the reasons his project ?Men, Mountains and the Sea? has taken this long. The work came o...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

26)NYTIMES h id:18:327591924:100.0:

[Brigade Takes on ISIS Allies in Libya](#)

brigade isil ally libya cairo fight islamic control libya city surt brigade neighboring city misurata report tunisia milit battle report sign milit effort islamic group surt month libya month conflict rival coalition milit region ideology vacuum islamic isil isil group fight libya province loyalty ...

Brigade Takes on ISIS Allies in Libya CAIRO ? Fighters aligned with the Islamic State who control the Libyan city of Surt have begun clashing with a brigade from the neighboring city of Misurata that...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

27)NYTIMES h id:19:327607184:100.0:

[Op-Ed Columnist: Go Ahead, Ruin My Day](#)

human forgive divine add human divine human wish notice outright fact ground laughingstock hope gap wide policy change policy middle east gap iran israel iraq choice country strike ala grandfather father iraq son daughter israel prime minister benjamin netanyahu likud labor leader isaac herzog race...

As the saying goes, ?to err is human, to forgive is divine,? to which I'd add: ?to ignore? is even more human, and the results rarely divine. None of us would be human if we didn?t occasionally get s...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

28)NYTIMES h id:20:327703131:100.0:

[Afghan First Vice President, an Ex-Warlord, Fumes on the Sidelines](#)

kabul security meeting vice president afghanistan warlord abdul dostum jarring somber national security council dostum mass hundred taliban prisoner roomful powerful president ashraf ghani national security adviser atmar return call gathering week country general advice scene odd transformation dos...

KABUL, Afghanistan ? The security meeting was almost at a close when the first vice president of Afghanistan , the former warlord Abdul Rashid Dostum , began crying. It was a jarring ending to the ty...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

29)NYTIMES s id:9:331300682:0.0:

[World Briefing: Uzbekistan: President Is Re-elected](#)

president islam karimov term landslide preliminary karimov percent vote percent ballot central election commission statement karimov soviet republic opposition organization security cooperation europe election foregone conclusion genuine political alternative karimov president fourth consecutive te...

President Islam Karimov has won another term in a landslide, as expected, according to preliminary results issued Monday. Mr. Karimov garnered 91 percent of the vote, with about 91 percent of the bal...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

30)NYTIMES s id:10:331527663:8.0:

[15 Bodies Pulled From Home Hit by Landslide in Northern India](#)

delhi rescue worker body hit landslide torrential rain india-administered jammu kashmir north official landslide cluster house budgam district dead heavy rain widespread flooding northern india jhelum river dangerous level water street srinagar capital water worker resident area official body famil...

NEW DELHI ? Rescue workers have recovered 15 bodies from a house that collapsed after

being hit by a landslide caused by torrential rains in the Indian-administered state of Jammu and Kashmir in the ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

31)NYTIMES s id:11:332850791:0.0:

[Op-Ed Columnist: Interview With a Christian](#)

debate religious freedom unfold interview imaginary representative member press conversation easter semi-reasonable christian terrible indiaa religious liberty bill original version experience law protect minority burden indiaa law vendor same-sex wedding protection business comer service reason re...

AFTER watching the debate about religious freedom unfold over the past week, I decided to subject myself to an interview by an imaginary ? but representative ? member of the press. Here is our conver...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

32)NYTIMES s id:12:333121315:0.0:

[Shilpa Patel and Miten Shah](#)

shilpa patel ketan shah parsippany pandit harendra dave hindu priest hilton parsippany rutgers bride degree public health psychology groom political science genetics master social study education bride program manager center study asian america health nyu langone medical center master public health...

Shilpa Patel and Miten Ketan Shah were married Saturday in Parsippany, N.J. Pandit Harendra R. Dave, a Hindu priest, officiated at the Hilton Parsippany. The couple, both 33, graduated from Rutgers, ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

33)NYTIMES s id:13:333795238:1.0:

[C.J. Wilson Throws 8 Sharp Inning, Angels Beat Mariners 2-0](#)

baseball wilson sharp inning angel mariner wilson sharp inning angel mariner seattle
wilson hit inning freese offense two-run homer los angel win seattle mariner wilson
outstanding wriggling second-inning jam final batter left-hander baserunners strikeout
seattle nelson cruz rickie week hit mariner...

Baseball | C.J. Wilson Throws 8 Sharp Inning, Angels Beat Mariners 2-0 C.J. Wilson Throws
8 Sharp Inning, Angels Beat Mariners 2-0 SEATTLE ? C.J. Wilson allowed two hits over eight
innings and David ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

34)WP h id:21:328019729:100.0:

[Tunisia promises ?merciless war? against terrorism](#)

tunisia merciless war terror tunis tunisia president country merciless war terror gunman
foreign tourist tunisia daylight attack tunisia museum assault stability lone country arab
spring success global condemnation world-famous national bardo museum tunis tunisia
prime minister habib gunman attack ...

Tunisia promises ?merciless war? against terrorism TUNIS, Tunisia ? Tunisia?s president
vowed Thursday to expand the country?s ?merciless war against terrorism? after gunmen
killed at least 18 foreig...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

35)WP h id:22:328313553:100.0:

[WorldViews: Petraeus: The Islamic State isn?t our biggest problem in Iraq](#)

petraeus islamic biggest problem iraq january file photo cia director petraeus testifies
intelligence committee committee hearing wide threat capitol hill washington dc karen
bleier afp getty general david troop surge iraq annual sulaimani forum get-together iraq
leader thinker academic america uni...

Petraeus: The Islamic State isn?t our biggest problem in Iraq In this January 31, 2012 file

photo, CIA Director David Petraeus, testifies before the US Senate Intelligence Committee during a full com...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

36)WP h id:23:328353441:100.0:

[Israel?s threat from within](#)

israel appearance cbs nation israel minister benjamin netanyahu white column prediction iran nuclear program netanyahu supreme leader ayatollah khamenei tweet way reason israel netanyahu attention threat reason clear iran supreme leader radical anti-western ideologue twitter hostility israel harm c...

Israel?s threat from within In an appearance on CBS News?s ? Face the Nation ? soon before he was reelected as Israel?s prime minister, Benjamin Netanyahu was asked if he was offended that the White ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

37)WP h id:24:328380276:100.0:

[Security bolstered across Tunisia as rallies decry reach of terrorism](#)

security tunisia rally reach terror tunis tunisia authority security vital installation country fear follow-up attack museum massacre islamic gunman security force site main mediterranean port capital headquarters radio threat broadcaster interior ministry spokesman police aware specific threat rad...

Security bolstered across Tunisia as rallies decry reach of terrorism TUNIS ? Tunisian authorities bolstered security at vital installations across the country Friday amid fears of follow-up attacks ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

38)WP h id:25:328491452:100.0:

[After Netanyahu win, Palestinians vow to press case in international arena](#)

netanyahu win palestine press case international arena ramallah west bank reelection israel prime minister benjamin netanyahu palestine leader move embarrass prosecute israel international arena risky disengagement decades-old security cooperation israel west bank palestine official spent israel co...

After Netanyahu win, Palestinians vow to press case in international arena RAMALLAH, West Bank ? Following the reelection of Israeli Prime Minister Benjamin Netanyahu, Palestinian leaders say they wi...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

39)WP s id:14:332234122:0.0:

[Chart: There will be almost as many Muslims as Christians in the world by 2050](#)

chart muslim christian study pew center population major religion decade finding islam fastest religion global population muslim christian chart project major undertaking pew comprehensive demographic analysis thousand data set census population register lead researcher conrad hackett tremendous am...

Chart: There will be almost as many Muslims as Christians in the world by 2050 A new study released Thursday by the Pew Research Center projected the populations of the world's major religions over t...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

40)WP s id:15:335307759:5.0:

[Paris set to bid on 2024 Olympics](#)

paris bid olympics miguel medina afp getty city hosting future olympics massive cost return official paris step bidding game city council resolution endeavor prospective bid federal government press formal intention international olympic committee sept paris game unsuccessful bid olympics city bid ...

Paris set to bid on 2024 Olympics (Miguel Medina/AFP/Getty Images) While many other

cities have balked at hosting future Olympics, citing massive costs and limited returns, officials in Paris announc...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

41)DAWN_PK s id:16:327118924:0.0:

[PTI to hold 'election rigging' rally in Lahore](#)

pti election rigging lahore lahore pakistan pti protest rally lahore general election dawnnews detail stage protest rally tribunal decision pti chairman imran muslim ayaz sadiq public government judicial commission election protest movement lahore sit-in staged parliament islamabad zone spokesperso...

PTI to hold 'election rigging' rally in Lahore. LAHORE: Pakistan Tehreek-i-Insaf (PTI) will soon hold a protest rally in Lahore against rigging in the 2013 general election, DawnNews reported on Mond...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

42)DAWN_PK h id:26:327327277:100.0:

[Israelis vote in tight race after last-ditch Netanyahu plea](#)

jerusalem israel election close-fought battle centre prime minister benjamin netanyahu palestine last-ditch appeal outcome race opinion netanyahu fewer seat centre-left zionist union determine prospect middle east peace talk israel relation voter eligible election israel member parliament poll pm g...

JERUSALEM: Israelis voted Tuesday in an election expected to be a close-fought battle between the centre left and Prime Minister Benjamin Netanyahu, who ruled out a Palestinian state in a last-ditch ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

43)DAWN_PK s id:17:327917517:4.0:

[Elected from Islamabad, Senator Rahila returns to Sindh](#)

islamabad senator rahila return hyderabad pakistan muslim league-nawaz senator dr rahila magsi upper parliament islamabad nomination paper sindh ruler president ali zardari senate campaign support sugar cane grower reporter tando allahyar press club worker jamshoro toll plaza karachi member upper m...

Elected from Islamabad, Senator Rahila returns to Sindh. HYDERABAD: Pakistan Muslim League-Nawaz Senator Dr Rahila Magsi ? who has recently been elected to the upper house of parliament from Islamaba...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

44)DAWN_PK s id:18:328434684:0.0:

[Govt agrees on judicial commission to probe election rigging](#)

govt judicial commission election islamabad pakistan muslim league nawaz pml-n form long-awaited judicial commission pakistan tehreek-i-insaf pti request irregularity general election joint press conference leader party consensus term reference discussion numerous month finance minister ishaq dar p...

Govt agrees on judicial commission to probe election rigging. ISLAMABAD: Pakistan Muslim League Nawaz (PML-N) agreed to form the long-awaited judicial commission ? on Pakistan Tehreek-i-Insaf?s (PTI)...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

45)DAWN_PK s id:19:328746569:1.0:

[ANP, JUI-F assail govt-PTI deal](#)

anp jui-f assail govt-pti deal finance minister ishaq dar contact leader party confidence agreement government pti formation judicial commission alleged poll anp jui-f reservation development finance ministry announcement dar spoke leader political party leader ppp muttahida qaumi movement mqm jama...

ANP, JUI-F assail govt-PTI deal. ISLAMABAD: As Finance Minister Ishaq Dar initiated

contacts with leaders of various parties on Saturday to take them into confidence over the agreement between the go...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

46)DAWN_PK s id:20:329491180:7.0:

[PA committee to probe illegal appointments case](#)

pa committee illegal appointment case peshawar khyber pakhtunkhwa matter alleged illegal appointment previous government pakhtunkhwa energy development organisation relevant concern probe pti mpa yaseen khalil matter relevant committee investigation question speaker asad qaisar announcement questio...

PA committee to probe illegal appointments case. PESHAWAR: The Khyber Pakhtunkhwa Assembly on Tuesday referred the matter of the alleged illegal appointments by the previous government to the Pakhtun...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

47)DAWN_PK s id:21:329515960:0.0:

[Permission from earlier wives not required for new marriage: Sherani](#)

permission earlier wife marriage sherani islamabad gap month chairman council islamic ideology cii maulana mohammad sherani talked-about topic marriage divorce woman reporter cci meeting maulana sherani men permission wife previous wife marriage anarchy discontent garb woman right consecutive decla...

Permission from earlier wives not required for new marriage: Sherani. ISLAMABAD: After a gap of few months the chairman of Council of Islamic Ideology (CII), Maulana Mohammad Khan Sherani, has spoken...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

48)DAWN_PK h id:27:329866510:100.0:

[Pakistan examining Saudi request to join anti-Houthi operation in Yemen: FO](#)

islamabad pakistan saudi arab gulf-led operation shia houthi rebel yemen foreign office spokesperson tasneem aslam reporter weekly medium islamabad saudi arab pakistan emergency basis invitation operation houthi rebel yemen detail request saudi matter aslam pakistan mission yemen alert touch pakist...

ISLAMABAD: Pakistan said on Thursday it was examining Saudi Arabia's request to join the Gulf-led operation against Shia Houthi rebels in Yemen. Foreign Office spokesperson Tasneem Aslam told reporte...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

49)DAWN_PK h id:28:329866515:100.0:

[Saudi Arabia, allies launch air strikes against Houthis in Yemen](#)

washington saudi arab muslim country egypt pakistan gulf-led milit coalition shia houthi rebel yemen jordan morocco sudan desire operation rebel kingdom firmness storm saudi spa agency saudi arab gulf bahrain kuwait qatar united arab emirate earlier decision president hadi yemen aggression shia hou...

WASHINGTON: Saudi Arabia said Thursday that five Muslim countries including Egypt and Pakistan want to participate in the Gulf-led military coalition against Shia Houthi rebels in Yemen. Together wit...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

50)DAWN_PK h id:29:329866573:100.0:

[IS sending advance guards to region: Ghani](#)

advance guard region ghani washington president ashraf ghani so-called islamic milit advance guard pak-afghanistan region ass vulnerability address ghani america public milit sacrifice country terror movement goal region base operation troop afghanistan obama afghanistan frontline terror boundary p...

IS sending advance guards to region: Ghani. WASHINGTON: Afghan President Ashraf Ghani

said on Wednesday that the so-called Islamic State militants were sending advance guards to the Pak-Afghan region...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

51)DAWN_PK h id:30:329898888:100.0:

[Iran condemns Saudi strikes in Yemen as 'dangerous step'](#)

tehran iran regional rival saudi arab launching air huthi rebel yemen dangerous step international responsibility national sovereignty foreign ministry spokeswoman marzieh afkham milit action situation spread crisil opportunity peaceful resolution yemen internal difference aggression result terror ...

TEHRAN: Iran condemned regional rival Saudi Arabia on Thursday for launching air strikes on Huthi rebels in Yemen, saying it was a dangerous step that violated international responsibilities and n...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

52)REUTERS h id:31:327068449:100.0:

[Iraq needs more air strikes to dislodge IS in Tikrit: officials](#)

iraq air strike tikrit official maggie fick chlorine-tinged cloud smoke rise air bomb iraqarmhi ite fight hashid shaabi force town al-alam salahuddin province march credit thailander al-sudani baghdad iraq tikrit offensive senior official air strike islamic milit explosive saddam hussein city centr...

Iraq needs more air strikes to dislodge IS in Tikrit: officials By Maggie Fick A chlorine-tinged cloud of smoke rises into the air from a bomb detonated by Iraqi army and Shi'ite fighters from Hashid...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

53)REUTERS h id:32:327171803:100.0:

[U.N. concerned by Islamic State's ability to unite Afghan insurgents](#)

islamic ability insurgent nicholas listens question conference kabul august credit sobhani
nation nation presence islamic afghanistan milit group power insurgent capability war-
torn country top official envoy nicholas security council afghanistan attempt broker
conflict taliban -led war foreign for...

U.N. concerned by Islamic State's ability to unite Afghan insurgents Nicholas Haysom
listens to a question during a news conference in Kabul August 8, 2012. Credit:
Reuters/Omar Sobhani UNITED NATION...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

54)REUTERS h id:33:327445449:100.0:

[Tunisia dismantles jihadi recruiting cell for Libya](#)

tunisia dismantles cell libya tunis tunisia recruiting cell jihad fight libya dozen tighter
security border control islamist milit violence libya islamic milit influence turmoil rival
government control tunisia largest contingent foreign fight syria iraq jihad conflict tunisia
north africa neighbor...

Tunisia dismantles jihadi recruiting cell for Libya TUNIS (Reuters) - Tunisia said on
Tuesday it had dismantled a recruiting cell sending jihadis to fight in Libya and arrested
dozens in part of tigh...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

55)REUTERS h id:34:327733838:100.0:

[Downing of U.S. drone suggests Syria imposing red lines on air war](#)

drone syria line air war beirut allowing air space islamic fight month syriaarmhooting
drone territory critical milit contact drone northwest syria incident latakia province
western region syria damascus control syriaarmource aircraft government-held territory
area islamic jihadt group daily bombar...

Downing of U.S. drone suggests Syria imposing red lines on air war BEIRUT (Reuters) -

After allowing the United States to use its air space to bomb Islamic State fighters for six months, the Syrian a...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

56)REUTERS h id:35:327900628:100.0:

[Gunmen storm Tunisian museum, kill 17 foreign tourists](#)

gunman tunisia museum kill foreign tourist tarek amara argoubi tunis mar edt tunis
gunman milit uniform tunisia national museum foreign tourist tunisia worst milit attack
country region arab spring turmoil japan visitor spain dead noon assault bardo museum
inside heavily parliament compound central...

Gunmen storm Tunisian museum, kill 17 foreign tourists By Tarek Amara and Mohamed Argoubi TUNIS Wed Mar 18, 2015 8:53pm EDT TUNIS (Reuters) - Gunmen wearing military uniforms stormed Tunisia's nation...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

57)REUTERS s id:22:327980872:0.0:

[Dutch ruling parties pledge to continue despite Senate loss](#)

holland ruling party senate loss thomas netherlands prime minister mark rutte gesture
europe twin challenge growth stability event switzerland mountain resort davos january
amsterdam leader holland ruling coalition government provincial election majority country
senate local ballot determines compo...

Dutch ruling parties pledge to continue despite Senate loss By Thomas Escritt Netherlands' Prime Minister Mark Rutte gestures during the Europe's Twin Challenges: Growth and Stability event in the Sw...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

58)REUTERS s id:23:328903864:4.0:

[Local bill on Muslim tribunal in Texas stirs sentiments](#)

irving bill foreign law court stir local islamic tribunal resolving civil city council irving texas
bill muslim islamic tribunal bill foreign law case sharia law court law case soundbite
england mayor beth van duyne brainier resolution constitution constitution straight
forward constitutional issue...

An Irving, Texas bill that says foreign law does not apply to U.S. courts stirs sentiment surrounding a local Islamic tribunal aimed at resolving civil disputes. A city council in Irving, Texas, has ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

59)REUTERS s id:24:330643418:4.0:

[Cosby show brings out protesters in Baltimore](#)

protester cosby goer baltimore bill cosby stage baltimore comedy 77-year-old protester
crowd standup routine demonstrator veteran comedian dozen woman soundbite katherine
bruce cosby protester mouth soundbite rachel perry-cook cosby protester reflects woman
teenager sunni welles accuser bill cosby ...

Protesters turn out to shame Cosby show goers in Baltimore. Bill Cosby takes the stage in Baltimore for his comedy show. The 77-year-old was heckled by several protesters in the crowd during his stan...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

60)FOX_NEWS h id:36:326855291:100.0:

[American intelligence analysts disagree about impact of airstrikes on al-Qaida cell in Syria](#)

america intelligence analyst impact airstrikes al-qaeda cell syria file sept file photoarmgen
william mayville jr director operation operation syria conference pentagon milit separate
target al-qaeda cell syria khorasan group official ongoing little-discussed air campaign
disrupting group capacity ...

American intelligence analysts disagree about impact of airstrikes on al-Qaida cell in Syria

FILE - In this Sept. 23, 2014, file photo, Army Lt. Gen. William Mayville, Jr., Director of Operations J-3...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

61)FOX_NEWS s id:25:326878520:2.0:

[NYC's plan to allow prayer breaks in pre-kindergarten classes raises church-state concerns](#)

nyc plan prayer break pre-kindergarten class church-state concern joey michael skaba rooftop playground sonny gindi barkai yeshivah brooklyn march jajati michael skaba rooftop playground sonny gindi barkai yeshivah brooklyn march elyse dweck joey raise pre-kindergarten sonny gindi barkai yeshivah b...

NYC's plan to allow prayer breaks in pre-kindergarten classes raises church-state concerns
Joey Arking, 5, left, Moshe Lib, 4, second from left, Michael Skaba, 4, second from right and Leo Jajati, 4,...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

62)FOX_NEWS h id:37:327086950:99.0:

[March Madness discussions ? so much more fun to talk and think about than Iran and nuclear weapons, right?](#)

pic scene otr interview senator lee senator marco rubio video step elevator nightmare vietnam march madness discussion iran nuclear weapon samaritan short heart determination click pic sudan pic pic tank evil president bashir milit nuba innocent campaign manager america israel pm netanyahu south af...

PICS: Behind the Scenes of OTR's interview with Senator Mike Lee and Senator Marco Rubio
VIDEO: Walking up remaining steps after getting elevator nightmare in Vietnam
March Madness discussions ? so m...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

63)FOX_NEWS h id:38:327109217:100.0:

[US, Iran working for nuclear pact; officials suggest a lesser announcement possible](#)

secretary john kerry break bilateral meeting iran foreign minister mohammad javad iran nuclear program lausanne switzerland march united iran negotiation effort decades-long standoff specter iran nuclear arsenal atomic arm race middle east israel milit intervention ap keystone jean-christophe bott ...

U.S. Secretary of State John Kerry takes a break during a bilateral meeting with Iranian Foreign Minister Mohammad Javad Zarif over Iran's nuclear program in Lausanne, Switzerland, Monday, March 16, ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

64)FOX_NEWS h id:39:327109231:100.0:

[Kurds probe 2 possible ISIS chemical weapon attacks](#)

march young volunteer militman battlefield islamic fight tikrit mile north baghdad iraq kurdish force iraq isil chemical weapon attack top official authority iraq offensive saddam hussein irbil iraq kurdish force iraq chemical weapon attack islamic group top official authority iraq offensive saddam...

March 15, 2015 - A young volunteer militiaman on his way to the battlefield against Islamic State fighters in Tikrit, 80 miles north of Baghdad, Iraq. Kurdish forces in Iraq are investigating 2 other...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

65)FOX_NEWS s id:26:327296488:0.0:

[Myanmar court jails New Zealand man for 2 years and 6 months for insulting Buddhism](#)

myanmar court jail zealand month insulting buddhism yangon myanmar myanmar court zealand bar manager business month prison buddhism flyer psychedelic depiction buddha headphone blackwood tun thurein htut ko ko lwin labor insulting religion month disobeying order public servant trial gastro manager ...

Myanmar court jails New Zealand man for 2 years and 6 months for insulting Buddhism
YANGON, Myanmar ? A Myanmar court has sentenced a New Zealand bar manager and his
business associates to two years ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

66)FOX_NEWS h id:40:327489007:100.0:

[Air Force vet charged with trying to join ISIS](#)

force vet isil america citizen force veteran airplane mechanic weapon system isil federal
prosecutor tairod webster asbury park jan plan black-clad jihadtarmauthority count
material support foreign terror organization obstruction justice attorney loretta lynch
release born country travel order terr...

Air Force vet charged with trying to join ISIS An American citizen and U.S. Air Force veteran
who worked as an airplane mechanic and was trained in weapons systems was charged
Tuesday with trying to ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

67)FOX_NEWS s id:27:327545407:5.0:

[Stunning meteor photo](#)

incredible bright tail meteor clear sky scotland ness absolute fluke alasdair macdonald tour
guide scotland highland rare shot wife sheer dumb luck complete fluke absolute fluke
macdonald offer private highland tour business hebridean explorer business facebook
social medium round evangelical chris...

It?s an incredible image ? the bright tail of a meteor streaking across the clear night sky
above Scotland?s Loch Ness. The image, which has been trending online, was an ?absolute
fluke,? according t...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

68)FOX_NEWS s id:28:329629022:0.0:

[Muslim groups rip Jeb Bush for consultant's anti-Shariah views](#)

muslim group jeb bush consultant anti-shariah view florida governor jeb bush talk medium
integra bioscience campaign hudson hampshire march stapleton tag health science
technology politics election rtr4tako coalition national muslim organization press release
jeb bush campaign-in-waiting senior con...

Muslim groups rip Jeb Bush for consultant's anti-Shariah views Former Florida Governor
Jeb Bush talks to the media after visiting Integra Biosciences during a campaign stop in
Hudson, New Hampshire M...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

69)FOX_NEWS s id:29:329735777:3.0:

[Ohio officer fined \\$500 after being convicted of misdemeanor K-9's heatstroke death](#)

ohio officer misdemeanor heatstroke death medina ohio municipal court judge northeast
ohio police officer fine finding guilty animal cruelty allowing police dog heatstroke cruiser
bench trial montville township police sgt brett harrison guilty count companion animal
cruelty guilty count harrison gu...

Ohio officer fined \$500 after being convicted of misdemeanor K-9's heatstroke death
MEDINA, Ohio ? A municipal court judge has ordered a northeast Ohio police officer to pay
a \$500 fine after finding...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

70)FOX_NEWS s id:30:330191027:0.0:

[Police say 10 people killed, dozens injured in a stampede at Hindu festival in Bangladesh](#)

police dozen stampede hindu festival bangladesh dhaka bangladesh dozen injured
stampede hindu religious gathering bangladesh police local police chief mazrul islam
accident langalbandh hindu pilgrimage spot bank brahmaputra river kilometer mile
southeast capital dhaka annual religious bathing ritua...

Police say 10 people killed, dozens injured in a stampede at Hindu festival in Bangladesh
DHAKA, Bangladesh ? At least 10 people were killed and dozens more injured Friday in a stampede during a Hind...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

71)THE_NEWS_PK h id:41:326982669:100.0:

[Is ISIS knocking on Karachis doors?](#)

sectarian fissure ethno-political polarisation karachi middle east-based jihadt group
islamic iraq syria isil network city worsen security analyst police official january isil daish
daulat-e-islamia pakistan taliban commander leader pakistan afghanistan hafiz chief
tehreek-e-taliban pakistan taliba...

Exploiting the sectarian fissures and ethno-political polarisation in Karachi, the Middle
East-based jihadist group the Islamic State of Iraq and Syria (ISIS) may organise its
network in the city and...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

72)THE_NEWS_PK s id:31:327023737:1.0:

[TOBA TEK SINGH City News](#)

arrest prayer leader sic strike faisalabad 18th toba tek singh sunni council chairman
sahibzada raza strike faisalabad march arrest sunni ulema province press conference
meeting sunni religious organisation score prayer leader case charge loudspeaker amplifier
arrest protest province roof cave roof...

Arrest of prayer leaders SIC to observe strike in Faisalabad on 18th TOBA TEK SINGH:
Sunni Ittehad Council chairman Sahibzada Hamid Raza has said that a strike will be
observed in Faisalabad on March...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

73)THE_NEWS_PK s id:32:327038787:0.0:

[ICU syndicates decision on appointment triggers concern](#)

yousaf ali peshawar selection board islamia college university person post verification panel interview syndicate approval exercise doubtful concern candidate case board position selection board september lapse selection board vice-chancellor prof dr qibla ayaz panel interview candidate recommendat...

Yousaf Ali PESHAWAR: Though the selection board of Islamia College University had okayed 42 persons for various posts after verification and panel interviews, the syndicate gave approval to around ha...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

74)THE_NEWS_PK s id:33:327049321:3.0:

[Anti-polio campaign from today](#)

anti-polio campaign three-day polio eradication campaign province health department team child age handout adviser chief minister health khawaja salman rafiq press statement health team three-day campaign mobile team polio vaccine child team entry exit point province permanent transit point termina...

Anti-polio campaign from today A three-day polio eradication campaign is starting in the province from today (Monday), during which Health Department teams will vaccinate 18.5 million children up to ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

75)THE_NEWS_PK h id:42:327257345:100.0:

[So what else should Christians do?](#)

islamabad diary home gojra provincial sarkar joseph colony lahore trumped-up pretence usual blasphemy tool passion mob government christian wife labourer brick kiln burning oven police christian church suicide attack lahore dead score road public property suspect guilt basis mob hearsay fire govern...

Islamabad diary When their homes were ransacked in Gojra, where was the provincial sarkar? When Joseph Colony in Lahore was ransacked on patently trumped-up pretences? the usual blasphemy tool emplo...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

76)THE_NEWS_PK s id:34:327495526:0.0:

[Oath-taking ceremony of student council held at Viqar-un-Nisa College](#)

oath-taking ceremony academic government viqar-un-nisa postgraduate college chief guest occasion abbasi aurangzeb mpa director college rawalpindi professor hamayun iqbal ceremony principal professor dr saira mufti oath member student council ex-members student council responsibility newly-elected m...

The oath-taking ceremony for the academic year 2015 was held here on Monday at Government Viqar-un-Nisa Postgraduate College. The chief guest on the occasion were Hanif Abbasi and Mrs. Tahira Aurangz...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

77)THE_NEWS_PK s id:35:327579425:1.0:

[Renovated orthopaedic department inaugurated at BBH](#)

orthopaedic department benazir hospital bbh simple ceremony renovation building fund bait-ul-mal director rawalpindi medical college principal professor dr muhammad umar chairman board management hospital muhammad aslam pakistan muslim league nawaz leader dr jamal nasir philanthropist sheikh ceremo...

The orthopaedic department was inaugurated at Benazir Bhutto Hospital (BBH) here on Tuesday in a simple ceremony after necessary renovation work on the building carried out through funds personally d...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

78)THE_NEWS_PK h id:43:327579370:100.0:

[Into the maze of global conflict](#)

view riyadh saudi mission syria awry isil force saudi obsession isil turn hold saudi crosshairs saud brand salafi islam heretical isil threat saudi north iraq isil large swathe saudi dominance salafi islam threat closer case shia bahrain eastern saudi arab shia crescent levant yemen reality riposte...

The view from Riyadh. The Saudi mission to topple Syria?s Assad has gone awry ? instead Isis is now a force to reckon with, and is the new Saudi obsession. Isis, in turn, holds the Saudis in its cros...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

79)THE_NEWS_PK s id:36:327578522:4.0:

[Eight injured in road accident](#)

injury van canal sundar police limit rescue team spot van lahore accident islam nagar stable condition wheelie city traffic police notice wheelie doer officer heavy bike wheelie handout chief traffic police officer cto tayyab hafiz cheema traffic police resource youth wheelie worth mentioning traff...

Eight people received injuries when a van fell into a canal near Sundar police limits on Tuesday. Rescue teams reached the spot and rescued the injured. A van was heading to Pattoki from Lahore when ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

80)THE_NEWS_PK h id:44:327989606:100.0:

[Only Moscow can stop it](#)

war russia choice department hadn coup ukraine president viktor yanukovych russia eu vital trade route nation continent economic integration asia europe plan high-speed rail china silk road eu clear danger share global gdp significance economy rival eu-russia equivalent towel future gradual persist...

The United States does not want a war with Russia, it simply feels that it has no choice. If the State Department hadn't initiated a coup in Ukraine to topple the elected president, Viktor Yanukovich...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

81)THE_NEWS_PK h id:45:328438370:100.0:

[Spring cleaning](#)

march mark gloomy anniversary 12th anniversary invasion iraq 5th anniversary nato intervention libya overthrew arab dictator local horrific strait nostalgia saddam hussein muammar qaddafi iraq dozen codepink colleague month invasion country draconian western sanction saddam hussein middle country w...

March 19 marks two gloomy anniversaries: the 12th anniversary of the US invasion of Iraq and the 5th anniversary of the Nato intervention in Libya. Both overthrew Arab dictators; both left the local ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

82)GUARDIAN_UK s id:37:326910180:0.0:

[Circular thinking: Stonehenge's origin is subject to new theory](#)

circular stonehenge subject theory wiltshire monument equivalent ancient mecca stilt forward museum director julian druid temple astronomical calendar centre healing mystery stonehenge source speculation debate dramatic theory prehistoric monument ancient mecca stilt megalith ceremony ground level ...

Circular thinking: Stonehenge's origin is subject to new theory. Wiltshire monument may have been equivalent of ?an ancient Mecca on stilts? according to an idea put forward by former museum director...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

83)GUARDIAN_UK s id:38:327190618:2.0:

[Fashion blogger Dina Torkia: ?There?s a fear factor around the hijab?](#)

fashion blogger dina torkia fear factor hijab hijabi blogger dina tokiu centre liberty bbc beauty pageant muslim woman zoella way torkia typical fashion blogger 25-year-old doll trend urban backdrop usual smattering self-promotion clothing range torkia blog tokiu requisite social medium stats follo...

Fashion blogger Dina Torkia: ?There?s a fear factor around the hijab?.Hijabi blogger ?Dina Tokio? started blogging while working in a call centre. Now she?s collaborating with Liberty and reporting f...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

84)GUARDIAN_UK s id:39:327258534:3.0:

[Don't say I'm oppressed because I'm an Arab woman. It denies me the right to my own experience](#)

arab woman experience generalisation dog whistle politics arab muslim woman vulnerable absolute generalisation binary rhetoric dog whistle politics team australia arab muslim muslim arab muslim arab terror terror arab woman subservient men arab muslim girl marriage stereotype public discourse decad...

Don't say I'm oppressed because I'm an Arab woman. It denies me the right to my own experience. We're living in a time of generalisations and dog whistle politics. But stereotyping Arab and Muslim wo...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

85)GUARDIAN_UK s id:40:327714035:0.0:

[Acts of the Assassins by Richard Beard review ? the Passion as police procedural](#)

act assassin richard beard review passion police procedural combine startling challenging hapless gumshoe fate jesus disciple christianity unusual religion way late century theologian christian christ single telling century canonical account account fewer life evangelist emphasis non-canonical acco...

Acts of the Assassins by Richard Beard review ? the Passion as police procedural. Past and present combine to startling effect in this challenging novel, in which a hapless gumshoe tries to unravel t...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

86)GUARDIAN_UK h id:46:327746734:100.0:

[Tunisia terror attack: eight dead after gunmen storm museum - live](#)

tunisia terror attack dead gunman storm museum tunisia foreign ministry foreigner tunisia attack bardo museum adjacent parliament building tunis gmt member security raid gunman officer security museum attack tunisia interior ministry ap gmt hostage tunisia force report official gmt al-jazeera inter...

Tunisia terror attack: eight dead after gunmen storm museum - live. Tunisian foreign ministry confirms seven foreigners and one Tunisian have been killed in an attack at the Bardo museum, adjacent to...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

87)GUARDIAN_UK h id:47:327760872:100.0:

[Binyamin Netanyahu victory causes international concern](#)

binyamin victory international concern binyamin netanyahu victory israel general election dismay nail coffin hope two-state solution israel-palestine conflict netanyahu eve-of-poll pledge creation independent palestine alongside israel arab citizen drove manipulative implacable israel election biny...

Binyamin Netanyahu victory causes international concern Binyamin Netanyahu's sweeping victory in Israel's general election is causing dismay internationally because it appears to be another nail in t...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

88)GUARDIAN_UK s id:41:328026120:1.0:

[Religion and money: is Islamic banking the way forward for Ethiopians?](#)

religion islamic banking forward ethiopia lack physical bank community access financial service ethical issue largest religion follower population muslim poverty islamic-compliant finance huge industry international talk access financial service country service islamic-compliant top agenda islamic-...

Religion and money: is Islamic banking the way forward for Ethiopians?. What happens when it?s not just a lack of physical banks preventing communities? access to financial services, but ethical issu...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

89)GUARDIAN_UK s id:42:328097463:1.0:

[The Tale of the Princess Kaguya is a joy but won't hold kids' attention](#)

tale princess kaguya joy kid attention tale princess kaguya tumultuous emotional journey attention-grabbing film hero tale princess kaguya latest release japan best-loved animation studio studio ghibli critical acclaim film festival circuit academy award kid doubt matinee screening fidgety throb b...

The Tale of the Princess Kaguya is a joy but won't hold kids' attention. The Tale of the Princess Kaguya is a ?tumultuous emotional journey? yet will struggle to compete with attention-grabbing films...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

90)GUARDIAN_UK h id:48:328139175:100.0:

[Create UN force to protect ancient heritage from Isis, says Italy](#)

create force ancient heritage isil culture minister creation force heritage site destruction ancient city iraq islamic isil milit sort blue helmet culture blue helmet intervene situation war dario franceschini referring signature uniform troop international rapid response force monument archaeologi...

Create UN force to protect ancient heritage from Isis, says Italy Italy?s culture minister has called for the creation of a UN peacekeeping force to protect the world?s heritage sites, following the ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

91)GUARDIAN_UK h id:49:328342745:100.0:

[Tunis gunmen trained with Libyan militia, says security chief](#)

tunis libya milit security chief gunman attack foreign tourist tunis museum training milit camp libya tunisia secretary security country december libya weapon rafik private hiwar ettounsi television channel gunman security force official yassine laabidi hatem khachnaoui laabidi detail view kill tun...

Tunis gunmen trained with Libyan militia, says security chief The two gunmen who killed 21 people in an attack on foreign tourists at a Tunis museum had received training at a militia camp in Libya, ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

92)GUARDIAN_UK h id:50:328373389:100.0:

[Drones aren?t just toys that cause a nuisance. They?re still killing innocent people](#)

drone nuisance innocent drone aircraft industry calling nature coverage normalisation milit drone medium interest interested writing toy drone white lawn white drone assassination arrest amateur pilot drone parliament britain drone syria assurance defence secretary michael fallon mission creep atte...

Drones aren?t just toys that cause a nuisance. They?re still killing innocent people F or anyone concerned about the use of drones, or remotely piloted aircraft as the industry insists on calling the...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

93)TIMES_OF_INDIA s id:43:326868986:0.0:

[Subramanian Swamy defends mosque 'not a religious place' remark](#)

mosque religious subramanian mosque religious subramanian swamy bjp leader
subramanian swamy massive contr subramanian swamy tharoor conspiracy subramanian
swamy shashi tharoor probe tharoor mehr tarar angle subramanian swamy cop janata bjp
leader subramanian swamy challe shashi tharoor sunanda pus...

Mosque is not a religious place, can be demolished anytime: Subramanian... Mosque is not a
religious place, can be demolished anytime: Subramanian Swamy BJP leader Subramanian
Swamy has sparked off a...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

94)TIMES_OF_INDIA s id:44:326931884:3.0:

[CM advocates mall for women SHGs](#)

woman move sudhir mungantiwar sanjay dhote jal yukt shivar chandrapur chief minister
devendra building mall woman group shgs market good rajura tehsil headquarter
programme woman shgs shop mall rotation basis group month good benefit move woman
shgs awareness drive government jal yukt shivar scheme...

Women | The move | Sudhir Mungantiwar Fadnavis | Sanjay Dhote | Jal Yukt Shivar
Chandrapur: Chief minister Devendra Fadnavis on Sunday advocated the idea for building a
mall exclusively for women sel...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

95)TIMES_OF_INDIA s id:45:326946850:0.0:

[JE vaccination drive to cover all 27 dists in Assam](#)

health department health minister japan encephalitis health minister elisa guwahati assam
country district japan encephalitis je vaccination child life casualty infant child adult health
department disease control health department laboratory district essential elisa kit
diagnosis japan encephaliti...

state health department | Says Health Minister | Japanese Encephalitis | Health Minister | Elisa Guwahati: Assam is all set to become the first state in the country to have all 27 of its districts co...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

96)TIMES_OF_INDIA s id:46:326960154:3.0:

[BJP keeps off Swamy's shrine remark; Gogoi sees red](#)

subramanian swamy prakash javadekar controversial remark chatra mukti sangram samiti bjp bjp leader union minister prakash javadekar sought distance colleague subramanian swamy controversial comment mumbai bjp leader union minister prakash javadekar sought distance colleague subramanian swamy contr...

Subramanian Swamy | Prakash Javadekar | controversial remarks | Chatra Mukti Sangram Samiti | BJP ?BJP?leader and Union Minister?Prakash?Javadekar?on Sunday sought to distance himself from his party ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

97)TIMES_OF_INDIA h id:51:326991974:100.0:

[US, Iran work for nuclear pact as deadline approaches](#)

nuclear deal mohammad javad lausanne john kerry secretary john kerry arrives geneva march iran foreign minister mohammad javad nuclear talk afp lausanne switzerland united iran negotiation effort decades-long standoff specter iran nuclear arsenal atomic arm race middle east milit intervention week ...

united states | Nuclear Deal | Mohammad Javad Zarif | LAUSANNE | John Kerry US secretary of state John Kerry arrives in Geneva on March 15, 2015 to meet Iranian Foreign Minister Mohammad Javad Zarif ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

98)TIMES_OF_INDIA s id:47:327200834:2.0:

[ASI-protected 14th century tomb gets a whitewash](#)

asi-protected 14th century tomb whitewash maria akram maria akram tnn defaced- tomb khan inside mehrauli archaeological park toi rajesh mehta delhi 14th century monument tomb khan inside mehrauli archaeological park monument archaeological survey india white smaller monument park monument official ...

ASI-protected 14th century tomb gets a whitewash Maria Akram Maria Akram ,TNN Defaced- Tomb of Khan Shahid, located inside the Mehrauli Archaeological Park. (TOI photo by Rajesh Mehta) NEW DELHI: A 1...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

99)TIMES_OF_INDIA s id:48:327235540:0.0:

[4 new H1N1 cases in Assam](#)

situation swine flu nazrul islam health minister guwahati kamrup metro district h1n1 positive influenza virus case total number h1n1 positive case assam stand screening government facility february h1n1 positive patient district prior health condition diabetes heavy blood pressure disease h1n1 case...

The situation | Swine flu | Nazrul Islam | heart | Health Minister Guwahati: The Kamrup (Metro) district reported three new H1N1 positive (influenza A virus) cases in the past 24 hours. The total num...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

100)TIMES_OF_INDIA h id:52:328082767:100.0:

[Barack Obama slams Benjamin Netanyahu's 'deeply divisive rhetoric' as US takes Iran and Hezbollah off terror list](#)

barack obama slam netanyahu divisive rhetoric iran hezbollah terror victoria richards richards independent zionist union terror obama-netanyahu spat israel election iran terror israel pm benjamin netanyahu delivers speech wife sara poll figure israel parliamentary

election march tel aviv netanyahu ...

Barack Obama slams Benjamin Netanyahu's 'deeply divisive rhetoric' as US takes Iran and Hezbollah off terror list Victoria Richards Victoria Richards ,The Independent Zionist Union | US terror list |...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

101)TIMES_OF_INDIA h id:53:328566015:100.0:

[Retreating Boko Haram leaves mass of throat-slit corpses near Nigerian town](#)

boko haram mass throat-slit corps nigeria town damasak boko haram milit haram thousand six-year insurgency establishing islamic caliphate northeast nigeria damasak nigeria soldier niger nigeria town damasak boko haram milit body throat bridge witness execution site islamist group body beneath concr...

Retreating Boko Haram leaves mass of throat-slit corpses near Nigerian town Damasak | Boko Haram militants Boko Haram has killed thousands of people in a six-year insurgency aimed at establishing an ...

- Conflict
- Non-Conflict
- Not-sure
- No Answer Recorded

--

102)TIMES_OF_INDIA h id:54:328725244:100.0:

[Shia rebels call for Yemen offensive; US troops evacuate](#)

shia rebel yemen evacuate yemen shia rebel al-qaeda militman loyal president rabbo hadi ride tank street aden yemen yemen yemen rebel arm force loyal country president troop southern air base crucial drone strike program al-qaeda milit nearby city turmoil yemen battle arab peninsula target drone pr...

Shia rebels call for Yemen offensive; US troops evacuate Yemen | Shia rebels | al-Qaida Militiamen loyal to President Abed Rabbo Mansour Hadi ride on a tank on a street in Aden, Yemen. ADEN, Yemen: Y...

- Conflict
- Non-Conflict
- Not-sure

No Answer Recorded

--

103)TIMES_OF_INDIA h id:55:328757204:100.0:

[IS posts pictures of 100 US military on kill list](#)

milit fight flag travel vehicle milit parade street syria raqqa province washington islamic name photo america milit service member brother america pentagon internet matter confirm validity defense official condition anonymity personnel opsec operation security force protection procedure official i...

Militant Islamist fighters waving flags, travel in vehicles as they take part in a military parade along the streets of Syria's northern Raqqa province. (Reuters file photo)

WASHINGTON: Islamic State...

Conflict

Non-Conflict

Not-sure

No Answer Recorded

--