ISSN 2278-3091
Volume 9 No.2, March -April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse225922020.pdf
https://doi.org/10.30534/ijatcse/2020/225922020

Intelligent Framework for Long-text Political Speeches Summarization and Visualization Using Sentiment Lexicons: A Study Directed at King Abdullah II Discussion Papers

Zaher Salah[1], Waleed T. Al-Sit[2], Hala Hani[3], Kamal Salah[4].
[1,3]Prince Al Hussein Bin Abdullah II faculty for Information Technology, Hashemite University,
P.O. Box 150459, Zarqa 13115, Jordan.
[2]Department of Computer Engineering, Mu'tah University, Al-Karak - Jordan
[4] Deanship of preparatory year and supporting studies, Imam Abdulrahman Bin Faisal University,
P.O Box 1982, Dammam, Saudi Arabia
[1]zahersalah@hotmail.com, [2]w_sitt@hotmail.com,[3]hala.hani1992.hh@gmail.com, [4]kisalah@iau.edu.sa.

ABSTRACT

In this paper the authors seek to propose a mechanism for conducting text summarization and visualisation with respect to long-text political speeches so as to simply their outcomes. To this end artificial intelligence approach are considered using sentiment analysis based on sentiment lexicon (SentiWordNet 3.0). Sentiment analysis and visualization approaches and supporting techniques is conducted by predicting the sentiment polarity of individual speech in political speeches (discussion paper) using transcripts taken from King Abdullah II Official Website. The resulted text visualizations indicate that the attitude of each speech can be effectively predicted and summarized using sentiment analysis and text visualization techniques. The authors then go on to consider the differences between the attitudes of consecutive speeches and to highlight the evolution of the sentiment polarity among the speeches using statistical linguistic analysis for political jargon that is a feature of the political speeches under the study.

Key words: Artificial Intelligence, Text Mining, Sentiment Analysis, Text Summarization, Text visualization.

1. INTRODUCTION

Text Summarization and Visualization is an approach to convert text to simplified images that make long-texts simpler to understand and to save time for researchers to search for specific work-related tasks more easily and efficiently. Sentiment analysis is an application of natural language processing for tracking the mood of the public about a particular product or topic or trend. Sentiment analysis, which is also called opinion mining, involves building a system to collect and examine opinions concerning some object of interest like products made in blog posts, comments, reviews, or tweets [1, 24, 25].

This paper presents a framework combines the sentiment and lexical contents in the context of long-text visualization for the seven discussion papers of His Majesty King Abdullah II Ibn Al Hussein. The seven discussion paper are long-text political speeches discussing various related topics, it also discusses the general challenges in the country. The research work described in this paper proposes a framework for visualizing long-text speech as a graph that summarizes transcripts of political speeches. The objective was to deploy sentiment analysis techniques for the extraction of speech graphs that will in turn allow for the graphical visualisation of the high level structure of this kind of long-text speeches showing the evolution and the distribution of sentiment intensity and polarity of the subjective text (positive and negative feelings) embedded within that individual speeches in order to automatically identify the subjectivity and orientation of text segments and to extract political attitudes or viewpoints from that transcripts. The operation of the framework was illustrated and evaluated using the seven long-text political speeches published on the King Abdullah II Official Website https://kingabdullah.jo/ (in a flat text form) [2].

2. LITERATURE REVIEW

We start by providing some relevant background for our study. We first introduce text visualization and then discuss some related work that has applied text summarizing and visualizing techniques, including sentiment analysis, to the analysis of political text.

Text Visualization is an approach to analyze a text in a clear image. Wang et al., Tunggawan et al., Young et al., Chatfield et al. focused on analysis of Tweets about U.S. presidential election campaign [3][4[5] in different manners. Wang et al. focused on the analysis of Tweets (15411 Tweets processed) about the 2012 U.S. presidential election campaign. Figure 1 and 2 represent different visualizations for trending words and for most positive, negative and frequent tweets. Figure 1 displays volume and sentiment by a candidate as well as trending words and a statistical summary. The top-left bar graph shows the number of positive and negative tweets about each candidate in the last five minutes as an indicator of sentiment towards the candidates. The top-right chart displays the number of tweets for each candidate every minute over the previous two hours. The bottom left shows system statistics, including the total number of tweets, the number of seconds since the system start, and the average data rate. The bottom right table shows trending words for the last five minutes.

Figure 2 displays the most positive, negative and frequent tweets, as well as some random neutral tweets. It also shows thetotal volume over time and a tag cloud of the most frequent words in the last five minutes across all candidates.

Tunggawan et al.[4] focused on the analysis of Tweets about 2016 U.S. presidential election campaign. The following figure (Figure 3) shows how the authors used and visualized twitter data.The distribution of tweet labels, most tweets related to Bernie Sanders, Donald Trump, and Hillary Clinton from a specific number of tweets. Figure 4 shows how the model accuracy is affected by several training data, where they trained more models for each candidate using first tweets and used them to predict the next 4000 tweets' sentiment.

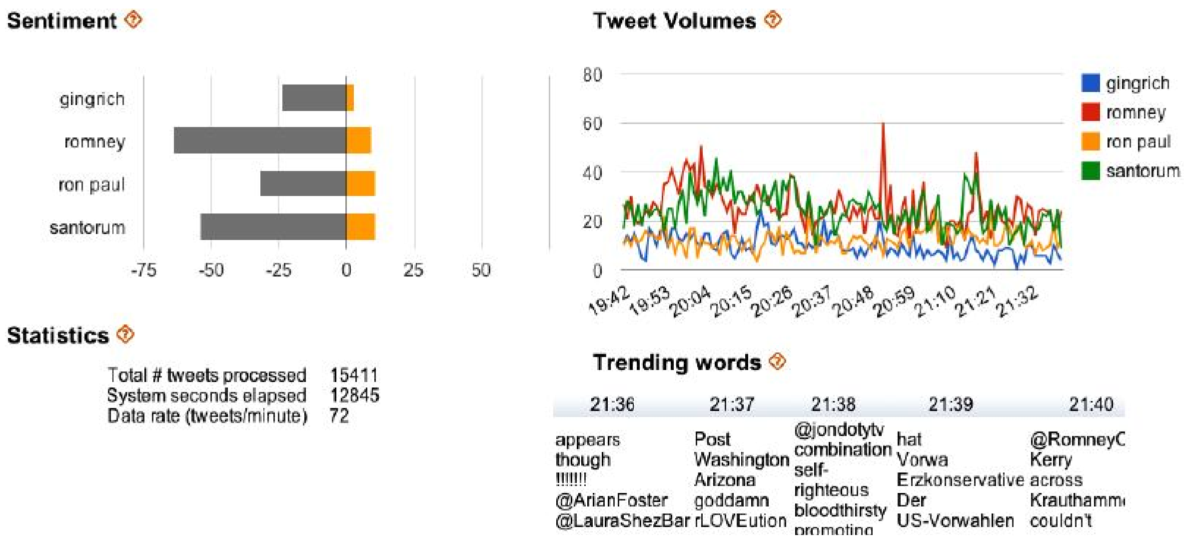


Figure 1: Dashboard for volume, sentiment and trending words

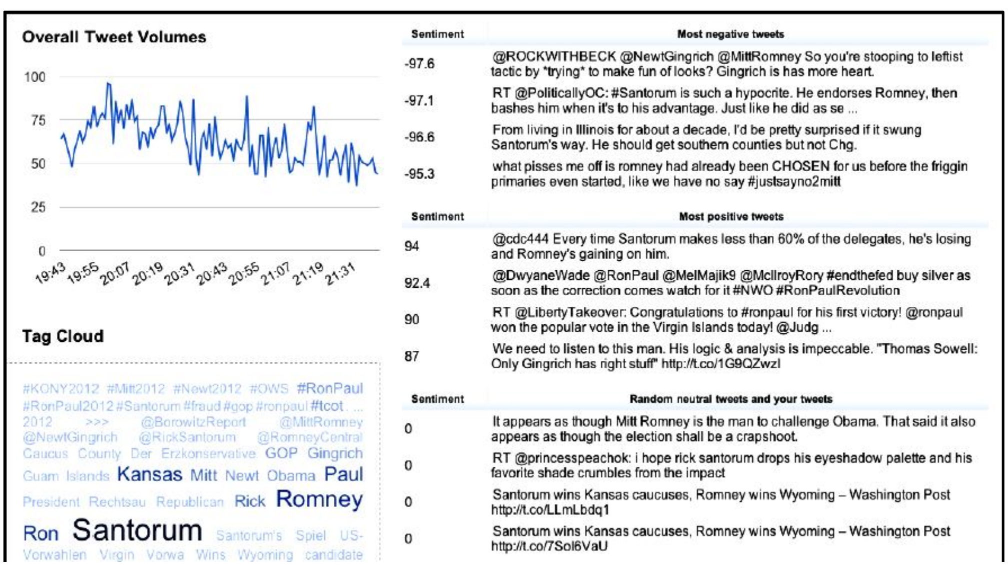


Figure 2: Dashed for most positive, negative and frequent tweets.

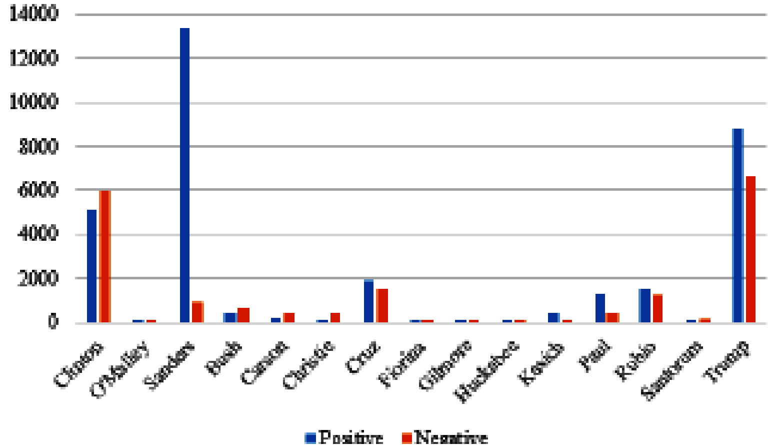


Figure 3: Sentiment Distribution by Candidates

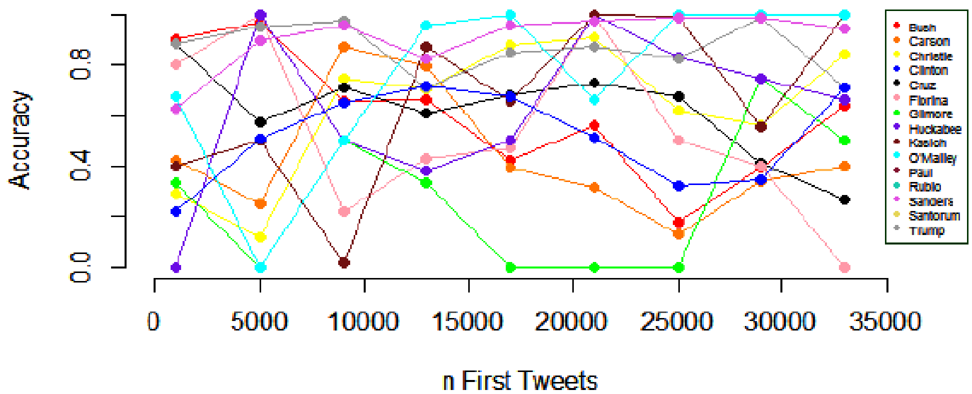


Figure 4: Model Test using n First Tweets as Training and 4000 Next Tweets as Test Data

YoungnyoJoa [5] collected data from different sources like social Twitter, print media, television networks, news magazines, online partisan media, online non-partisan media, and political commentators. Figures 5 and 6 show the visualization image in the context of negative and positive collection data. YoungnyoJoa [5] then used the results to draw two directed graphs of positive and negative word as shown in figure 7 and 8.

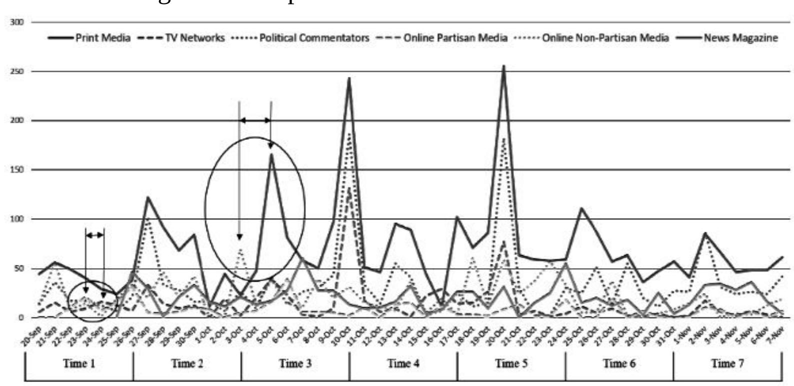


Figure 5: Aggregated Time Series of Negative Sentiment Words by Media Type.

In Figure 7, the edges represent significant relationships found in Granger causal tests for each media Twitter account-media group pair comparison. Granger causality is a statistical concept of causality used to investigate causality between two variables in a time series. The arrows denote the direction of

Granger causality relationship. The size of nodes was adjusted to represent the out-degree centrality calculated in the network.

In figure 8, the edges represent significant relationships found in Granger causal tests for each media Twitter account-media group pair comparison. The arrows denote the direction of Granger causality relationship. The size of nodes was adjusted to reflect the out-degree centrality calculated in the network.

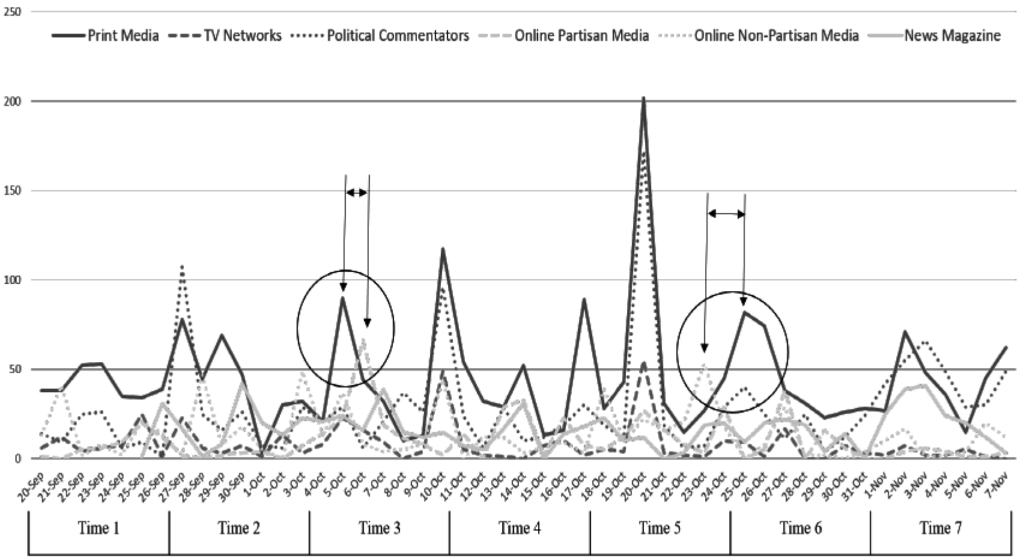


Figure 6: Aggregated Time Series of Positive Sentiment Words by Media Type.

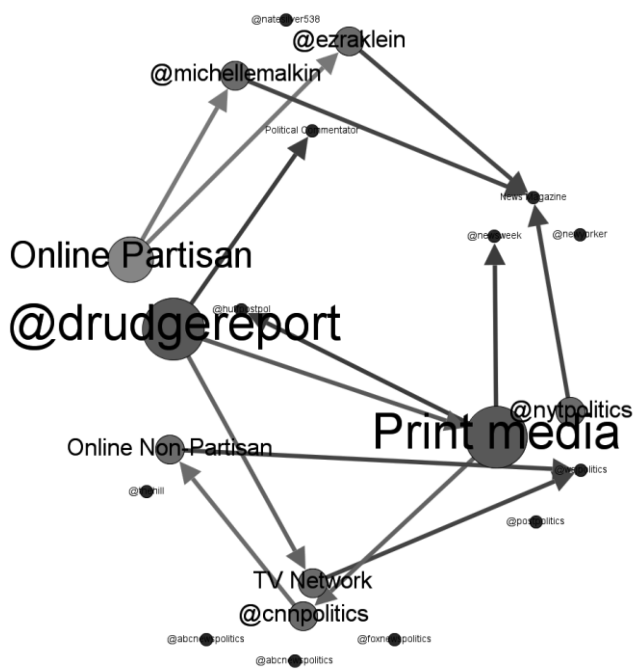


Figure 7: Directed Granger Causality Graph: Negative Sentiment.

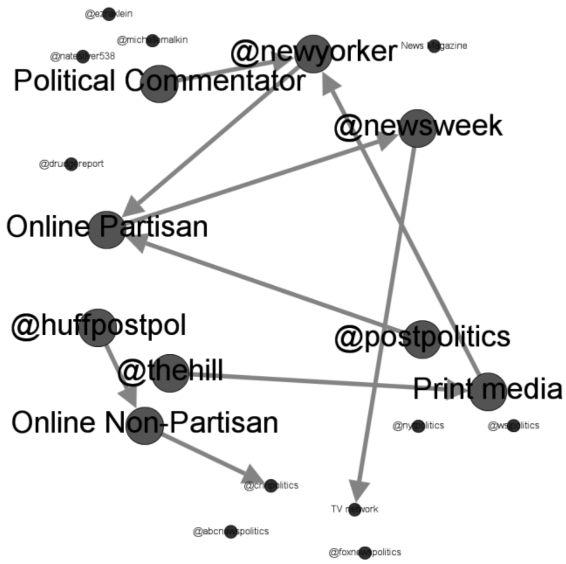


Figure 8: Directed Granger Causality Graph: Positive Sentiment.

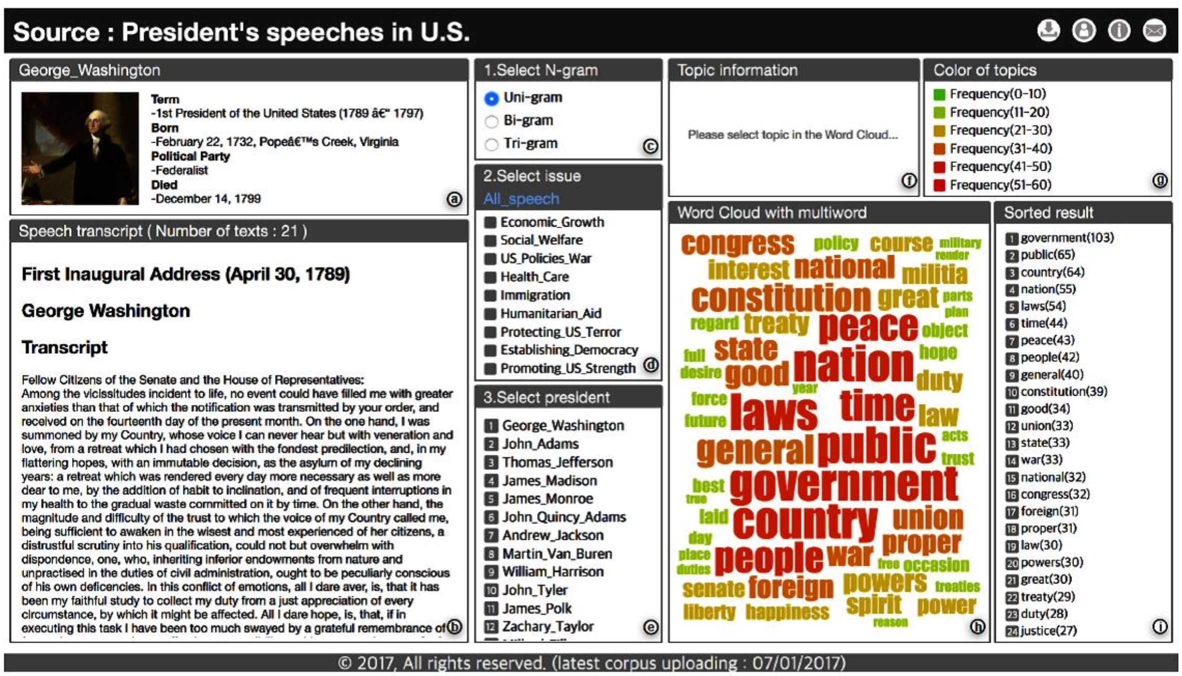


Figure 9: U.S. presidential speeches visualization system interface.

Mun et al.[6] developed an interface of a text visualization system dedicated for a corpus data of 957 political speeches for 43 U.S. presidents from George Washington to Barack Obama. Figure 9 shows that developed text visualization interface for the presidential speech.

Michaela Edwards [7] conducted interesting analysis on Twitter data speeches, where six statements to Trump other American presidents and compare between them with respect to their RNC (Republican National Convention) speeches. Presidential speeches were analyzed through word occurrence of specific ultimate terms revealing similarities and differences in Trump's efficient method and those before him. Figure 10 shows representative Chart Images of Word Occurrence for six different RNC speeches.

González et al. [9] used a natural language processing based approach employing the latent Dirichlet allocation (LDA) algorithm, which is a generative statistical model, to conduct text mining techniques on almost 900 presidential "state-of-the-union"—type speeches from 10 Latin American countries from 1819 to 2016. Recently, the abundant amount of documents reference to poverty and inequality shows that how poverty and inequality are discussed differently. Figure 11 plots the relationship between average poverty and inequality and relative frequency of usage of poverty and inequality in presidential speeches. For each country, they calculate the average poverty rate and Gini coefficient for the period 2000-2015, using all years for which data are available.

Sim et al. [10] measured political candidates' ideological positioning from their speeches. Ideological cues were inferred using a domain-informed Bayesian HMM from a labelled corpus of political texts annotated with predefined ideologies. Figure 12 represents visualizations of the speeches for Obama, Romney, and McCine. The proportion of time spent in each ideology by McCain, Romney, and Obama in the 2008 and 2012 presidential election seasons are highlighted and represented visually.

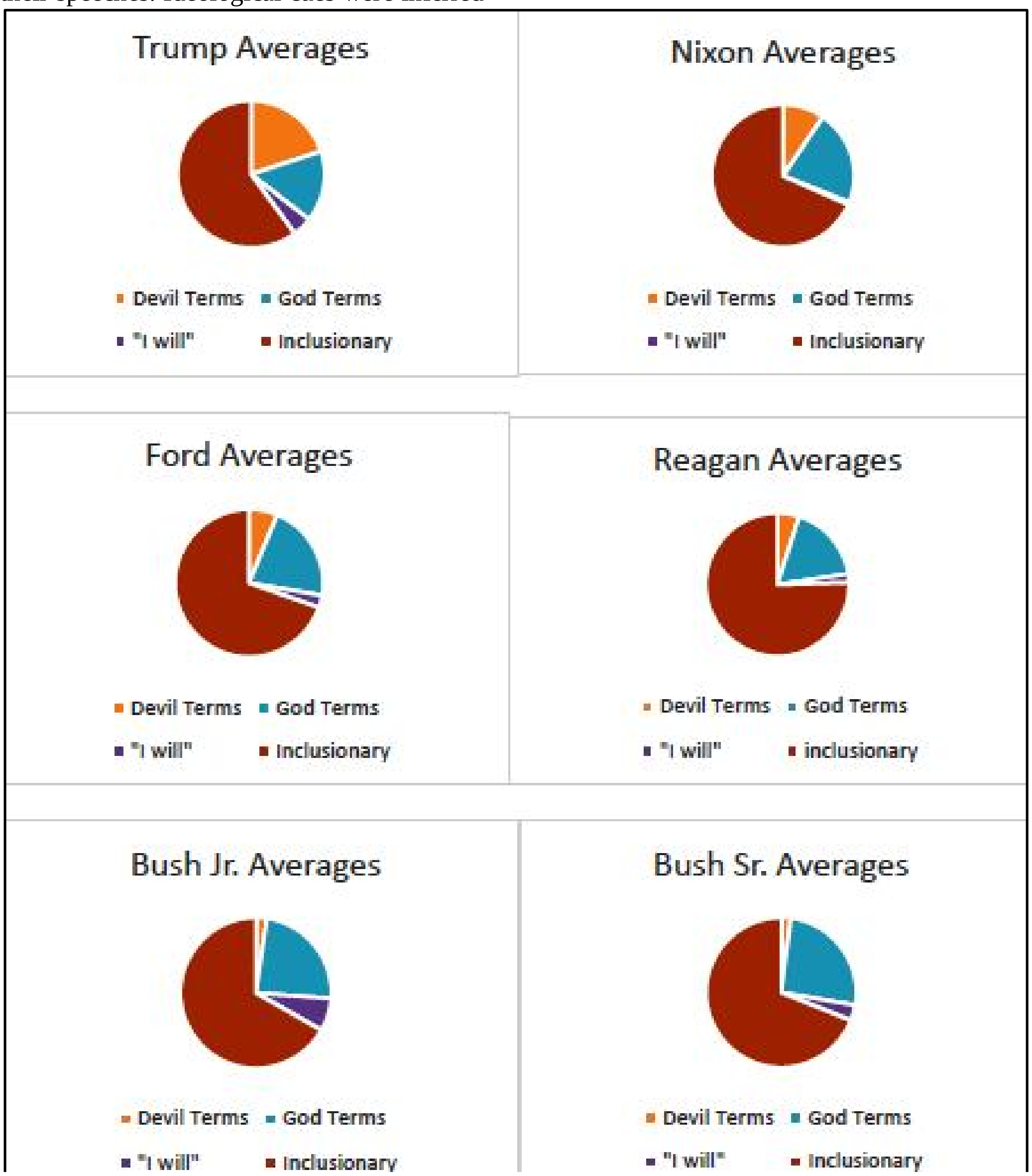


Figure 10: Pie Chart Images of Word Occurrence. Wordcount analysis in six different RNC speeches.

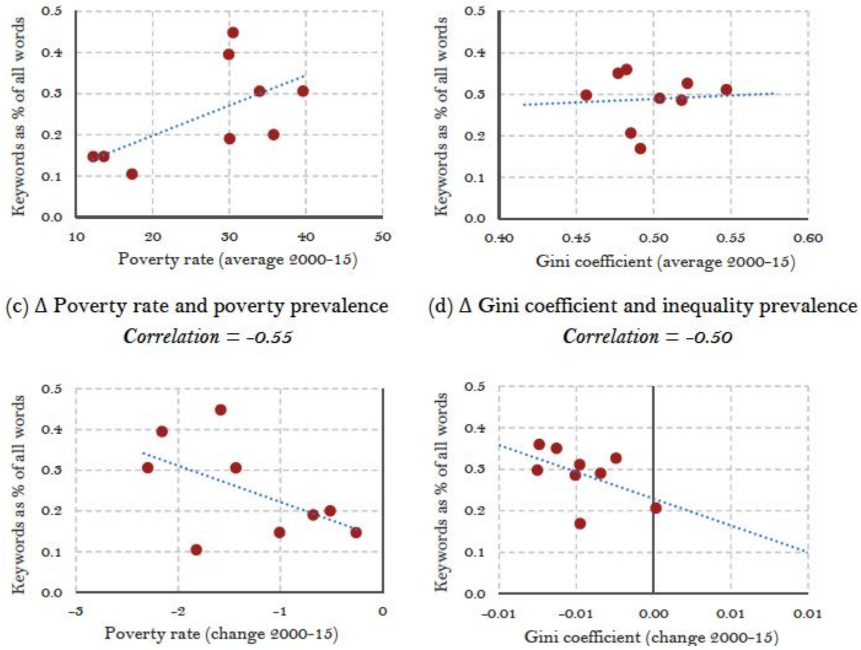


Figure 11: The relationship between average poverty and inequality and relative frequency of the usage of poverty and inequality in presidential speeches.

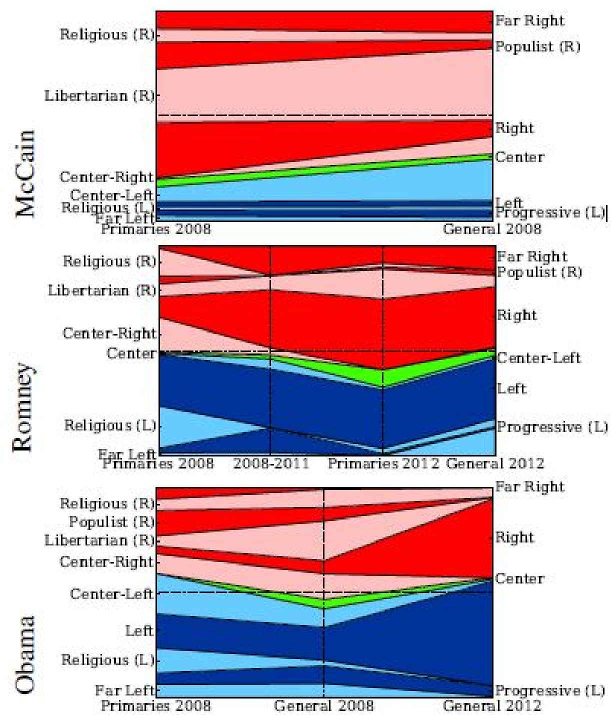


Figure 12: Proportion of time spent in each ideology by McCain, Romney, and Obama during the 2008 and 2012 Presidential election seasons.

Santos et al. [11] conducted a sentiment based visualization approach using tweets about the World Cup in 2014. User's sentiments about Brazil vs. Germany game and hashtags mostly used online by Brazilians. Figures 13 and 14 discuss the visualization approach used.Cui et al. [23] used the word cloud to analyze and visualize documents. The top center of Figure 15 presents a significance trend chart viewer, which shows a curve extracted from a collection of papers with different time stamps. The x-axis represents the time, and the y-axis represents the significance of the word clouds. The green curve in the chart represents the measured significance of the word clouds at different time stamps. Five-word clouds ((a)-(e) in the figure) are created using their algorithm for five selected time points where high significance values are observed.

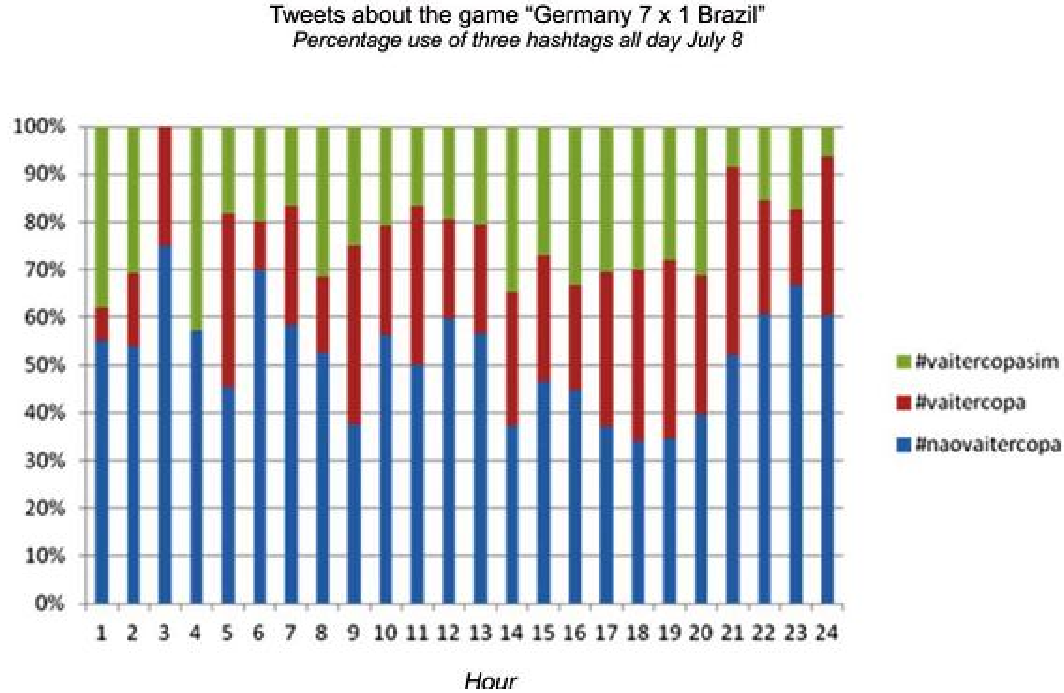


Figure 13: Percentage of tweets with the three hashtags that suggest whether there will or will not be World Cup.

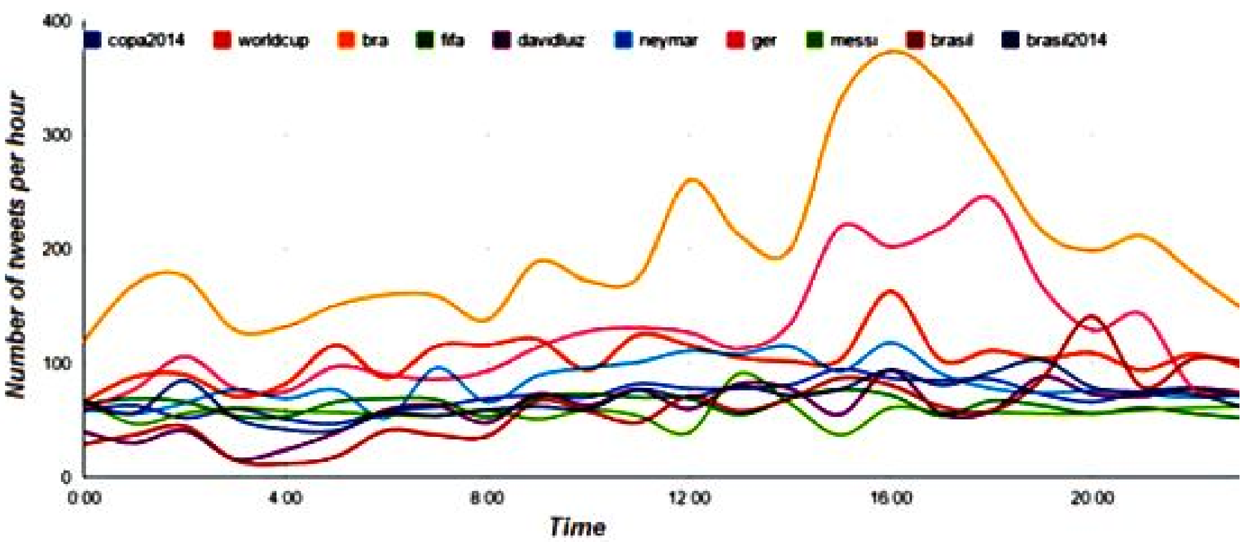
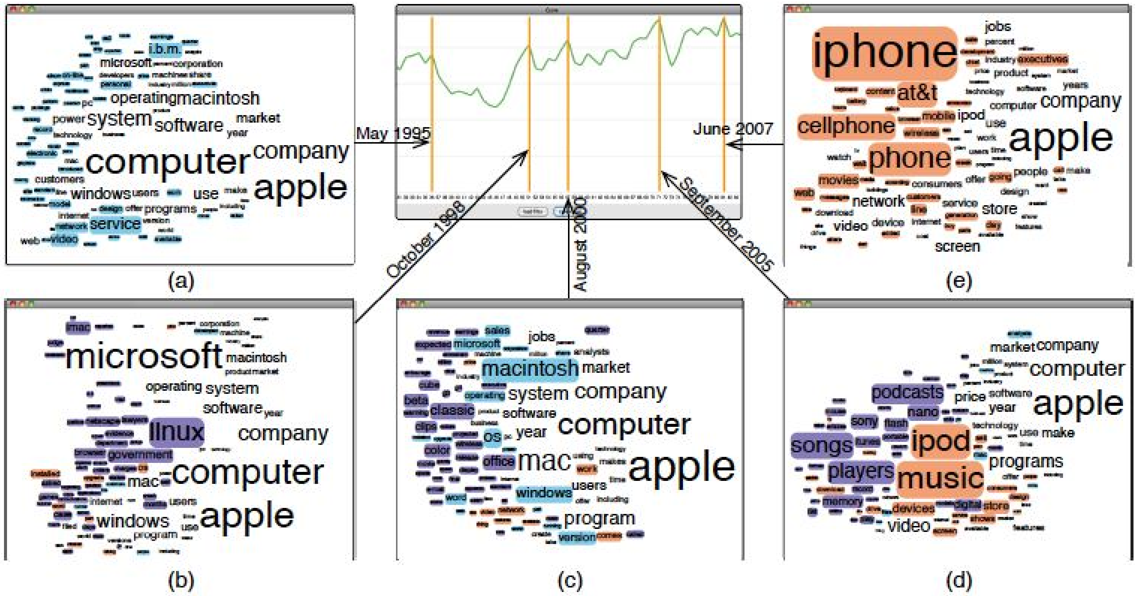


Figure 14: Most frequent hashtags that appeared on July 8th



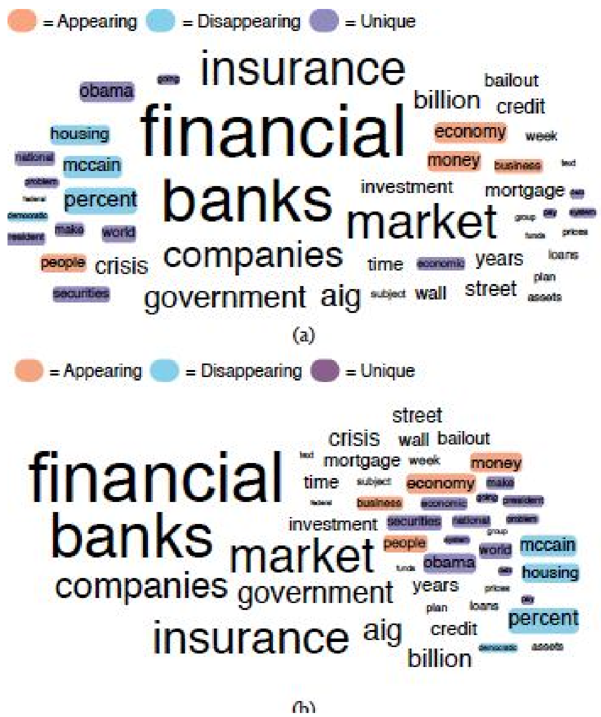


Figure 15: Two word cloud layouts (a) and (b) generated by the importance criterion and the co-occurrence criterion, respectively.

3. DATA SET

To act as a focus for the work described in this paper the seven long-text political speeches published online were used. The authors extracted the speeches associated with the seven discussion papers from the King Abdullah II Official Website https://kingabdullah.jo/. Figure 16 shows the seven discussion papers and their titles. King Abdullah of the Hashemite Kingdom of Jordan, since ascending the Throne in 1999, His Majesty has established a clear vision for comprehensive reform and the future of democracy in Jordan. King Abdullah II has sought to inspire a national dialogue on the reform endeavor and the democratic transformation process that Jordan is undergoing, intending to reach a consensus, encouraging public participation in decision-making, and sustaining the constructive momentum around the on-going reform process.

Discussion Papers

Since ascending the Throne in 1999, His Majesty King Abdullah II ibn Al Hussein has established a clear vision for comprehensive reform and the future of democracy in Jordan. In a series of discussion papers, King Abdullah II has sought to inspire a national dialogue on the reform endeavour and the democratic transformation process that Jordan is undergoing, with the aim of reaching a consensus, encouraging public participation in decision-making and sustaining the constructive momentum around the on-going reform process.

First Discussion Paper:	Our Journey to Forge Our Path Towards Democracy
Second Discussion Paper:	Making Our Democratic System Work for All Jordanians
Third Discussion Paper:	Each Playing Our Part in a New Democracy
Fourth Discussion Paper:	Towards Democratic Empowerment and Active Citizenship
Fifth Discussion Paper:	Goals, Achievements and Conventions: Pillars for Deepening Our Democratic Transition
Sixth Discussion Paper:	Rule of Law and Civil State
Seventh Discussion Paper:	Developing Human Resources and Education Imperative for Jordan's Progress

Figure 16: The Seven Discussion Papers published on the King Abdullah II Official Website https://kingabdullah.jo/.

Our Journey to Forge Our Path Towards Democracy

By Abdullah II ibn Al Hussein
29 December 2012

The Coming Campaign
National lists and candidates across the country have begun their election campaigns for the next Parliament, launching an intense, short election period, in which every day matters, and every citizen matters, because it is your active participation, as citizens, that will breathe life into our democracy.

Candidates are not running for the right to sit in Parliament in Amman and earn personal benefits. They are running to be given a responsibility and a privilege: the national duty of making key choices on some of the most important decisions facing our country, decisions that will impact the future of every Jordanian.

My goal and responsibility within this national course is to encourage debate about our progress as a nation in democratic development. This paper* is part of efforts towards that goal. Today, and in a series of other discussion papers in the next few months, I seek to stimulate debate among citizens about the most important issues we face as a country. A few weeks ago, in an interview with *Al-Rai* and *The Jordan Times* newspapers, I outlined in detail my vision for Jordan's democratic future and the roadmap to get there. Today, I dedicate this paper to share my vision for the principles and values needed to help us progress in our democratisation journey, under our constitutional monarchy.

Now is the time for us to move actively towards key practical milestones in that

Figure 17: Fragment First Discussion Paper: "Our Journey to Forge Our Path Towards Democracy" as published on the King Abdullah II Official Website https://kingabdullah.jo/.

4. TEXT VISUALIZATION FOR KING ABDULLAH SEVEN DISCUSSION PAPERS

In this work, we have used SentiWordNet 3.0 (sentiment lexicon) for purpose of sentiment analysis. Given a text, the goal is to classify subjective words, expressed as bag-of-words (BOW), into two distinct sentiment labels; positive or negative. The core step to predict the sentiment label for a given word is the lookup process, where each word included in each discussion paper is looked up in the sentiment lexicon. If a word is not included in the lexicon it will be considered as a neutral word as described in following Algorithm 1. The Sentiment Label of a word is categorized according to its sentiment score found in the sentiment lexicon. Accordingly, the Sentiment Label Set S is categorized as positive or negative; where: positive indicates a positive polarity and negative indicates negative polarity.

Algorithm 1: Words Sentiment Identification Using Sentiment Lexicon

```
INPUT: SentiWordNet Lexicon
       {T, BOW}
OUTPUT: Set of Sentiment Labels S = {Positive, Negative}
PosCount=  Number of words having Positive Sentiment Polarity
NegCount= Number of words having Negative Sentiment Polarity
PosScore =   The accumulated Positive Sentiment Intensities for each discussion paper
NegScore =  The accumulated Positive Sentiment Intensities for each discussion paper
for all τi∈T  do
        retrieve sentiment intensity for Term τi
                if intensity > 0
                        then τi= Positive
                else if intensity < 0
                        then τi = Negative
                else [i.e intensity = 0]
                        then τi = Neutral
                end if
end for
```

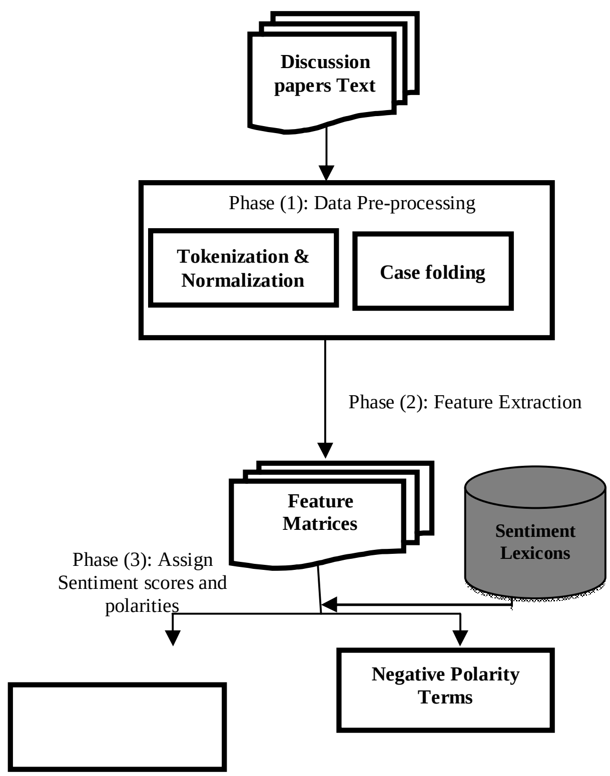


Figure 17: Text Extraction and Polarity Prediction Framework

Table 1: most frequently words in discussion paper

Number of paper						
First	**Second**	**Third**	**Fourth**	**Fifth**	**Sixth**	**Seventh**
10 Most Frequent Words						
our 46	our 32	our 45	Political 24	our 46	our 46	our 23
we 45	parliamentary 22	parliamentary 39	our 32	government 30	government 30	education 16
democracy 15	government 20	government 38	we 19	political 27	political 27	we 15
citizens 12	we 17	political 38	society 13	national 22	national 22	future 9
practices 12	system 12	role 26	democratic 12	parliamentary 18	parliamentary 18	knowledge 6
country 9	citizens 7	parties 24	democratic 10	we 16	we 16	educational 5
election 9	majority 7	national 21	Jordanians 10	continue 14	continue 14	human 5
democratic 8	parliament 7	parliament 20	active 8	system 14	system 14	Jordan 5
Jordanians 8	parties 7	citizens 17	citizens 8	democratic 13	democratic 13	modern 5
future 7	transition 7	Jordanians 13	civic 7	elections 13	elections 13	resources 5
						students 5

Table 2: positive and negative word in discussion paper

Paper Number	Paper title	Date	#of word	Positive Word	Negative Word
First	Our Journey to Forge Our Path Towards Democracy	29/12/2012	1949	167	64
Second	Making Our Democratic System Work for All Jordanians	16/1/2013	1299	113	44
Third	Each Playing Our Part in a New Democracy	2/3/2013	3844	251	112
Forth	Towards Democratic Empowerment and Active Citizenship	2/6/2013	1784	158	58
Fifth	Goals, Achievements and Conventions: Pillars for Deepening Our Democratic Transition	3/10/2014	2928	212	86
Sixth	Rule of Law and Civil State	16/10/2016	2966	215	84
Seventh	Developing Human Resources and Education Imperative for Jordan's Progress	15/4/2017	1062	120	41

5. CONCLUSION AND FUTURE WORK

In this paper we have described a framework for generating long-text speech graph from transcripts of political speeches. The objective of the research described was to deploy sentiment analysis techniques for the extraction of speech graphs that will in turn allow for the graphical visualisation of the high level structure of such speeches showing the evolution and the distribution of sentiment intensity and polarity of the subjective text within the speeches.

The operation of the framework was illustrated and evaluated using seven long-text political speeches published on the King Abdullah II Official Website https://kingabdullah.jo/ (in a flat text form). The promising results obtained so far indicate that: (i) it is possible to capture the speech structure representing the attitude of each speech; (ii) it is possible to use lexicon based opinion mining techniques (such as SentiWordNet) to identify the attitudes embedded within long-text political speeches or discussions, although dedicated political lexicons might need to be used to improve overall accuracy. Future extensions will be directed at the adoption of machine learning techniques instead of lexicon based technique that used in this research work, in addition to use deeper linguistic analysis techniques for extracting latent meanings and information embedded in the speeches.

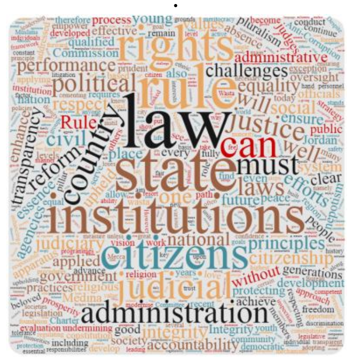
(1) First Discussion paper Word Cloud.

(2) Second Discussion paper Word Cloud
.

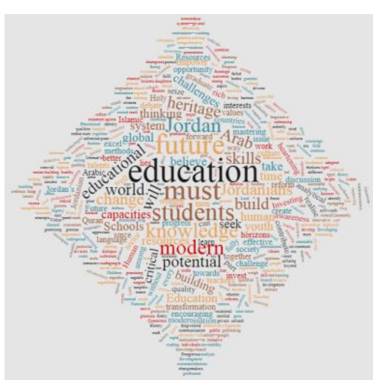
(3) Third Discussion paper Word Cloud.

(4) Fourth Discussion paper Word Cloud
.

(5) Fifth Discussion paper Word Cloud.

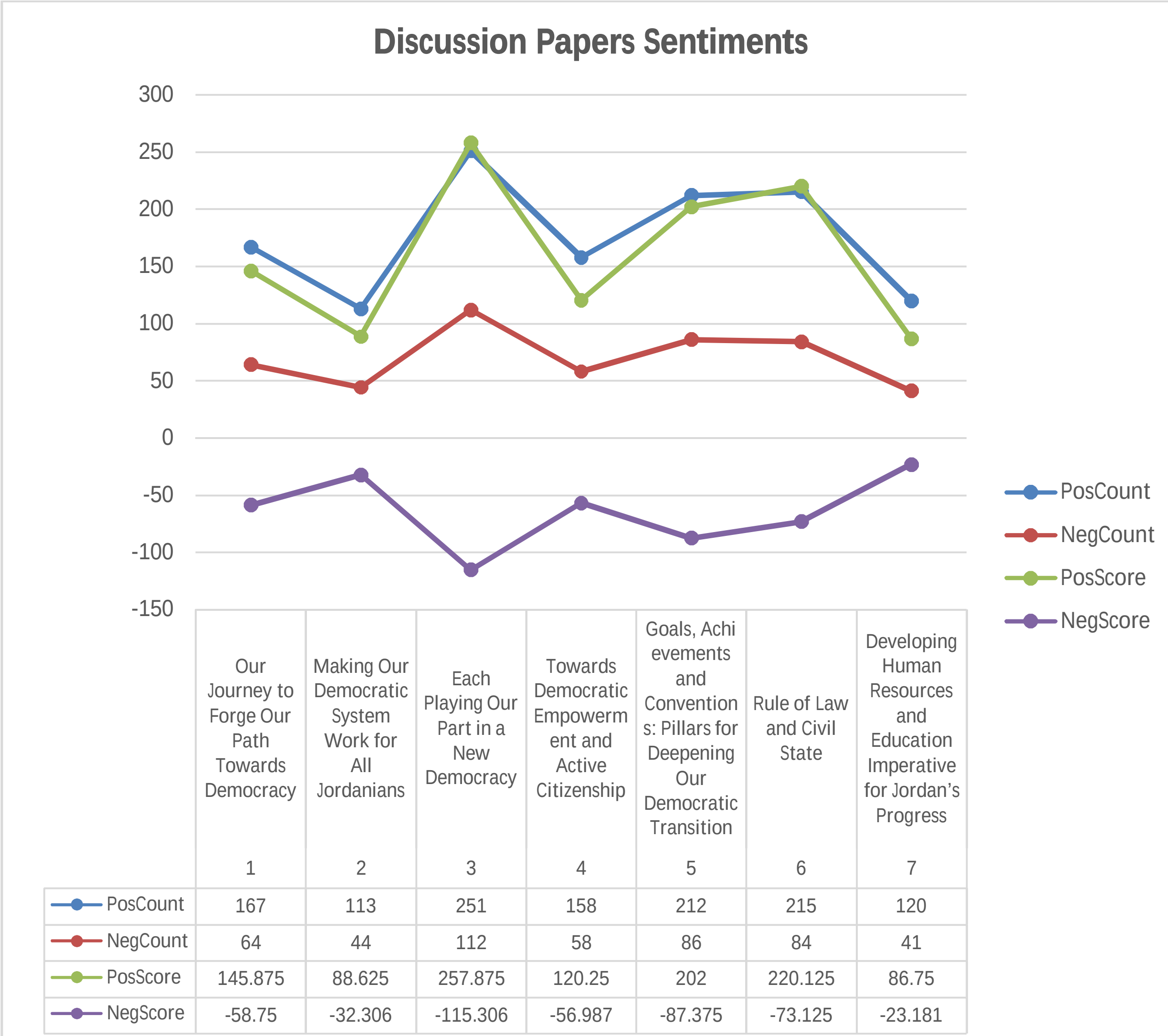


(6) Sixth Discussion paper Word Cloud.



(7) Seventh Discussion paper Word Cloud.

Figure 18: Text Visualization Using Word Cloud (https://www.wordclouds.com/) for the seven Discussion papers.



	Our Journey to Forge Our Path Towards Democracy 1	Making Our Democratic System Work for All Jordanians 2	Each Playing Our Part in a New Democracy 3	Towards Democratic Empowerment and Active Citizenship 4	Goals, Achievements and Conventions: Pillars for Deepening Our Democratic Transition 5	Rule of Law and Civil State 6	Developing Human Resources and Education Imperative for Jordan's Progress 7
PosCount	167	113	251	158	212	215	120
NegCount	64	44	112	58	86	84	41
PosScore	145.875	88.625	257.875	120.25	202	220.125	86.75
NegScore	-58.75	-32.306	-115.306	-56.987	-87.375	-73.125	-23.181

Figure 19: The resulted summary visualization showing the differences between the attitudes of consecutive speeches and to highlight the evolution of the sentiment polarity among the speeches using statistical linguistic analysis for political jargon

REFERENCES

1. Vinodhini, G., and R. M. Chandrasekaran.**Sentiment analysis and opinion mining: a survey,** *Int J Adv Res ComputSciTechnol*,2.6 (2012): 282-292.
2. https://kingabdullah.jo/en/vision/discussion-papers
3. Wang, Hao, et al. **A system for real-time twitter sentiment analysis of 2012 us presidential election cycle**, *Proceedings of the ACL 2012 System Demonstrations. Association for Computational Linguistics*, 2012.
4. Tunggawan, Elvyna, and Yustinus Eko Soelistio.**And the winner is…: Bayesian Twitter-based prediction on 2016 US presidential election.**, *Computer, Control, Informatics and its Applications (IC3INA), 2016 International Conference on. IEEE*, 2016. https://doi.org/10.1109/IC3INA.2016.7863019
5. Joa, Youngnyo. **A Hyperlink and Sentiment Analysis of the 2016 Presidential Election: Intermedia Issue Agenda and Attribute Agenda Setting in Online Contexts.***Diss. Bowling Green State University*, 2017.
6. Mun, Seongmin, Guillaume Desagulier, and Kyungwon Lee. **A visual approach for text analysis using multiword topics,** *EuroVis 2017. Anna Puig Puig; Tobias Isenberg*, 2017.
7. Edwards, Michaela. Gotham.**A Rhetorical Analysis of President Donald J. Trump's Republican National Committee Acceptance Speech.***Diss. University Honors College, Middle Tennessee State University*, 2017.

8. Guha, Sreya. **Doris: A tool for interactive exploration of historic corpora (Extended Version).** arXiv preprint arXiv:1711.00714(2017).
9. Calvo-González, Oscar, Axel Eizmendi, and German Jeremias Reyes. **The changing way governments talk about poverty and inequality: evidence from two centuries of Latin American presidential speeches.** (2018). https://doi.org/10.1596/1813-9450-8311
10. Sim, Yanchuan, et al. **Measuring ideological proportions in political speeches,** *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing.* 2013.
11. Santos, Caroline Q., et al. **Can Visualization Techniques Help Journalists to Deepen Analysis of Twitter Data? Exploring the, Germany 7 x 1 Brazil Case.",** *System Sciences (HICSS), 2016 49th Hawaii International Conference on. IEEE*, 2016. https://doi.org/10.1109/HICSS.2016.245
12. Alieva, Iuliia. **Data journalism and democracy: how American mass media framed the 2016 presidential campaign in the United States using data visualization.***Diss. University of MissouriColumbia*, 2017.
13. Yeuk-Yin, C. H. A. N., and Q. U. Huamin. **FinaVistory: Using Narrative Visualization to Explain Social and Economic Relationships in Financial News.**
14. Queiroz Santos, Caroline, et al. **Media professionals' opinions about interactive visualizations of political polarization during Brazilian presidential campaigns on Twitter.** (2017). https://doi.org/10.24251/HICSS.2017.229
15. Salloum, Said A., Mostafa Al-Emran, and Khaled Shaalan. **Mining social media text: extracting knowledge from facebook,** *International Journal of Computing and Digital Systems*, Vol. 6. No. 2. University of Bahrain, 2017. https://doi.org/10.12785/IJCDS/060203
16. Elghazaly, Tarek, Amal Mahmoud, and Hesham A. Hefny. **Political sentiment analysis using twitter data.** *Proceedings of the International Conference on Internet of things and Cloud Computing*. ACM, 2016. https://doi.org/10.1145/2896387.2896396
17. SVAMPA, M. Lucila. **Public uses of dictatorial pasts: Visualizations in Germany.** *História (São Paulo)* 35 (2016).
18. Handler, Abram, and Brendan O'Connor. **Rookie: A unique approach for exploring news archives.** arXiv preprint arXiv:1708.01944 (2017).
19. Mostafa, Mohamed M., and Nicolás Roser Nebot. **Sentiment Analysis of Spanish Words of Arabic Origin Related to Islam: A Social Network Analysis.** *Journal of Language Teaching and Research*, 8.6 (2017): 1041-1049.
20. Öztürk, Nazan, and Serkan Ayvaz. **Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis.** *Telematics and Informatics* (2017). https://doi.org/10.1016/j.tele.2017.10.006
21. Claster, William B., Malcolm Cooper, and Philip Sallis. **Thailand--Tourism and conflict: Modeling sentiment from Twitter tweets using naïve Bayes and unsupervised artificial neural nets.** *Computational Intelligence, Modelling and Simulation (CIMSiM), 2010 Second International Conference on. IEEE*, 2010.
22. Sörensen, Jil. **The German federal president in the press: Public officeholder or private citizen?.** *European Journal of Communication* 31.3 (2016): 243-259.
23. Cui, Weiwei, et al. **Context preserving dynamic word cloud visualization.** *Visualization Symposium (PacificVis), 2010 IEEE Pacific. IEEE*, 2010. https://doi.org/10.1109/PACIFICVIS.2010.5429600
24. Arun Krishna Chitturi and Saravanakumar Kandasamy. **Survey on Abstractive Text Summarization using various approaches,** *International Journal of Advanced Trends in Computer Science and Engineering,* Volume 8, No.6, November – December 2019, pp. 2956-2964. https://doi.org/10.30534/ijatcse/2019/45862019
25. MagantiSyamala and N.J.Nalini. **A Deep Analysis on Aspect based Sentiment Text Classification Approaches,** *International Journal of Advanced Trends in Computer Science and Engineering,* Volume 8, No.5, September - October 2019, pp. 1795-1801. https://doi.org/10.30534/ijatcse/2019/01852019