AN EXPANDABLE ARABIC LEXICON AND VALENCE SHIFTER RULES FOR SENTIMENT ANALYSIS ON TWITTER

BAHA` NAJIM SALMAN IHNAINI

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2019

Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

BAHA' NAJIM SALMAN IHNAINI

calon untuk Ijazah PhD
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"AN EXPANDABLE ARABIC LEXICON AND VALENCE SHIFTER RULES FOR SENTIMENT ANALYSIS ON TWITTER"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : ***24 Januari 2019.***
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on: ***January 24, 2019.***

Pengerusi Viva: *(Chairman for VIVA)*	**Prof. Dr. Huda Hj Ibrahim**	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	**Assoc. Prof. Dr. Nazlia Omar**	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	**Dr. Azizi Ab Aziz**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Dr. Massudi Mahmuddin**	Tandatangan *(Signature)*

Tarikh:
(Date) ***January 24, 2019***

Permission to Use

In presenting this thesis in fulfillment of the requirements for a postgraduate degree from University Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this project in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor or in his absence by the Assistant of Vice Chancellor of College of Arts and Sciences. It is understood that any copying or publication or use of this project or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to University Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences

UUM College of Arts and Sciences

Universiti Utara Malaysia

06010 UUM Sintok

Abstrak

Analisis sentimen (SA) merujuk kepada pengkomputeran dan teknik pemprosesan bahasa tabii yang digunakan untuk mengekstrak maklumat subjektif dalam sebaris teks. Dalam kajian SA ini, tiga pemasalahan utama dikenalpasti: a) ketiadaan sumber pada dialek bahasa Arab Palestin (PAL), b) kewujudan perkataan sentimen baru sehingga mengurangkan prestasi model analisis sentimen apabila diterapkan pada twit yang dikumpulkan, dan c) mengendalikan perkataan pengubah valens yang tidak ditangani dengan teliti dalam analisis sentimen bahasa Arab. Oleh itu, kajian ini bertujuan untuk membangunkan leksikon PAL untuk twit Palestin dan membina leksikon yang boleh diperbaharui dan terkini untuk bahasa Arab (EULA). Satu peratuan pengubah valens yang baru bagi meningkatkan prestasi analisis sentimen berasaskan leksikon terhadap twit bahasa Arab turut dibina. Dalam kajian ini, leksikon PAL telah dibina dengan menggunakan algoritma pemadanan fonologi manakala EULA dibina dengan memanfaatkan leksikon umum pada set data twit untuk mencari istilah baru dan meramalkan polariti melalui beberapa peraturan linguistik. Tambahan pula, satu set peraturan telah dicadangkan untuk mengendalikan perkataan pengubah valens. Dengan menggunakan peraturan untuk mencari skop perkataan, dan nilai peralihan yang dihasilkan oleh perkataan tersebut. Set data twit Palestin dan Arab dari bulan Mac hingga Mei 2018 telah digunakan bagi menilai idea yang dicadangkan. Hasil eksperimen menunjukkan bahawa leksikon PAL yang dicadangkan telah menghasilkan keputusan yang lebih baik berbanding dengan leksikon lain apabila diuji pada set data Palestin. Sementara itu, EULA dapat meningkatkan prestasi pendekatan berasaskan leksikon untuk bersaing dengan pendekatan pembelajaran mesin. Malahan lagi, penggunaan peraturan pengubah valens yang dicadangkan telah meningkatkan prestasi purata keseluruhan sebanyak 5%. Leksikon sentimen PAL baru yang dicadangkan dapat mengendalikan dialek Palestin. Tambahan pula, EULA telah mengatasi kelemahan kewujudan perkataan slang baru dalam media sosial. Selain itu, peraturan pengubah valens yang dibina mampu mengatasi penafian, intensifikasi, dan kontras dalam meningkatkan prestasi analisis sentimen bahasa Arab.

Kata Kunci: Walaupun moden Arabic, Pendekatan berasaskan leksikon, Aturan peraturan shifter.

Abstract

Sentiment analysis (SA) refers as computational and natural language processing techniques used to extract subjective information expressed in a text. In this SA study, three main problems are addressed: a) absence of resources on Palestinian Arabic dialect (PAL), b) emergence of new sentiment words, hence decreases the performance of sentiment analysis models when applied on tweets collected, and c) handling valence shifter words were not thoroughly addressed in Arabic sentiment analysis. Therefore, this study aims to construct a PAL lexicon for Palestinian tweets and to design an Expandable and Up-to-date Lexicon for Arabic (EULA). A new valence shifter rules in enhancing the performance of lexicon-based sentiment analysis on Arabic tweets is also been constructed. In this study, a PAL lexicon is built by using phonology matching algorithm while EULA is constructed by harnessing a general lexicon on a tweets dataset to find new terms and predict its polarity through some linguistic rules. Furthermore, a set of rules are proposed to handle the valence shifters words by applying rules to find the scope of words, and shifting value that is produced by these words. Palestinian and Arabic tweets datasets from March to May 2018 are used to evaluate the proposed idea. Experimental results indicate that the proposed PAL lexicon has produced better results compared to other lexicons when tested on Palestinian dataset. Meanwhile, EULA enhanced the performance of lexicon-based approach to be competitive with machine learning approach. Moreover, applying the proposed valence shifter rules have increased overall performance of 5% on average. The new proposed PAL sentiment lexicon is able to handle Palestinian's dialects. Furthermore, the EULA has overcome the emergence of new slang words in social media. Moreover, the constructed valence shifter rules are capable to handle negation, intensifiers and contrasts in enhancing the performance of Arabic sentiment analysis.

Keywords: Arabic sentiment analysis, Palestinian dialect lexicon, Lexicon-based approach, Valence shifter rules, Twitter.

Acknowledgment

First and for most, thank you Almighty Allah for giving me the health, courage, patience, and all the power to continue this journey through all hard times.

My eternal partner, cheerleader, forever interested, encouraging and always enthusiastic, my wife Suha, I owe it all to you. I will always remember your screams of joy whenever a significant milestone was reached. Many Thanks!

I am grateful to my mother Fathiyah Maarouf, who has provided me with moral and emotional support, tears, and prayers for me through all nights. I am also thankful to my father Najim Ihnaini for his continuous push in order to reach this point. Thanks also goes to my other family members, brothers and sisters who have supported me along the way.

With a special gratitude to my supervisor Dr. Massudi Mahmuddin for all the guidance and support. And finally, last but by no means least, big appreciation also to everyone in the InterNetworks Laboratory, chaired by Prof. Dr. Suhaidi Hassan, it was great sharing laboratory with all of you during my Ph.D. journey.

Thanks for all your encouragement.

Table of Contents

Permission to Use ..... **ii**

Abstrak ..... **iii**

Abstract ..... **iv**

Acknowledgment ..... **v**

List of Tables ..... **xi**

List of Figures ..... **xiii**

List of Abbreviations ..... **xv**

CHAPTER ONE INTRODUCTION **1**

1.1 Background 1

1.2 Research Motivation 3

1.3 Problem Statement 4

1.4 Research Questions 6

1.5 Research Objectives 7

1.6 Scope of the Research 7

1.7 Research Contributions 8

1.8 Thesis Organization 9

CHAPTER TWO LITERATURE REVIEW **12**

2.1 Sentiment Analysis of Arabic 12

2.2 Arabic Language 14

2.2.1 Palestinian Dialect 16

2.2.2 Arabic Tweets 17
2.3 Tweets Collection 19
2.4 Pre-processing 21
2.4.1 Tweets Cleaning 21
2.4.2 Tokenization 22
2.4.3 Normalization 23
2.4.4 Stemming 23
2.4.5 Stop Words Removal 24
2.5 Sentiment Analysis Approaches 27
2.5.1 The Machine Learning Approach 27
2.5.2 The Lexicon-Based Approach 30
2.5.3 The Hybrid Approach 40
2.6 Valence Shifters 42
2.6.1 Negation Words 42
2.6.2 Intensification Words 45
2.6.3 Contrast Words 46
2.7 Latest Researches on Arabic Lexicon-Based Sentiment Analysis 46
2.8 Research Gap 66
2.9 Summary 68
CHAPTER THREE RESEARCH METHODOLOGY 69
3.1 Introduction 69

3.2 Research Phases 69
3.3 Theoretical Study 70
3.4 Experimental Design 71
3.4.1 Crawling Tweets 71
3.4.2 Arabic Tweets Datasets 75
3.4.5 Pre-processing and Cleaning 78
3.4.3 Lexicons Construction 83
3.4.4 Arabic Sentiment Lexicons 85
3.4.6 Features Extraction 88
3.4.7 Rules Implementation 88
3.5 Evaluation Measurement 94
3.6 Summary 98
CHAPTER FOUR AN ENHANCED LEXICON CONSTRUCTION AND VALENCE SHIFTER RULES 99
4.1 Introduction 99
4.2 Tweets Pre-processing 102
4.3 Lexicons Construction 102
4.3.1 Construction of Basic Lexicon 103
4.3.2 Construction of PAL Lexicon 107
4.3.3 EULA Construction 111
4.3.4 Valence Shifter Lexicons Construction 121
4.4 Valence Shifter Rules 121

4.4.1 Contrast Rules 122

4.4.2 Negation Rules 124

4.4.3 Intensifier Rules 128

4.4.4 Predictor Words Rules 130

4.5 Benchmarking with Latest Related Researches 130

4.6 Summary 133

CHAPTER FIVE RESULTS AND DISCUSSION 134

5.1 Overview 134

5.2 Experimental Results of PAL Lexicon 136

5.3 EULA Experimental Results 142

5.3.1 Performance and Evaluation of EULA-L 143

5.3.2 Performance and Evaluation of EULA-U 148

5.4 Experimental Results of Valence Shifter Rules 151

5.5 Summary 159

CHAPTER SIX CONCLUSION AND FUTURE WORK 161

6.1 Summary of Research 161

6.2 Achievements 163

6.2.1 New PAL Lexicon 163

6.2.2 Expandable and Updated EULA 164

6.2.3 Enhanced Valence Shifter Rules 164

6.3 Research Limitations 165

6.4 Future Work 166

6.4.1 Lexicons for other Arabic Dialects 166

6.4.2 Stop Words List from EULA 166

6.4.3 Multi-Classification Approach 167

6.4.4 Handling Sarcasm 167

6.4.5 Building Larger Dataset 167

6.5 Summary 168

References 170

List of Appendices 195

Appendix A Tweepy Code for Collecting Tweets 195

Appendix B Code of Expanding EULA 196

Appendix C Implementation of Contrast Rules 198

Appendix D Implementation of Intensifier Rules 202

Appendix E Implementation of Negation Rules 204

Appendix F Snapshot of Data 208

Appendix G Valence Shifter Lists 212

Appendix H Experts Biography 216

Appendix I Links to Datasets 217

List of Tables

Table 2.1 Summary of Pre-processing Tools in the Literature 26

Table 3.1 Agreement Table between Linguists 73

Table 3.2 Manual Validation of the Automatic Annotation 75

Table 3.3 Datasets Used for Evaluation Purposes 77

Table 3.4 Example on Pre-processing a Tweet 82

Table 3.5 Lexicons Used for Benchmarking Purposes 87

Table 3.6 Example of Negation`s Scope When Polarity Changes 90

Table 3.7 Example of Negation`s Scope When Polarity Doesn`t Change 91

Table 4.1 Overall Process of Lexicons Construction 101

Table 4.2 Experiment`s Results of Combining Lexicons to Form the Basic Lexicon 105

Table 4.3 Terms from Proposed PAL Lexicon with Translation to English 111

Table 4.4 Example of Unlabeled Tweets 120

Table 4.5 Process of Expanding EULA-U through Unlabeled Tweets 120

Table 4.6 Predicted Polarity Before and After Expanding EULA-U 120

Table 4.7 Finding Window Size of the Negation`s Scope 125

Table 4.8 Examples on Polarity Shifting of Negative Prefixes Words 126

Table 4.9 Examples of Polarity Shifting by Negation Word. 127

Table 4.10 Summary of benchmark Methods on Arabic Language 132

Table 5.1 Evaluation Results when Simple Lexicon-Based Approach is applied on Levantine Datasets 138

Table 5.2 Evaluation Results when Simple Lexicon-Based Approach is applied on Palestinian Dataset 141

Table 5.3 Datasets Split: Training and Testing 144

Table 5.4 Best F-score as Reported in Benchmark Researches 145

Table 5.5 Performance Measurements of using EULA-L when Expanded by the Same Dataset 146

Table 5.6 Performance Measurements of using EULA-L Expanded by EMAR-Tweets Dataset 147

Table 5.7 Terms from EULA-L with Translation to English 148

Table 5.8 Reported F-scores from the Literature of Lexicon-Based Approach on Arabic Tweets Datasets 150

Table 5.9 Performance Measurements of using EULA-U Expanded by EMAR-Tweets Dataset 151

Table 5.10 Results obtained without using Negation Rules, with using Switch Negation, and with using Researcher`s Negation 154

Table 5.11 Results of not Applying Rules, and Results of Applying Contrast Rules 155

Table 5.12 Results without Applying Rules, and Results with Intensification Rules 157

Table 5.13 Results without Applying Rules, and Results with Applying all Valence Shifter Rules 158

List of Figures

Figure 2.1. Sentiment Classification Techniques.....27

Figure 2.2. Research Problems and Solutions.....67

Figure 3.1. Research Phases 70

Figure 3.2. The Experimental Design 71

Figure 3.3. Tweets Pre-processing Stages 79

Figure 4.1. Proposed Lexicon-Based Sentiment Analysis System 99

Figure 4.2. Pre-processing Steps Sequence.....102

Figure 4.3. Hierarchy of All Proposed Lexicons 103

Figure 4.4. Steps of Constructing the Basic Lexicon.....105

Figure 5.1. Testing all lexicons by Simple Lexicon-Based Approach using Levantine Datasets 136

Figure 5.2. Accuracy Rates of all Lexicons using Simple Lexicon-Based Approach when applied on Levantine Datasets.....139

Figure 5.3. Testing all Lexicons By Simple Lexicon-Based Approach using PAL-Tweets Dataset 140

Figure 5.4 Accuracy Rate of all Lexicons using Simple Lexicon-Based Approach when applied on Palestinian Dataset.....142

Figure 5.5. Testing EULA-L by Simple Lexicon-Based Approach.....143

Figure 5.6. 5-Fold Cross-Validation 144

Figure 5.7. Accuracy Rates without using Negation Rules, using Switch Negation, and using Researcher`s Negation.....153

Figure 5.8. Accuracy Rates of no Rules against Applying Contrast Rules.....156

Figure 5.9. Accuracy Rates of no Rules against Applying Intensification Rules....156

Figure 5.10. Accuracy Rates of no Rules against Applying All Valence Shifter Rules 159

List of Abbreviations

AEL	Arabic Emoticons Lexicon
AHL	Arabic Hashtags Lexicon
AFINN	Affective Lexicon by Finn Arup Nielsen
AMT	Amazon Mechanical Turk
ANEW	Affective Norms for English Words
API	Application Programming Interface
BEP	Break Even Point
DA	Dialect Arabic
DAHL	Dialectical Arabic Hashtags Lexicons
EULA	Expandable and Updated Lexicon for Arabic
EWN	English WordNet
FN	False Negative
FP	False Positive
KNN	K-Nearest Neighbor
MaxEnt	Maximum Entropy
ML	Machine Learning
MPQA	Multi-Perspective Question Answering
MSA	Modern Standard Arabic
NB	Naïve Bayes
NLP	Natural Language Processing
PAL	Palestinian Arabic Dialect
PANAS	Positive Affect Negative Affect Schedule
PMI	Point-wise Mutual Information
POS	Part of Speech
RT	Re-Tweet
SA	Sentiment Analysis
SAMAR	Subjectivity and Sentiment Analysis of Arabic Social Media
SLSA	Standard Arabic Sentiment Lexicon
SVM	Support Vector Machines
TF	Term Frequency
TN	True Negative
TP	True Positive

URL	Uniform Resource Locator
VADER	Valence Aware Dictionary for Sentiment Reasoning
UWOM	Un-Weighted Opinion Mining

CHAPTER ONE

INTRODUCTION

1.1 Background

People all over the world are getting used to express feelings and present their own opinions using different social media platforms with more than five hundred millions of tweets per day by millions of people on Twitter only. This has been a good destination for organizations to investigate objectives, to study people`s reactions and opinions on several things in life. This has attracted researchers to benefit more from the data produced from social media for analyzing aims, using techniques as language processing, sentiment analysis, text mining, text processing, and information extraction on Twitter and any other microblogging services.

In this thesis, sentiment analysis has been under investigation. In order to study sentiment analysis, the word "sentiment" should be defined as terms like opinion, emotion, sentiment, evaluation and belief, also, expressions that are not related to objective observations or verification. Yet, the diversity in these terms could make beginners in this area misunderstand the nature of this term or become uncertain about it. Mainly, the informational sentence has an objective meaning and one with personal opinion and feelings is called a subjective sentence. Therefore, sentiment analysis is to view subjective information that will be extracted from a given text (Turney, 2002).

Different names are given to sentiment analysis such as subjectivity analysis, review mining, opinion mining, and appraisal extraction (Pang & Lee, 2008). More officially, sentiment analysis can be well-defined as: Given a text *t* from a text set *T*,

References

Abd-Elhamid, L., Elzanfaly, D., & Eldin, S. (2016). Feature-based sentiment analysis in online Arabic reviews. Proceedings of 2016 11th International Conference on Computer Engineering and Systems, ICCES 2016, 260–265.

Abdul-Mageed, M. (2017). Modeling Arabic subjectivity and sentiment in lexical space. Information Processing and Management, 1, 1–17.

Abdul-Mageed, M., & Diab, M. (2012). Awatif: A multi-genre corpus for modern standard Arabic subjectivity and sentiment analysis. Proceedings of the Language Resources and Evaluation (LREC'12), Istanbul, (November), 3907–3914.

Abdul-Mageed, M., & Diab, M. (2014). SANA: A large scale multi-genre, multi-dialect lexicon for Arabic subjectivity and sentiment analysis. Proceedings of the Language Resources and Evaluation Conference, 1162–1169.

Abdul-Mageed, M., Kübler, S., & Diab, M. (2012). SAMAR: A system for subjectivity and sentiment analysis of Arabic social media. 12 Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, (July), 19–28.

Abdul-Mageed, M., Diab, M., Korayem, M. (2013). Subjectivity and Sentiment Analysis of Modern Standard Arabic and Arabic Microblogs. Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, (3), 55–64.

Abdulla, N. A., Ahmed, N. A., Shehab, M. A., Al-Ayyoub, M., Al-Kabi, M. N., & Al-rifai, S. (2014). Towards improving the lexicon-based approach for Arabic sentiment analysis. International Journal of Information Technology and Web Engineering. (IJITWE), 9(3), 55-71.

Abdulla, N. A., Ahmed, N. A., Shehab, M. A., & Al-ayyoub, M. (2013). Arabic sentiment analysis: Corpus-based and lexicon-based. Proceedings of Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT´13), 6(12), 1–6.

Abdulla, N. A., Majdalawi, R., Mohammed, S., Al-Ayyoub, M., & Al-Kabi, M. (2014). Automatic lexicon construction for Arabic sentiment analysis. Proceedings of 2014 International Conference on Future Internet of Things and Cloud, FiCloud 2014, 547–552.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. Proceedings of the Workshop on Languages in Social Media, (June), 30–38.

Al-Ayyoub, M., Essa, S. B., & Al-Smadi, I. (2015). Lexicon-based sentiment analysis of Arabic tweets. International Journal of Social Network Mining, 2, 101–114.

Al-Aziz, A., Gheith, M., & Eldin, A. S. (2016). Lexicon based and multi-criteria decision making (MCDM) approach for detecting emotions from Arabic microblog text. Proceedings - 1st International Conference on Arabic Computational Linguistics: Advances in Arabic Computational Linguistics, ACLing 2015, 100–105.

Al-Harbi, W. & Emam, A. (2016). Effect of Saudi dialect preprocessing on Arabic sentiment analysis. International Journal of Advanced Computer Technology (IJACT), 91–99.

Al-Hasan, A. (2016). Building a sentiment lexicon for the Palestinian dialect. (Master`s thesis). The Islamic University - Gaza, Gaza, Palestine.

Al-Horaibi, L., & Khan, M. B. (2016). Sentiment analysis of Arabic tweets using text mining techniques. International Journal of Computing & Information Sciences, 12(2), 100111F.

Al-Kabi, M., Al-Ayyoub, M., Al-Smadi, I., & Wahsheh, H. (2016). A prototype for a standard Arabic sentiment analysis corpus. International Arab Journal of Information Technology, 13(1A), 163–170.

Al-Kabi, M., Al-Qudah, N., Al-Smadi, I., Dabour, M., & Wahsheh, H. (2013). Arabic/English sentiment analysis: An empirical study. The Fourth International Conference on Information and Communication Systems (ICICS 2013), (October 2015), 23-25.

Al-Kabi, M., Gigieh, A., Al-Smadi, I., Wahsheh, H., & Haidar, M. (2013). An opinion analysis tool for colloquial and standard Arabic. The Fourth International Conference on Information and Communication Systems (ICICS 2013), (April), 23-25.

Al-Kabi, M. N., Gigieh, A. H., Alsmadi, I. M., Wahsheh, H. A., & Haidar, M. M. (2014). Opinion mining and analysis for Arabic language. IJACSA)

International Journal of Advanced Computer Science and Applications, 5(5), 181-195.

Al-Moslmi, T., Al-Bared, M., Al-Shabi, A., Omar, N., & Abdullah, S. (2017). Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis. Journal of Information Science, 44(3), 345-362.

Al-Osaimi, S., & Badruddin, K. M. (2014). Role of emotion icons in sentiment classification of Arabic tweets. Proceedings of the 6th International Conference on Management of Emergent Digital EcoSystems - MEDES '14, (September), 167–171.

Al-Saffar, A., Sabri, B., Tao, H., Awang, S., Abdul-Majid, M., & Al-Saiagh, W. (2016). Sentiment analysis in Arabic social media using association rule mining. Journal of Engineering and Applied Sciences, (Special Issue 2), 3239–3247.

Al-Twairesh, N., Al-Khalifa, H., Al-Salman, A., & Al-Ohali, Y. (2017). Sentiment Analysis of Arabic Tweets : Feature Engineering and A Hybrid Approach, 1–10.

Al-Twairesh, N., Al-Khalifa, H., & Al-Salman, A. (2016). AraSenTi : Large-scale twitter-specific Arabic sentiment lexicons. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016), 1, 697–705.

Alayba, A. M., Palade, V., England, M., & Iqbal, R. (2017, April). Arabic language sentiment analysis on health services. In Arabic Script Analysis and Recognition (ASAR), 2017 1st International Workshop on (pp. 114-118). IEEE.

Albraheem, L., & Al-Khalifa, H. S. (2012). Exploring the problems of sentiment analysis in informal Arabic. Proceedings of the 14th International Conference on Information Integration and Web-Based Applications & Services - IIWAS '12, 415-418.

Aldayel, H. K., & Azmi, A. M. (2015). Arabic tweets sentiment analysis - a hybrid scheme. Journal of Information Science, 42(6), 782-797.

Alotaibi, S. S. (2015). Sentiment analysis in the Arabic language using machine learning (Doctoral dissertation.) Colorado State University, Fort Collins, Colorado, USA.

Alotaibi, S., & Khan, M. B. (2017). Sentiment analysis challenges of informal Arabic language. (IJACSA) International Journal of Advanced Computer Science and Applications, 8(2), 278–284.

Althobaiti, M., Kruschwitz, U., & Poesio, M. (2014). AraNLP: A Java-based library for the processing of Arabic text. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), 4134–4138.

Altrabsheh, N. (2016). Sentiment analysis on students' real-time feedback. Diss. University of Portsmouth.

Amolik, A., Jivane, N., Bhandari, M., & Venkatesan, M. (2016). Twitter sentiment analysis of movie reviews using machine learning technique. International Journal of Engineering and Technology, 7(6), 2038–2044.

Amram, A., Ben-David, A., & Tsarfaty, R. (2018). Representations and Architectures in Neural Sentiment Analysis for Morphologically Rich Languages: A Case Study from Modern Hebrew. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 2242-2252).

Appel, O., Chiclana, F., & Carter, J. (2015). Main concepts, state of the art and future research questions in sentiment analysis. Acta Polytechnica Hungarica, 12(3), 87–108.

Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. Knowledge-Based Systems, 108, 110–124.

Araque, O., Corcuera, I., Román, C., Iglesias, C. A., & Sánchez-Rada, J. F. (2015). Aspect Based Sentiment Analysis of Spanish Tweets. In TASS@ SEPLN (pp. 29-34).

Assiri, A., Emam, A., & Al-Dossari, H. (2017). Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis. Journal of Information Science, 44(2), 184-202.

Assiri, A., Emam, A., & Al-Dossari, H. (2015). Arabic sentiment analysis: A survey. (IJACSA) International Journal of Advanced Computer Science and Applications, 6(12), 75–85.

Assiri, A., Emam, A., & Al-Dossari, H. (2016). Saudi twitter corpus for sentiment analysis. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 10(2), 272-275.

Astudillo, R. F., Amir, S., Ling, W., Martins, B., Silva, M., & Trancoso, I. (2015). INESC-ID: A regression model for large scale twitter sentiment lexicon induction. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 613–618.

Awwad, H., & Alpkocak, A. (2016, September). Performance Comparison of Different Lexicons for Sentiment Analysis in Arabic. In 2016 Third European Network Intelligence Conference (ENIC) (pp. 127-133). IEEE.

Badaro, G., Baly, R., & Hajj, H. (2014). A large scale Arabic sentiment lexicon for Arabic opinion mining. Arabic Natural Language Processing Workshop Co-Located with EMNLP 2014, 176–184.

Baly, R., Khaddaj, A., Hajj, H., El-Hajj, W., & Shaban, K. (2018). ArSentD-LEV : A multi-topic corpus for target-based sentiment analysis in Arabic Levantine tweets. Proceedings of the 3rd Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT).

Batrinca, B., & Treleaven, P. C. (2014). Social media analytics: A survey of techniques, tools, and platforms. AI and Society, 30(1), 89–116.

Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An overview of sentiment analysis in social media and its applications in disaster relief. Studies in Computational Intelligence, 639, 313–340.

Benevenuto, F., Araújo, M., & Ribeiro, F. (2015). Sentiment analysis methods for social media. Proceedings of the 21st Brazilian Symposium on Multimedia and the Web - WebMedia '15, 11–11.

Bennett, P. R. (1998). Comparative Semitic linguistics: a manual. Eisenbrauns.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1–8.

Bradley, M. M., & Lang, P. P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. Psychology, Technical(C-1), 30(1), 25-36.

Brahimi, B., Touahria, M., & Tari, A. (2016). Data and text mining techniques for classifying Arabic tweet polarity. Journal of Digital Information Management, 14(1), 15–25.

Bravo-Marquez, F., Mendoza, M., & Poblete, B. (2013). Combining strengths, emotions, and polarities for boosting Twitter sentiment analysis. Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining - WISDOM '13, 1–9.

Brody, S., & Diakopoulos, N. (2011). Cooooooooooooooollllllllllllll !!!!!!!!!!!!!! using word lengthening to detect sentiment in microblogs. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 562–570.

Buckwalter, T. 2002. Arabic Morphological Analyzer (AraMorph). (2002).

Burnap, P., Gibson, R., Sloan, L., Southern, R., & Williams, M. (2016). 140 Characters to victory?: Using Twitter to predict the UK 2015 general election. Electoral Studies, 41, 230–233.

Guerra, P. H., Veloso, A., Meira, Jr., W., & Almeida, V. (2011). From bias to opinion: A transfer-learning approach to real-time sentiment analysis. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 150–158.

Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010). SenticNet : A publicly available semantic resource for opinion mining. Artificial Intelligence, 10, 14–18.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. Proceedings of the 27th Annual Meeting on Association for Computational Linguistics -, 76–83.

D'Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools, and applications for sentiment analysis implementation. International Journal of Computer Applications, 125(3), 26–33.

Dasgupta, S., & Ng, V. (2009). Topic-wise, sentiment-wise, or otherwise?: Identifying the hidden dimension for unsupervised text classification. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2 (pp. 580-589). Association for Computational Linguistics.

Davidov, D., Tsur, O., & Rappoport, A. (2010). Semi-supervised recognition of sarcastic sentences in Twitter and Amazon. Fourteenth Conference on Computational Natural Language Learning, (7), 107–116.

Diab, M., Hacioglu, K., & Jurafsky, D. (2007). Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks. HLT-NAACL 2004: Short Papers, 149–152.

Ding, C., Li, T., Peng, W., & Park, H. (2006). Orthogonal nonnegative matrix t-factorizations for clustering. Proceedings of the 12th ACM SIGKDD, 126–135.

Dodds, P. S., & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. Journal of Happiness Studies, 11(4), 441–456.

Duwairi, R. M., Ahmed, N. A., & Al-Rifai, S. Y. (2015). Detecting sentiment embedded in Arabic social media - A lexicon-based approach. Journal of Intelligent and Fuzzy Systems, 29(1), 107–117.

Eisenstein, J. (2013). What to do about bad language on the internet. Proceedings of the 2013 Conference of the North American Chapter of the association for computational linguistics: Human language technologies Naacl-Hlt, 359–369.

El-Beltagy, S. R., & Ali, A. (2013). Open issues in the sentiment analysis of Arabic social media: A case study. In Innovations in Information Technology (IIT), 2013 9th, 1–6.

El-Beltagy, S. R. (2016). NileULex: A phrase and word level sentiment lexicon for Egyptian and modern standard Arabic. In Proceedings of LREC 2016, (4), 2900–2905.

El-Beltagy, S. R. (2017). WeightedNileULex: A scored arabic sentiment lexicon for improved sentiment analysis. Language Processing, Pattern Recognition, and Intelligent Systems. Special Issue on Computational Linguistics, Speech& Image Processing for Arabic Language, (2).

El-Khair, I. (2006). Effects of stop words elimination for Arabic information retrieval: A comparative study. International Journal of Computing & Information Sciences, 4(3), 119–133.

El-Masri, M., Altrabsheh, N., & Mansour, H. (2017). Successes and challenges of Arabic sentiment analysis research: A literature review. Social Network Analysis and Mining, 7(1).

Elarnaoty, M., Abdel-Rahman, S. & Fahmy, A. (2012) A machine learning approach for opinion holder extraction in Arabic language. International Journal of Artificial Intelligence & Applications (IJAIA), 3(2), 45–63.

Elawady, R. M., Barakat, S., El-Bakry, H. M., & Elrashidy, N. M. (2015). Sentiment analysis for Arabic and English dataset. International Journal of Intelligent Computing and Information Science, 15(1), 10–14.

Elhawary, M., & Elfeky, M. (2010). Mining Arabic business reviews. Proceedings of IEEE International Conference on Data Mining, ICDM, 1108–1113.

Elsahar, H., & El-Beltagy, S. R. (2014). A fully automated approach for Arabic slang lexicon extraction from microblogs. International Conference on Intelligent Text Processing and Computational Linguistics, (4), 79–91.

Eskander, R., & Rambow, O. (2015). SLSA: A sentiment lexicon for standard Arabic. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), (9), 2545–2550.

Farra, N., Challita, E., Assi, R. A., & Hajj, H. (2010). Sentence-level and document-level sentiment mining for Arabic texts. Proceedings of IEEE International Conference on Data Mining, ICDM, 1114–1119.

Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., & Smith, N. (2010). Part-of-speech tagging for twitter: Annotation, features, and experiments. Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science.

Gligoric, K., Anderson, A., & West, R. (2018). How Constraints Affect Content: The Case of Twitter's Switch from 140 to 280 Characters. arXiv preprint arXiv:1804.02318.

Golder, S. A., & Macy, M. W. (2013). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science, 333(6051), 1878-1881.

Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2013). Comparing and combining sentiment analysis methods. Proceedings of the First ACM Conference on Online Social Networks - COSN '13, 27–38.

Habash, N., Jarrar, M., Alrimawi, F., Akra, D. F., Zalmout, N., Bartolotti, E., & Arar, M. A. (2016). Palestinian Arabic conventional orthography guidelines.

Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. Procedia Computer Science, 17, 26–32.

Hailong, Z., Wenyan, G., & Bo, J. (2014). Machine learning and lexicon based methods for sentiment classification: A survey. In 2014 11th Web Information System and Application Conference (pp. 262-265). IEEE.

Hamouda, A. E., & El-Taher, F. E. (2013). Sentiment analyzer for Arabic comments system. International Journal of Advanced Computer Science and Applications, 4(3), 99–103.

Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. Proceedings of the 35th Annual Meeting on Association for Computational Linguistics, 174–181.

Heerschop, B., Goossen, F., Hogenboom, A., Frasincar, F., Kaymak, U., & Jong, F. (2011). Polarity analysis of texts using discourse structure. Proceedings of the 20th ACM International Conference on Information and Knowledge Management - CIKM ’11, 1061-1070.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD,(8), 168-177.

Hu, X., Tang, J., Gao, H., & Liu, H. (2013). Unsupervised sentiment analysis with emotional signals. International Conference on World Wide Web, 607–617.

Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-14), 216–225.

Ibrahim, H. S., Abdou, S. M., & Gheith, M. (2015). MIKA: A tagged corpus for modern standard Arabic and colloquial sentiment analysis. Proceedings of 2015 IEEE 2nd International Conference on Recent Trends in Information Systems, 4(2), 353–358.

Ibrahim, H. S., Abdou, S. M., & Gheith, M. (2016). Automatic expandable large-scale sentiment lexicon of modern standard Arabic and colloquial. Proceedings of 1st International Conference on Arabic Computational Linguistics: Advances in Arabic Computational Linguistics, ACLing 2015, (4), 94–99.

Ihnaini, B., & Mahmuddin. M., 2018. Lexicon-Based Sentiment Analysis of Arabic Tweets: A Survey. Journal of Engineering and Applied Sciences, 13: 7313-7322.

Itani, M., Roast, C., & Al-Khayatt, S. (2017). Corpora for sentiment analysis of Arabic text in social media. Proceedings of 2017 8th International Conference on Information and Communication Systems, ICICS 2017, 64–69.

Jarrar, M., Habash, N., Akra, D., & Zalmout, N. (2014). Building a corpus for Palestinian Arabic : A preliminary study. Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), 18–27.

Jarrar, M., Habash, N., Al-Rimawi, F., Akra, D., & Zalmout, N. (2017). Curras: An annotated corpus for the Palestinian Arabic dialect. Language Resources and Evaluation, 51(3), 745–775.

Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011). Target-dependent Twitter sentiment classification. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies,1, 151–160.

Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. Computational Intelligence, 22(2), 110–125.

Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. Decision Support Systems, 57(1), 245–257.

Khan, F. H., Qamar, U., & Bashir, S. (2016). SentiMI: Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection. Applied Soft Computing Journal, 39, 140–153.

Kharde, V. A., & Sonawane, S. S. (2016). Sentiment analysis of Twitter data: A Survey of techniques. International Journal of Computer Applications, 139(11), 5–15.

Khatua, A., Ghosh, K., & Chaki, N. (2015). Can #Twitter-Trends predict election results? Evidence from 2014 Indian general election. Proceedings of the Annual Hawaii International Conference on System Sciences,1676–1685.

Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research, 50, 723-762.

Korayem, M. (2016). Sentiment/subjectivity analysis survey for languages other than English. Social Network Analysis and Mining, 6(1), 1–26.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. Journal of Manufacturing Science and Engineering, Transactions of the ASME, 31, 249–268.

Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg! Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 11(164), 538–541.

Larkey, L., Ballesteros, L., & Connell, M. (2007). Light stemming for Arabic information retrieval. Arabic Computational Morphology, 221–243.

Liu, B. (2011). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

Mataoui, M. H., Zelmati, O., & Boumechache, M. (2016). A proposed lexicon-based sentiment analysis approach for the vernacular Algerian Arabic. Research in Computing Science, 110, 55-70.

Medhat, W., Hassan, A., & Korashy, H. (2014a). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113.

Medhat, W., Hassan, A., & Korashy, H. (2014b). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113.

Media, A. S. (2017). Social Media and the Internet of Things.

Mobarz, H., Rashown, M., & Farag, I. (2014). Using automated lexical resources in Arabic sentence subjectivity. International Journal of Artificial Intelligence and Applications (IJAIA), 5(6), 1-14.

Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. Proceedings of the Seventh International Workshop on Semantic Evaluation Exercises, 2(SemEval), 321–327.

Mohammad, S. M., Salameh, M., & Kiritchenko, S. (2016a). How translation alters sentiment. Journal of Artificial Intelligence Research, 55(1), 95–130.

Mohammad, S. M., Salameh, M., & Kiritchenko, S. (2016b). Sentiment lexicons for Arabic social media. Tenth International Conference on Language Resources and Evaluation, LREC 2016, (9), 33–37.

Mudinas, A., Zhang, D., & Levene, M. (2012). Combining lexicon and learning based approaches for concept-level sentiment analysis. Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining - WISDOM, 12, 1–8.

Mustafa, M., Alsamahi, A. S., & Hamouda, A. (2017). New Avenues in Arabic Sentiment Analysis. International Journal of Scientific and Engineering Research, 8(5), 907–915.

Nabil, M. (2015). ASTD : Arabic Sentiment Tweets Dataset, (September), 2515–2519.

Nagar, Y., & Malone, T. W. (2011). Making business predictions by combining human and machine intelligence in prediction markets. International Conference on Information Systems ICIS 2011, 1–16.

Narayanan, V., Arora, I., & Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced Naive Bayes model. International Data Engineering and Automated Learning, Lecture Notes in Computer Science, 8206, 194–201.

Nielsen, F. A. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. CEUR Workshop Proceedings, 718, 93–98.

Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 1320–1326.

Paltoglou, G., Thelwall, M., & Buckely, K. (2010). Online textual communication annotated with grades of emotion strength. Proceedings of the Third International Workshop on EMOTION Satellite of LREC Corpora for Research on Emotion and Affect, 25–31.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 1(2), 1–135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. Empirical Methods in Natural Language Processing (EMNLP), 10(7), 79–86.

Patel, S. N., & Choksi, M. J. B. (2015). A survey of sentiment classification techniques. Journal for Research| Volume, 1(01).

Patra, B. G., Kundu, A., Das, D., & Bandyopadhyay, S. (2012). Classification of interviews-A case study on cancer patients. In Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology (pp. 27-36).

Previtali, F., Arrieta, A. F., & Ermanni, P. (2015). Double-walled corrugated structure for bending-stiff anisotropic morphing skins. Journal of Intelligent Material Systems and Structures, 26(5), 599–613.

Purver, M., & Battersby, S. (2012). Experimenting with Distant Supervision for Emotion Classification. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 482–491.

Refaee, E., & Rieser, V. (2014). An Arabic twitter corpus for subjectivity and sentiment analysis. Proceedings of the Language Resources and Evaluation Conference, (spring 2013), 2268–2273.

Refaee, E., & Rieser, V. (2015). Benchmarking Machine Translated Sentiment Analysis for Arabic Tweets. Proceedings of NAACL-HLT 2015 Student Research Workshop (SRW), 71–78.

Refaee, E., & Rieser, V. (2016). iLab-Edinburgh at SemEval-2016 Task 7 : A Hybrid Approach for Determining Sentiment Intensity of Arabic Twitter Phrases. Proceedings of SemEval-2016, 474–480.

Reitan, J., Faret, J., Gambäck, B., & Bungum, L. (2015). Negation Scope Detection for Twitter Sentiment Analysis. Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA'15), (Wassa), 99–108.

Rodriguez-Penagos, C., Atserias, J., Codina-Filbà, J., Garcia-Narbona, D., Grivolla, J., Lambert, P., & Sauri, R. (2013). FBM: Combining lexicon-based ML and heuristics for Social Media Polarities. Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), 2(SemEval), 483–489.

Salameh, M., Mohammad, S., & Kiritchenko, S. (2015). Sentiment after translation: A case-study on Arabic social media posts. In Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies (pp. 767-777)

Saleh, S. A. (2015). Sentiment Analysis in the Arabic Language Using Machine Learning. Colorado State University. Colorado State University.

Schouten, K., & Frasincar, F. (2016). Survey on Aspect-Level Sentiment Analysis. IEEE Transactions on Knowledge and Data Engineering, 28(3), 813–830.

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. Information Sciences, 311(August 2015), 18–38.

Sheela, L. (2016). A Review of Sentiment Analysis in Twitter Data Using Hadoop. International Journal of Database Theory and Application, 9(1), 77–86.

Shoukry, A., & Rafea, A. (2012). Sentence-level Arabic sentiment analysis. Proceedings of the 2012 International Conference on Collaboration Technologies and Systems, CTS 2012, 546–550.

Siddiqui, S., Monem, A. A., & Shaalan, K. (2016, October). Towards improving sentiment analysis in Arabic. In International Conference on Advanced Intelligent Systems and Informatics (pp. 114-123). Springer, Cham.

Silva, I. S., Gomide, J., Veloso, A., Meira, W. J., & Ferreira, R. (2011). Effective sentiment stream analysis with self-augmenting training and demand-driven projection. Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, 11, 475–484.

Smailović, J. (2014). Sentiment analysis in streams of microblogging posts (Doctoral dissertation, Ph.D. thesis, Jožef Stefan International Postgraduate School, Ljubljana, Slovenia).

Socher, R., Pennington, J., & Huang, E. (2011). Semi-supervised recursive autoencoders for predicting sentiment distributions. Conference on Empirical Methods in Natural Language Processing, EMNLP, (7), 151–161.

Suttles, J., & Ide, N. (2013). Distant supervision for emotion classification with discrete binary values. Proceeding of International Conference on Intelligent Text Processing and Computational Linguistics, 121–136.

Taboada, M., Brooke, J., & Tofiloski, M. (2011). Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2), 267–307.

Taboada, M. (2016). Sentiment analysis: an overview from linguistics.

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 63(1), 163–173.

Thelwall, M., Buckley, K., Paltoglou, G., & Cai, D. (2010). Sentiment strength detection in short informal text. The American Society for Informational Science and Technology, 61(12), 2544–2558.

Thelwall, M., & Prabowo, R. (2009). Sentiment analysis: A combined approach. Journal of Informetrics, 3(2), 143–157.

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In Fourth international AAAI conference on weblogs and social media. 10(1), 178-185.

Turney, P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), (July), 417–424.

Turney, P., & Littman, M. (2003). Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems, 21(4), 315–346.

Vinodhini, G., & Chandrasekaran, R. (2012). Sentiment analysis and opinion mining: A survey. International Journal of Advanced Research in Computer Science and Software Engineering, 2(6), 282–292.

Volkova, S., Wilson, T., & Yarowsky, D. (2013). Exploring sentiment in social media: Bootstrapping subjectivity clues from multilingual Twitter streams.

Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics 2, 505–510.

Wang, X., Wei, F., Liu, X., Zhou, M., & Zhang, M. (2011). Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach. Proceedings of the 20th ACM International Conference on Information and Knowledge Management, 1031–1040.

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. Language Resources and Evaluation, 39(2–3), 165–210.

Wilson, T., Wiebe, J., & Hoffman, P. (2005). Recognizing contextual polarity in phrase level sentiment analysis. Proceedings of the conference on human language technology and empirical methods in natural language processing, 347-354.

Xianghua, F., Guo, L., Yanyan, G., & Zhiqiang, W. (2013). Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. Knowledge-Based Systems, 37, 186-195.

Yang, J., & Leskovec, J. (2011). Patterns of temporal variation in online media. Proceedings of the Fourth ACM International conference on Web search and data mining, 177-186.

Yuan, B. (2016). Sentiment analysis of Twitter data. Linguistic Rules Lexicon-based.

Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2015). Combining lexicon-based and learning-based methods for Twitter sentiment analysis. International

Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE), 89, 1–8.

List of Appendices

Appendix A

Tweepy Code for Collecting Tweets

```
#!/usr/bin/env python

# -*- coding: utf-8 -*-

from __future__ import unicode_literals

from tweepy import Stream

from tweepy import OAuthHandler

from tweepy.streaming import StreamListener

import tweepy

import csv

import pandas as pd

#consumer key, consumer secret, access token, access secret.

ckey = "consumer key"

csecret = "consumer secret"

atoken = "access token"

asecret = "access secret"

auth = OAuthHandler(ckey, csecret)

auth.set_access_token(atoken, asecret)

api = tweepy.API(auth,wait_on_rate_limit=True)

# Open/Create a file to append data

csvFile = open('Dataset1.csv', 'a')
```

```
#Use csv Writer

csvWriter = csv.writer(csvFile)



for     tweet     in     tweepy.Cursor(api.search,q="key-word     or
emoticon",count=10000,lang="ar",since="DATE").items():

    print (tweet.created_at, tweet.text)

    csvWriter.writerow([tweet.text.encode('utf-8')])
```

Appendix B

Code of Expanding EULA


```
    def insertToEula(self,data,pol):

        logger.debug("insert To Eula(%s,%s)" ,data,pol,totalP,totalN)

        word = data[0]

        counterP = data[1]

        counterN = data[2]

        counterSum = data[3]

        polarity = 0

        stepper = 1

        if pol == "negative":

            counterN += stepper

        if pol == "positive":

            counterP += stepper
```

```
            counterSum = counterP + counterN

            tfP = counterP / (counterP + counterN)

            tfN = counterN / (counterP + counterN)

           idfP = (totalP + totalN)/totalP

            idfN = (totalP + totalN)/totalN

            if ( counterP / totalP) > ((counterN / totalN)+alpha):

           polarity = tfP * idfP * counterP / totalP

            elif (( counterP / totalP)+ alpha) < ((counterN / totalN):

             polarity = tfN * idfN * counterN / totalN      else :

           polarity = 0

           query  = '''  UPDATE  eula_words  SET  counterP=?,counterN=?,
polarity=? WHERE word=?

           '''

           data = (counterP,counterN,polarity,word)

           run = self.writeSingle(query,data)

       def sendToEula(self,word,pol):

           data = self.isEula(word)

           if data == False:

               self.insertNewEula(word,pol)

           else:

               self.insertExistsEula(data,pol)
```

Appendix C

Implementation of Contrast Rules

```
import sys

import logging

logger = logging.getLogger("mylog")

class ContrastRule:

    def __init__(self,tweetparser,db):

        logger.info("Applaying contrast rule")

        self.db = db

        self.tp = tweetparser

        self.contrast_index_1 = []

        self.contrast_index_2 = []

        self.searchContrastIndex()

        self.reConstructSentence()

        self.applyRule(self.contrast_index_1)

        self.applyRule(self.contrast_index_2)

    def searchContrastIndex(self):

        logger.debug("search for Contrast()")

        for i in xrange(self.tp.getTweetSentenceLength()):

            tuple_index = ()

            for j in xrange(self.tp.getWordLength(i)):

                data = self.db.isContrast(self.tp.getWordOf(i,j))
```

```
                if data != "null":
                    if type(data) == tuple:
                        if str(data[2]) == "null":
                            tuple_index = (i,j,str(data[3]))
self.contrast_index_1.append(tuple_index)
                        if str(data[2]) != "null":
                            second_word = self.tp.getNextWord(i,j)
                            logger.debug("Comparing %s with
%s",second_word,data[2])
                            if second_word == str(data[2]):
                                tuple_index = (i,j,str(data[3]))
                                logger.debug("Found 2 indexes of
contrast word.")
self.contrast_index_2.append(tuple_index)
                    if type(data) == list:
                        for k in data:
                            logger.debug("Comparing %s with
%s",self.tp.getNextWord(i,j),k[2])
                            if str(k[2]) == "null":
                                tuple_index = (i,j,str(k[3]))
self.contrast_index_1.append(tuple_index)
                            if k[2] == self.tp.getNextWord(i,j):
                                logger.debug("Found 2 indexes of
contrast word.")
```

```
                                tuple_index = (i,j,str(k[3]))

self.contrast_index_2.append(tuple_index)



    def reConstructSentence(self):
        logger.debug("Checking for reconstruction of sentence")
        if self.contrast_index_2:
            logger.debug("Reconstructing the sentence")
            for contrast in reversed(self.contrast_index_2):
                i = contrast[0]
                j = contrast[1]
                contType = contrast[2]
                temp_word_1 = self.tp.getWordOf(i,j)
                temp_word_2 = self.tp.getNextWord(i,j)
                new_word = temp_word_1 + ' ' + temp_word_2
                self.tp.removeWordfromIndex(i,j+1)
                self.tp.setWord(i,j,new_word)
                self.setContrast(i,j,contType)
        if self.contrast_index_1:
            for i in self.contrast_index_1:
                self.setContrast(i[0],i[1],i[2])
```

```
def applyRule(self,contrast):
    if contrast:
        for i in contrast:
            if i[2] == "C1":
                index_i = i[0]
                index_j = i[1]
                self.contrastRule1(index_i,index_j)
            else:
                index_i = i[0]
                index_j = i[1]
                self.contrasRule2(index_i,index_j)

def contrastRule1(self,index_i,index_j):
    logger.debug("Applying rule 1 for sentence %s",index_i)
    for i in xrange(0,index_j):
        score = self.tp.getScore(index_i,i)
        logger.debug("Index: %s with score : %s",i,score)
        if score:
            score = score*3.0
            logger.debug("New score : %s",score)
            self.tp.resetScore(index_i,i,score)


def contrasRule2(self,index_i,index_j):
```

```
        logger.debug("Applying rule 2")
        for i in xrange(index_j+1,self.tp.getWordLength(index_i)):
            score = self.tp.getScore(index_i,i)
            logger.debug("Index: %s with score : %s",i,score)
            if score:
                score = score*3.0
                self.tp.resetScore(index_i,i,score)

    def setContrast(self,index_i,index_j,contType):
        self.tp.setType(index_i,index_j,"Contrast-"+str(contType))
    def getContrastIndex(self):
        return 'Index 1:' + str(self.contrast_index_1) + ' Index 2' +
str(self.contrast_index_2)
```


Appendix D
Implementation of Intensifier Rules

```
import sys
import logging
logger = logging.getLogger("mylog")
class Intensify:
    def __init__(self,tweetparser,db):
        self.tp = tweetparser
        self.db = db
        self.intens_index = []
```

```
    logger.info("Applying intensifier rule")


    self.searchIntens()
    for i in self.intens_index:
        self.setIntense(i[0],i[1])
    self.applyIntenseRule()
def searchIntens(self):
    for i in xrange(self.tp.getTweetSentencesLength()):
        tuple_index = ()
        for j in xrange(self.tp.getWordLength(i)):
            data = self.db.isIntense(self.tp.getWordOf(i,j))
            if data:
                tuple_index = (i,j,data[0])
                self.intens_index.append(tuple_index)
def applyIntenseRule(self):
    logger.debug("Applying intensifier rule")
    if self.intens_index:
        for i in self.intens_index:
            next_score = self.tp.getScore(i[0],i[1]+1)
            if next_score:
                new_score = next_score * (i[2]/100)
                self.tp.resetScore(i[0],i[1]+1,new_score)
def setIntense(self,index_i,index_j):
```

```
        self.tp.setType(index_i,index_j,"Intensifier")

    def getIntenseIndex(self):

        return self.intens_index
```

Appendix E

Implementation of Negation Rules

```
import sys

import logging

logger = logging.getLogger("mylog")

class NegationRule:

    def __init__(self,tweetparser,db,flag):

        logger.info("Applying Negation Rule 1")

        self.flag = flag

        self.db = db

        self.tp = tweetparser

        self.negation_index = []

        self.searchNegation()

        for i in self.negation_index:

            self.setNegation(i[0],i[1])

        if self.flag == 1:

            self.applyRule2()

        if self.flag == 0:

            self.applyRule1()

    def searchNegation(self):
```

```
    for i in xrange(self.tp.getTweetSentencesLength()):
        tuple_index = ()
        for j in xrange(self.tp.getWordLength(i)):
            if self.db.isNegation(self.tp.getWordOf(i,j)):
                tuple_index = (i,j)
                self.negation_index.append(tuple_index)
def applyRule1(self):
    logger.debug("Applying rule 1")
    if self.negation_index:
        for i in self.negation_index:
            for j in xrange(1,5):
                logger.debug("Checking: (%s,%s)",i[0],i[1]+j)
                next_type = self.tp.getTypeByIndex(i[0],i[1]+j)
                if next_type == "Negation"
                    break
                if next_type:
                    next_score = self.tp.getScore(i[0],i[1]+j)
                    if next_score:
                        if next_score > 2.0:
                            next_score= next_score -6
                        elif next_score <= -2.0:
                            next_score= next_score +5
                        elif next_score > 0 and next_score <= 2:
```

```
                                next_score= next_score -4
                            elif next_score > -2 and next_score < 0:
                                next_score= next_score +3
                            elif next_score == 0:
                                next_score= next_score -1
                            else:
                                pass
self.tp.resetScore(i[0],i[1]+j,next_score)

    def applyRule2(self):
        logger.debug("Applying rule 2")
        if self.negation_index:
            for i in self.negation_index:
                next_score = self.tp.getScore(i[0],i[1]+1)
                logger.debug("Got score :%s", next score)
                if next_score:
                    next_score = next_score*-1.0
                    logger.debug("Thus, new score %s",next_score)
                    self.tp.resetScore(i[0],i[1]+1,next_score)
                else:
                    pass
    def setNegation(self,index_i,index_j):
```

```
        self.tp.setType(index_i,index_j,"Negation")

    def getNegationIndex(self):

        return self.negation_index
```

Appendix F

Snapshot of Data

Example of EMAR tweets without emoticons:

Negative EMAR tweets
كنت دائم الخوف افقد مكاني الصدمة الكبرى عندما اكتشفت أن لا 😭😭 مكان لي
😭😐😢 ما ابي نحاف💔😕💔😕 ابغي سالب مربرب او متين
😕😐 انتي أصلا كلك اوفرايتد
ده ايه الغلب ده 💔يعني النهاردة التلات وكمان السيسي جاي عندنا 😢ياربي
😐😄😄 حد يقول حاجه والله زهقانه
الله لو كنا جيران والبيت يسلم عالبيت نتخاصم وأقفل الشباك وأفتحه اذا 😢😢 حنيت
😢😢 💔 لربما بعد الغياب بتسألين💔
اول اسباب افلاس المبتعثين برنامج بوستن😢 😕 😕 😕
😭😭 محتاجه الشنطه فالجامعه
💔💔 😕😕😕 واصلني هالفيديو وانا بمكتب الدوام مو حرام

Positive EMAR tweets
يارب شعور إني أكون مروقه ومبسوطه يكون كل يوم
هذا المخلوق هو اكثر واحد وثقت فيه فحياتي شكرا لاسطورة و شكرا زيزو
انا مو حولك بس استودعتك اله ليلك ونهارك نومك وصحوتك بفرحك وبحزنك استودعتك اله من كل شيء يضرك
هذا انتم يالدواسر حقين طقطقه ارحب ابو الشموخ منور يا غالي
الف مبروك اختي تستاهلين ولمليون وندعمك ان شاء اله
هذه ثقافه أنتقلت من السعوديه لإمارات علمتني الحياه
عطوني ارقامكم و بتصل بسوي عمري ماعرفكم و بنتعرف
معلومة سريعة وبريئة برشلونة بايرن ليستر يوفنتوس موناكو اتلتيكو دورتموند أبطال ريال مدريد وحده

من هيلينا الي حفصه استعجلت بالاسم يا ابن العربي
عندما يتوقف العتاب فجأة تكون النفس قد طابت

Example of EMAR tweets with emoticons:

Negative EMAR tweets
كنت دائم الخوف افقد مكاني الصدمة الكبرى عندما اكتشفت أن لا مكان لي
😔 ابغي سالب مربرب او متين ما ابي نحاف
نتي أصلا كلك اوفرايتد
يعني النهاردة التلات وكمان السيسي جاي عندنا ده ايه الغلب ده ياربي
😔💔💔 حد يقول حاجه والله زهقانه
الله لو كنا جيران والبيت يسلم عالبيت نتخاصم وأقفل الشباك وأفتحه اذا حنيت
😔😔 لربما بعد الغياب بتسألين
اول اسباب افلاس المبتعثين برنامج بوستن
😔محتاجه الشنطه فالجامعه
واصلني هالفيديو وانا بمكتب الدوام مو حرام

Positive EMAR tweets
😘يارب شعور إني أكون مروقه😮 ومبسوطه يكون كل يوم
هذا المخلوق هو اكثر واحد وثقت فيه فحياتي شكرا لاسطورة 😘و شكرا زيزو
انا مو حولك ❤❤بس استودعتك اله ليلك ونهارك نومك وصحوتك بفرحك😊😊💕 وبحزنك استودعتك اله من كل شيء يضرك
هذا انتم يالدواسر😊😊 حقين طقطقه ارحب ابو الشموخ منور يا غالي
👍👍الف مبروك😮 اختي تستاهلين ولمليون وندعمك😮😮💕 ان شاء اله
💕هذه ثقافه أنتقلت من السعوديه لإمارات علمتني الحياه

عطوني ارقامكم 👈👈و بتصل بسوي عمري ماعرفكم و بنتعرف
معلومة سريعة وبريئة😎😎 برشلونة بايرن ليستر يوفنتوس موناكو اتلتيكو دورتموند أبطال😎😎💅 ريال مدريد وحده
من هيلينا الى حفصه 😂😂استعجلت بالاسم يا ابن العربي
💋عندما يتوقف العتاب فجأة🔪 تكون النفس قد طابت

Example of PAL tweets:

Negative PAL tweets
لازم تعرف الكلام المناسب عشان تواسي حد، وإلا انطم لا تواسي
طب سدي بوزك يالله روحي احلبي الغنم
!لا تاخد عني فكرة غلط • انا بلبس نظارة لأني ما بشوف منيح مو لأني نيرد
)': معلش سامحنا يا سعادة المحامي بس انهرى وبرنا امتحانات كمان
انتو مستوعبين؟ يضربوها ويقتلوها وهيا حامل؟
صايرة اشوف نفسيات كتير، والعياذ بالله أنا بحب الناس المش نفسيات يعني
جيراننا ولعت معهم , ضرب مرتو وانهال عليها بوابل من الشتائم
ومن الأول للآخر، يلعن ابو المصاري شو انها وسخ ايدين
رح يصرلي 4 أيام مش دايقة طعم الأكل، سدت نفس مش طبيعية مشان هيك تعبانة والامتحانات قتلاني
لا هادا الجزء مني انفرك بالليمون فرك والوجع جزء اساسي صار

Positive PAL tweets
القدس بوصلتنا #قادتنا فخرنا #المصالحة#
يارب هاليوم يحمل لنا الفرج لأوضاعنا ويكون فاتحة خير لهالبلاد
ربي يوفق الجميع لما فيه خير العباد والبلاد #المصالحة
وحدة لزيزة بصراحة وطيوبة وعفويتها جميلة
"اذا كان بدك تستريح شو ما شفت قول منيح

الله علي صوتك و أحساسك موطني للفنان #محمد_عساف
♥ ما ازكى الزيت الي معصور توي وانت الي ملقط الزتون
❤ هههههه دخيل الله انا محلاك ومحلا هبالك يا زلمه
حبيبتي يا سحر .. والله بتشرف بهالحكي ويارب يقدرني بس
راح نتحرر، مية بالمية رح نتحرر احنا

Appendix G

Valence Shifter Lists

Sample of stop words

Word	English Translation	Word	English Translation
انا	Me	سبع	Seven
انت	You	لحد	Until
كلمة	Word	لقاء	Meeting
واو	And	لالمم	Nations
ذاك	That	لمّا	When
كالم	Talk	لهذا	For that
اتلكا	Both	ماذا	What
وجه	Face	مارس	March
وحدة	One	مازال	Still
كله	All	مافتئ	Keeps
كيف	How	مائة	Hundred
كان	Be	متى	When
انكا	Was	مثل	Like
على	On	مذ	Since
في	In	مرّة	Once
قال	Said	مساء	Evening
كى	So	تلك	That
وقت	Time	سوف	Will
الن ه	Because	هناك	There

Intensifier words

Word	English Translation	Intensify value
امامت	Exactly	150
دياتبال	Definitely	150
لئاه	Tremendously	150
ةصخ	Specially	150
يائزشا	Extraordinary	150
اقح	Really	150
ةقبد	Precisely	150
اغلبا	Mostly	150
ريثك	A lot	150
اجد	Extremely	150
ريثك	Numerous	150
رثك	Much more than	150
ةدشب	very	150
يائزشا	Extraordinary	150
رامب	Overwhelmingly	150
قاحس	Overwhelmingly	150
مظعا	Greater than	150
طرفم	over	150
ىلإأ	More	150
اريقت	Almost	50
داكلبا	Nearly	50
اعون	Sort of	50
ادرا	Rarely	50

قليل	Little bit	50
قلما	Rarely	50

Contrast words

Contrast 1	English Translation	Contrast 2	English Translation
رغم ان	although	رغم ذلك	Although
مع ان ه	Even though	ورغم	Despite
مع ان	Even though	اال اذا	Unless
مع اني	Even though I	اال ان ه	Only it
مع الرغم	Though	بس	But
على الرغم	Although	لكن	But
حتى وان	Even if	بل على عكس	On the contrary
بدل	Substitute	مع ذلك	Even though
وان	Even though	بل	Yet

Negation words

Word	English Translation	Word	English Translation
ما	No	ما عاد	Do not
مش	Not	لم	Did not
وال	Neither	دون	Without
ال	Don’t/ No	غير ك	Other than you
فش	Nothing	غير ها	Other than
بدون	Without	فما	No
بال ش	No need	لغير	Other than
ومش	And not	ليس	Not

لن	Will not	ليسوا	Not
لاء	No	مازال	No still
مشعتم	I didn't	نوت	Not
مشيلم	I'm not	ولن	And will not
مش	Not	وليس	and not
غير	Other than	بال	Without
وما	And not	وما	And not
بلدا	never	مو	Not
ومو	And not	ليست	Not
مطلقا	ever		

Special words

word	English Translation	word	English Translation
عديم	Lacking	فقد	Lost
عدم	Without	يفتقر	Lacks
سوء	Mis-(prefix)	نقص	Deficiency
يفتقد	Lacks	قلة	Shortage of
فقد	Lose	ندرة	Shortage of
فقدت	Lose	فتقر	Lacking

Appendix H
Experts Biography

- Laila Al-Haneene

Laila was born in Palestine in1990, attended Al-Ittihad School in Nablus Palestine, and graduated from the Beit Dajan School. She enrolled in the University of Al-Najah and focused her studies on linguistics and literature, earning her bachelor degree in linguistics and master degree in Educational & Psychological Counseling.

- Ahmad Ali

Ahmad is a Palestinian linguist who was born in Jordan in1983, attended Al-Karamah School in Ramallah, and graduated in 2005 from Al-Najah University from the Arabic Language and Literature department with a bachelor degree.

- Suha Foqaha'

Suha is a Palestinian engineer, was born in Palestine in 1991, attended Al-Beirah school in Ramallah, and graduated from Birzeit University in 2015 from the Architectural Engineering department with a bachelor degree in urban planning, with an interest in novel writing and poetry.

Appendix I

Links to Datasets

The link to the Palestinian dataset (PAL-Tweets) is :

https://github.com/baha2107/Palestinian-Tweets

The link to the EMAR-Tweets is :

https://github.com/baha2107/EMAR-Tweets