Exploring the Effects of Word Roots for Arabic Sentiment Analysis

Shereen M. Oraby
College of Engineering
and Technology
AAST, Alexandria, Egypt
shereen.oraby@aast.edu

Yasser El-Sonbaty
College of Computing
and Information Technology
AAST, Alexandria, Egypt
yasser@aast.edu

Mohamad Abou El-Nasr
College of Engineering
and Technology
AAST, Alexandria, Egypt
mnasr@aast.edu

Abstract

The inherent morphological complexity of languages such as Arabic entails the exploration of language traits that could be valuable to the task of detecting and classifying sentiment within text. This paper investigates the relevance of using the roots of words as input features into a sentiment analysis system under two distinct domains, in order to tailor the task more suitably to morphologically-rich languages such as Arabic. Different word-rooting solutions are employed in conjunction with a basic sentiment classifier, in order to demonstrate the potential of mapping Arabic words to basic roots for a language-specific development to the sentiment analysis task, showing a noteworthy improvement to baseline performance.

1 Introduction

An increasing need for quick and effective analysis of huge masses of text has sparked a revolution in the requirements of natural language processing systems, demanding an ability to handle varied types and formats of textual data for a wide range of language analysis tasks, both on the syntactic and semantic levels. The task of sentiment analysis in particular presents a unique form of text analytics due to the flourish of new opinionated web data in social media and otherwise, dealing with the detection and of opinions within a text, and then further with distinguishing their polarity.

Two main tasks are of great importance with respect to the classification of opinions in text, regardless of the language under inspection: the tasks of *subjectivity detection* in a set of statements to differentiate between purely *objective* reporting of information in the form of facts, as opposed to a *subjective* account of the information; and the task of *sentiment analysis*, which entails classifying the resultant subjective statements into a set of classes, *positive*, *negative*, and *neutral*, depending on the polarity of the opinion expressed. With respect to the level of analysis performed, individual tasks may be more relevant than others: while subjectivity analysis is relevant at the sentence level to sort out opinionated statements, sentiment analysis can be appropriate at both the sentence and the document level, if the excerpts are already defined to be subjective, and the task is to distinguish the *polarity* of the opinion being expressed (Liu, 2012).

While much research has been attributed to the task of sentiment analysis in English, fewer attempts tackle the task in other more morphologically complex languages such as Arabic, and increasing amounts of information available in these languages makes the task of Arabic sentiment analysis a very relevant one, albeit a challenge for classification systems. Such language processing tasks are made more difficult in Arabic due to the lack of resources and tools available as well, despite a growing user and content base in the language.

This paper explores the implications of reducing words to their roots in order to find common, basic, sentiment-bearing components that will relate many words to a single source, and thus help to classify a larger number of words to aid in the Arabic sentiment analysis task. This is presented through comparisons within two distinct datasets where opinions are classified based on their sentiment using roots derived from three different rooting libraries. Section 2 discusses related background information on Arabic language morphology, and some intuition behind the use of rooting as an aid to such a classification task. Section 3 details related work in the fields of subjectivity and sentiment analysis. The root-based methodology proposed is presented in Section 4, followed by

International Joint Conference on Natural Language Processing, pages 471–479,
Nagoya, Japan, 14-18 October 2013.

results, evaluation, and analysis of the performed experiments in Section 5. Section 6 presents conclusions and possible directions for future work.

2 Background Information

The challenges behind the natural language processing of languages such as Arabic stem from rich morphologies, or internal word structures (Habash, 2008), and the intricate construction of words from roots and patterns specific to the language's grammar. While Modern Standard Arabic (MSA) is the standard form of communication for written and broadcasted Arabic (Ryding, 2006), spoken Arabic exists in the form of many different dialects, all of which diverge significantly from written MSA (Habash, 2008). This makes standardizing language processing tasks in Arabic even more complicated, in addition to the problem of diacritization and text normalization for data retrieved from unregulated sources, which is often the case for the mining of data appropriate for tasks such as sentiment analysis. The following section gives a basic outline of some of the details of Arabic grammar and morphology relevant to the opinion classification task at hand as background for the proposed algorithm.

2.1 Roots in Morphologically-Rich Languages

In derivational languages such as Arabic, words are derived from sets of "roots", which are commonly two, three, or four letter words that describe a basic idea (StudyQuran, 2004). Full words in Arabic are then derived from these roots by adding vowels (and/or other consonants) around the basic root, called *affixes*, which change the word pronunciation and form word derivations (Albraheem and Al-Khalifa, 2012; Ryding, 2006). These affixes can be attached to a base, stem, or root, as either *prefixes* (inserted before the word), *infixes* (inserted within the word), or *suffixes* (inserted after the word).

As an example, the Arabic letters س ل م (*siin-laam-miim*) serve as the root for several words, including *salām*, *ʾisalām*, and *muslim* (StudyQuran, 2004), as shown in Table 1. By sharing the same root word, these three derivations also share a common base meaning. This pattern is a result of the word formation scheme in Arabic, where a root such as ك ت ب (*kaaf-ta'-ba'*) means having to do with "writing", and where most other Arabic words "having to do with writing" are derived from additions and modifications to this basic three-letter root, such as كتاب (*kitāb*, meaning *book*), كاتب (*kātib*, meaning *writer*), مكتب (*maktab*, meaning *desk*), and مكتبة (*maktabatu*, meaning *library*) (Ryding, 2006).

The idea of words carrying meaning from their basic root derivatives is different to the system of word derivation in concept-based languages, where only some subset of words, but not most, can be likened to the root system. Such patterns do exist in languages such as English, but are not a general rule for word derivation. For example, the English "consonant sequence" *s-ng*, which can be used to compose various derived words, including *s-i-ng*, *s-a-ng*, *s-u-ng*, and *s-o-ng*, each of which shares a common base meaning having to do with "vocal music". Attaching various prefixes and suffixes to these derivations also results in a wider array of words, including *sing-ing*, *sing-er*, and *un-sung* (Ryding, 2006).

This concept maps the English consonant sequence to the Arabic *root*, and the English derivations resultant from the addition of vowels and affixes to the concept of an Arabic *pattern*. The consistence of this word-derivation scheme across most of a language gives root-based languages such as Arabic a well-defined clarity for word formation, which could be used to classify words based on common meaning or sense.

2.2 Intuition Behind Root-Based Matching

Due to this word formulation and root-based derivation scheme that is prevalent Arabic, many words bearing similar meanings come from the

Arabic Word	Transliteration	English Gloss
سلام	*salām*	peace
إسلام	*ʾislām*	submission, compliance, conformance, surrender
مسلم	*muslim*	one who submits, complies, conforms, surrenders

Table 1: Various Arabic Word Derivations for the Root Word *Siin-Laam-Miim* (StudyQuran, 2004).

same root, which in itself holds the "idea" that the derivations express.

It is this morphological property that can be exploited to enhance the efficiency of an automatic sentiment classification system. The proposed method seeks to use different rooting techniques to reduce input feature words to their most basic roots, thus mapping a larger number of words to matching source roots. Sentiment-bearing roots, once found recurrently in a positive or negative context, can be used to classify many more words than the derivations themselves, allowing for classification of a broader feature set.

Two sets of sentiment-bearing words that are derived from the positive-sentiment root ن ج ح (*nun-jim-ha´*, having to do with "success") and the negative-sentiment root ق ت ل (*qaf-taʾ-lam*, having to do with "killing"), are shown in Table 2 (with their respective transliterations and translations). Various derivations are shown in matching positive and negative contexts, where the root word is the same, and the meaning of the sentence, likewise, retains the same sentiment orientation.

Because the task of sentiment analysis is not enclosed at the word-level, and because the surrounding words in a phrase may change the meaning of the phrase significantly as in the case of polarity incrementing or decrementing words (or even entirely, as in the case of negation words) a root-matching scheme on its own is not sufficient for consistently accurate sentiment classification. One common handling of such problems as negations in English is to consider all words between the negation and the next clause-level punctuation mark as negative (Pang et al., 2002; Sanjiv and Chen, 2001). In an Arabic context, a more flexible free word-ordering makes such a method difficult to consistently match, so the task would require a more elaborate handling scheme. Still, the initial root-matching task can be used to enhance results as a building block for an automatic Arabic sentiment analysis system.

3 Related Work

While there has been much work on sentiment analysis in English, few examples of work on the

Arabic Word	**Positive Word Context Excerpt**
انجح *ānǧḥ* "the most successful"	يمكن اعتبار [...] من انجح المدربين *ymkn āʿtbār [...] mn ānǧḥ ālmdrbyn* "[...] can be considered one of the most successful coaches"
النجاحات *ālnǧāḥāt* "successes"	النجاحات العديدة التي حقّقها مع المنتخب *ālnǧāḥāt ālʿdydh ālty ḥqqhā mʿ ālmntḫb* "the many success that [he] achieved with the team"
نجحت *nǧḥt* "succeeded"	نجحت [...] في تجديد عقود ابرز نجوم الفريق *nǧḥt [...] fy tǧdyd ʿqwd ābrẓ nǧwm ālfryq* "[...] succeeded to renew the contracts of the most prominent team stars"

Arabic Word	**Negative Word Context Excerpt**
قتل *qtl* "were killed"	قتل ١٠٧ اشخاص في المواجهات *qtl 107 āšḫāṣ fy ālmwāǧhāt* "107 people were killed in clashes"
مقتل *mqtl* "the killing [of]"	كانت حصيلة اولية اشارت مساء الجمعة الي مقتل اثنين *kānt ḥṣylh āwlyh āšārt msāʾ ālǧmʿh āly mqtl āṯnyn* "the initial toll on Friday evening indicated the killing of two"
قتلي *qtly* "victims"	اعلن [...] مسوءوليته عن هجومين [...] اوقعا خمسة قتلي *āʿln [...] msūʾwlyth ʿn hǧwmyn [...] āwqʿā ḫmsh qtly* "[he] claimed responsibility for two attacks leaving five victims"

Table 2: Two Sets of Sentiment-Bearing Words Derived from Common Roots, in Context. (Excerpts from the PATB Part 1 v 4.1 (Maamouri et al., 2010))

task for morphologically complex language such as Arabic are available, and possibly even more rare are data sets and corpora suitable for work on Arabic sentiment classification tasks.

Pang et al. (2002) tackled the classic problem of positive and negative two-class sentiment classification of English movie reviews from the Internet Movie Database (IMDB) corpus, highlighting the effectiveness of machine-learning techniques for sentiment classification, and paving the way for further research to enhance the efficiency of such automatic classification systems. With respect to the varied levels of granularity used (term, phrase, sentence, and document) in the classification task, individual and process-oriented approaches have been addressed, where information acquired from one level of analysis can be passed on to the next level, as observed by Turney and Littman (2003) and Dave et al. (2003). At the sentence-level, the work of Kim and Hovy (2004) addresses the topic of detecting sentiment towards a specific topic.

For feature selection and optimization, Yu et Hatzivassiloglou (2003) use N-gram based features and a polarity lexicon at the sentence level to determine subjectivity of sentences on the Wall Street Journal (WSJ) corpus, while Bruce and Wiebe (1999) use the same corpus, but employ additional lexical, part-of-speech (POS), and structural features. As a more profound paradigm shift, recent research has shifted from keyword and lexical-based approaches to concept-based sentiment analysis approaches, where semantic networks and entity ontologies are employed to achieve a more semantically-oriented "understanding" of text (Cambria et al., 2013; Grassi et al., 2011; Olsher, 2012).

With respect to the task of Arabic subjectivity and sentiment analysis in specific, the work of Abdul-Mageed et al. (2011) addresses the task in Modern Standard Arabic (MSA), where a manually-annotated corpus of MSA is presented from Part 1 v 3.0 of the Penn Arabic Treebank (PATB) (Maamouri et al., 2004), in addition to a wide-scale polarity lexicon tailored to the newswire domain under analysis. By using various stemming and lemmatization settings with a rich feature-set under an SVM classifier, it is shown that taking language-specific morphological features and traits into consideration for complex languages such as Arabic results in significant improvements in performance, achieving test results of 71.54% F (16.44% higher than the baseline) for subjectivity detection, and 95.52% F (37.14% higher than the baseline) for the sentiment analysis task using a newswire domain-specific lexicon, as compared to 57.84% F in development without the lexicon.

In the domain of positive and negative movie reviews, Rushdi-Saleh et al. (2011) present the Opinion Corpus for Arabic (OCA) movie review corpus, compiled from various Arabic web pages. Classification was performed using both Naive-Bayes and SVM classifiers, with combinations of N-grams, stemming, and stop-word removal preprocessing, and achieved a best result of 90% accuracy under SVM, as compared to a similar classification task in English with the Pang et al. (2002) IMDB corpus, which obtained 85.35% accuracy with various N-gram models.

Abbassi et al. (2008) explore the task of feature selection for the opinion classification task, using an Entropy Weighted Genetic Algorithm (EWGA), which incorporates both syntactic and stylistic features and the information gain heuristic to classify text on the document level. An accuracy of 93.6% is reported on a compiled Middle-Eastern web forum dataset. Other problems of sentiment analysis in informal and dialectical Arabic are also addressed by Albraheem and Al-Khalifa (2012) and Shoukry and Rafea (2012), for a more specific approach to the classification problem, tailored to a regional, social-media setting.

As compared to the discussed work on the sentiment analysis task in Arabic, the proposed root-based technique employes the commonalities between words of the same root to map sets of words to the same base meaning. Rather than taking a domain-specific approach to the problem, the proposed technique is tested on two corpora from very different domains: the PATB newswire corpus annotated by Abdul-mageed et al. (2011), and the OCA movie review corpus by Rushdi-Saled et al. (2011), with focus on language characteristics to enhance classification results.

4 Proposed Algorithm

The following section presents a detailed description of the datasets and rooting libraries used for experimentation, the various experimentation settings undergone, and the proposed classification method applied for the task sentiment analysis on the two studied domains.

4.1 Datasets

The proposed sentiment classification method was conducted on two different datasets, to test the root-based approach on a generic level, unconstrained by the domain of the data itself.

4.1.1 Penn Arabic Treebank (PATB Part 1 v 4.1) Newswire Corpus

The first corpus used pertains to the tokenized newswire-domain text from the latest version of the PATB, Part 1 v 4.1 (Maamouri et al., 2010). The corpus consists of 734 newswire stories from the Agence France Presse (AFP) with various tags attached to each token, including part-of-speech information, morphology, English gloss, treebank annotation, and vocalization.

For the purposes of the sentiment analysis task, the applied section of the dataset comes from the compiled corpus of Abdul-Mageed et al. (2011), where the first 2855 sentences (comprising 54.5% of the Part 1 v 3.0 dataset[1], in 400 documents) were each manually annotated into one of four labels (Abdul-Mageed and Diab, 2011): objective (OBJ), subjective positive (S-POS), subjective negative (S-NEG), and subjective neutral (S-NEUT), depending on whether the information being conveyed in the sentence was to objectively inform, or offer a subjective sense (Wiebe et al., 1999). The number of sentences with each of the four respective tags are shown in Table 3.

Tag Class	Number of Sentences
OBJ	1281
S-POS	491
S-NEG	689
S-NEUT	394

Table 3: Distribution of Sentiment Classes in the Manually-Tagged Portion of the PATB Corpus (Abdul-Mageed et al., 2011).

4.1.2 Opinion Corpus for Arabic (OCA) Movie Review Corpus

For another perspective for testing the proposed root-based method, the distinctly subjective OCA corpus (Rushdi-Saleh et al., 2011) was also experimented with. The corpus is comprised of 500 movie reviews (250 positive and 250 negative) which were collected from 15 Arabic web pages, after which a series of spelling correction, tokenization, basic stop-word and special character removal, and stemming processes were performed. In addition, normalization of the rating schemes used for each site was conducted to appropriately partition the reviews into positive and negative categories, and prepare the text for the classification task (Rushdi-Saleh et al., 2011).

[1]The differences between the two versions of the PATB Part 1 lie in improvements to the organization of the data, and updates to certain aspects of the annotation (Maamouri et al., 2010).

4.2 Rooting Libraries

Three different rooting libraries were applied to derive the roots of each of the input words in the classification examples, each applying a slightly different approach to the complex Arabic rooting problem.

4.2.1 Khoja Arabic Stemmer

The Khoja Arabic Stemmer (Khoja and Garside, 1999) is a fast Arabic stemmer that works by removing the longest prefix and suffix present in the input word and then matching the rest of the word with known verb and noun patterns using a root library. The stemmer attempts to take into account the unavoidable irregularities in the language in order to extract the correct root from words that do not follow the general rules. The Khoja stemmer has been used in various Arabic natural language processing tasks, and has been noted to produce good improvements to various natural language tasks, despite many tagging errors (Larkey and Connell, 2001).

4.2.2 Information Science Research Institute (ISRI) Arabic Stemmer

The Information Science Research Institute's (ISRI) stemmer (Taghva et al., 2005) uses a similar approach to word rooting as the Khoja stemmer, but does *not* employ a root dictionary for lookup. Additionally, if a word cannot be rooted, the ISRI stemmer normalizes the word and returns a normalized form (for example, removing certain determinants and end patterns) instead of leaving the word unchanged. The ISRI stemmer has been shown to give good improvements to language tasks such as document clustering, as opposed to a non-stemmed approach (Bsoul and Mohd, 2011).

4.2.3 Tashaphyne Light Arabic Stemmer

The Tashaphyne Light Arabic Stemmer (Tashaphyne, 2010) works by first normalizing words in preparation for the "search and index" tasks required for stemming, including removing diacritics and elongation from input words. Next, segmentation and stemming of the input is performed using a default Arabic affix lookup list, allowing for various levels of stemming and rooting (Tashaphyne, 2010).

4.3 Experimental Setup

Two different sets of experiments were conducted to test the effect of the root-based method for the sentiment analysis task, varying somewhat for the different corpora under analysis.

For the PATB corpus, only the subjective data was taken into consideration for the two-class positive and negative sentiment classification problem (the 1180 sentences in total: 491 S-POS and 689 S-NEG sentences). The task was conducted at the sentence-level, with 5-fold cross-validation splits across the dataset (Abdul-Mageed et al., 2011).

For the OCA corpus, with already-defined opinions in the form of movie reviews, the entire dataset was used for classification. The task was conducted at the document-level (with each document composed of sets of sentences, ranging from an average of 13 sentences in the positive review sets, and 20 sentences in the negative reviews), with 10-fold cross-validation splits (Rushdi-Saleh et al., 2011).

Each of the corpora was tested using the basic words as features to the classifier, then by iteratively adding roots using each of the three rooting libraries. After experimentation with parameters, the classification task was performed using a linear kernel (Abdul-Mageed et al., 2011) under the SVM^{light} classifier (Joachims, 2008). Precision and recall values are reported for the average of the K-fold runs (5 folds for the first corpus, and 10 folds for the second), along with F-measure (F1) and accuracy results for each respective experiment.

4.4 Pre-processing

As pre-processing to prepare the text for the classification and analysis tasks, the already undiacritized corpus sentences were tokenized from the set of documents into word sets, and testing with stop-word removal (the removal of commonly used words) was done to filter out words that could be unnecessary for the task of opinion classification.

4.5 Feature Sets

For each document, the basic unigrams (individual words) composing the document were used as initial input features to the SVM, after which the three rooting libraries (Khoja, ISRI, and Tashaphyne) were then iteratively used to derive the roots of each of the input word features. Finally, the resultant roots were then added as additional features (along with the unigrams) to each of the examples. Binary presence vectors were used to indicate the existence of a feature (Abdul-Mageed et al., 2011). For the purposes of exploring the effect of adding the root-based features, only independent unigrams and unigrams with roots were experimented with for the basic evaluation task.

5 Results and Evaluation

The results of the sentiment analysis tasks on the two datasets are illustrated in Table 4 for the PATB corpus, and Table 5 for the OCA corpus, detailing the task statistics. The basic results using a standard *unigram* feature (the encountered word itself) are depicted initially, along with a *baseline* result (in F-measure and accuracy, as available for the two datasets, respectively) without the use of root features, as presented by previous work with the datasets (Abdul-Mageed et al., 2011; Rushdi-Saleh et al., 2011).

With respect to the sentiment analysis task on the PATB newswire-domain corpus shown in Table 4, all three of the individual rooting libraries resulted in improvements to the initial unigram results. The largest observable improvement to all measures reported came from the Khoja stemmer, with a 4.9% increase in F-measure, and a 4.7% increase in accuracy as compared to the unigram result. Also, a 3.4% increase in F-measure is observed from the sentiment baseline of 57.8% (previously achieved on the dataset using various morphological features, without a domain-specific lexicon (Abdul-Mageed et al., 2011)).

For the OCA movie-domain corpus shown in Table 5, slight improvements can be seen by adding root features to the unigram classifier input, particularly with the Tashaphyne rooting library. An increase of 3.2% accuracy after the addition of root features is observed from the baseline accuracy of 90.0% (Rushdi-Saleh et al., 2011).

While an increase in overall accuracy and F-measure is notable in the task of basic two-class opinion classification, two main points are of im-

	Precision	Recall	F1	Accuracy
Unigrams (No Roots)	58.1	58.1	56.3	63.8
+ Khoja Roots	**63.8**	**61.9**	**61.2**	**68.5**
+ ISRI Roots	61.5	60.4	59.3	66.8
+ Tashaphyne Roots	61.7	58.8	58.8	67.0
Baseline			57.8	

Table 4: PATB Sentiment Classification Results for the Proposed Method under Three Rooting Libraries. (Baseline: Abdul-Mageed et al. (2011))

	Precision	Recall	F1	Accuracy
Unigrams (No Roots)	90.0	95.2	92.8	92.6
+ Khoja Roots	90.7	94.4	92.3	92.2
+ ISRI Roots	90.5	95.6	92.8	92.6
+ Tashaphyne Roots	**91.1**	**96.0**	**93.4**	**93.2**
Baseline				90.0

Table 5: OCA Sentiment Classification Results for the Proposed Method under Three Rooting Libraries. (Baseline: Rushdi-Saled et al. (2011))

portance: the nature of the words in the dataset under analysis, and the efficiency of the stemming systems themselves. The divergence in the most effective rooting library on each of the corpora can be attributed to various factors, including the style of writing used in the datasets, the formality of the text, and the existence of irregular words and words that cannot be rooted, depending on the accuracy and robustness of the employed stemming library.

The PATB news domain corpus, for example, is expected to have less opinion-bearing content than the OCA movie review corpus, due to the less subjective nature of the domain. The overall accuracy and F-measure results for the OCA movie corpus are thus significantly higher than those observed in the PATB corpus. Another difference between the corpora lies in the formality of the language employed: while the PATB corpus uses a strict Modern Standard Arabic (MSA), the use of slang and dialect-specific language is frequent in the OCA corpus. This type of varied language presents a layer of difficulty for sentiment classification in general, as well as for the rooting systems applied for language mapping.

Furthermore, with respect to the stemming tools themselves, the overall inaccuracy of current stemmers is another important consideration. The best-performing stemming libraries, Khoja and Tashaphyne, for each of the two domains, are those that employ some form of root-lookup dictionary in order to verify the correctness of the affixes and resultant roots generated. Another consideration is the limitation imposed by employing only unigrams enriched with the root features, while features beyond word level could be used to further predict sentiment patterns changes over a more complex language structure.

6 Conclusion

The composition scheme and complex morphology of Arabic make the task of root-extraction to normalize words to their basic functional units a very relevant one for various natural language processing tasks. With respect to the sentiment analysis tasks presented in this paper, some notable improvements to the classification performance when using various rooting libraries as input features can be observed, warranting further research on enhancements to existent rooting techniques and handling of the intricacies of the Arabic language structure to predict more sentence forms and correctly classify their polarity.

As detailed, the reasoning behind the root-based method and its enhancement of the sentiment classification task in the two explored domains relies on the semantic similarities between different word derivations, allowing for a broader map of interconnections between words with similar polarity orientation to be created. Such valid interconnections between words also warrants the exploration of semantic expansion of words and their synonyms (Magdy et al., 2013), expanding the word map and serving to better connect and understand sentiment-bearing ideas and expressed opinions.

By applying various rooting schemes at different granularities in two separate domains, it is also shown that word roots can serve to enhance the sentiment analysis task results on a more generic level, instead of using a domain-specific approach that may not always be applicable. Thus, using root derivation techniques such as that presented for Arabic sentiment analysis in particular are applicable and valid to help enhance the performance of natural language processing tasks on morphologically rich and complex languages.

References

A. Abbasi, H. Chen, and A. Salem. 2008. Sentiment analysis in multiple languages: feature selection for opinion classification in web forums. *ACM Trans. Inf. Syst.*, 26(3):12:1–12:34, June.

M. Abdul-Mageed and M. Diab. 2011. Subjectivity and sentiment annotation of modern standard arabic newswire. In *Proceedings of the 5th Linguistic Annotation Workshop*, LAW V '11, pages 110–118, Stroudsburg, PA, USA. Association for Computational Linguistics.

M. Abdul-Mageed, M. Diab, and M. Korayem. 2011. Subjectivity and sentiment analysis of modern standard arabic. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2*, HLT '11, pages 587–591, Stroudsburg, PA, USA. Association for Computational Linguistics.

L. Albraheem and H. Al-Khalifa. 2012. Exploring the problems of sentiment analysis in informal arabic. In *Proceedings of the 14th International Conference on Information Integration and Web-based Applications; Services*, IIWAS '12, pages 415–418, New York, NY, USA. ACM.

R. Bruce and J. Wiebe. 1999. Recognizing subjectivity: a case study in manual tagging. *Nat. Lang. Eng.*, 5(2):187–205, June.

Q. Bsoul and M. Mohd. 2011. Effect of isri stemming on similarity measure for arabic document clustering. In *Information Retrieval Technology*, volume 7097 of *Lecture Notes in Computer Science*, pages 584–593. Springer Berlin Heidelberg.

E. Cambria, B. Schuller, Y. Xia, and C. Havasi. 2013. New avenues in opinion mining and sentiment analysis. *Intelligent Systems, IEEE*, 28(2):15–21.

K. Dave, S. Lawrence, and D. Pennock. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, WWW '03, pages 519–528, New York, NY, USA. ACM.

M. Grassi, E. Cambria, A. Hussain, and F. Piazza. 2011. Sentic web: A new paradigm for managing social media affective information. *Cognitive Computation*, 3(3):480–489.

N. Habash. 2008. *Introduction to Arabic Natural Language Processing*. Synthesis lectures on human language technologies. Morgan & Claypool Publishers.

T. Joachims. 2008. Svmlight: Support vector machine, http://svmlight.joachims.org/.

S. Khoja and R. Garside. 1999. Stemming arabic text (tech. rep.). Computer Department, Lancaster University, Lancaster.

S. Kim and E. Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

L. Larkey and M. Connell. 2001. Arabic information retrieval at umass in trec-10. University of Massachusetts.

B. Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

M. Maamouri, A. Bies, and W. Mekki. 2004. The penn arabic treebank: building a large-scale annotated arabic corpus. pages 102–109.

M. Maamouri, A. Bies, S. Kulick, F. Gaddeche, W. Mekki, S. Krouna, B. Bouziri, and W. Zaghouani. 2010. Arabic treebank: Part 1 v 4.1.

A. Magdy, M. Kholief, and Y. El-Sonbaty. 2013. Qasim: Arabic language question answer selection in machines. *Conference and Labs of the Evaluation Forum.*

D. Olsher. 2012. Full spectrum opinion mining: integrating domain, syntactic and lexical knowledge. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*, pages 693–700.

B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on empirical methods in natural language processing - Volume 10*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

M. Rushdi-Saleh, M. Martin-Valdivia, L. Urena-Lopez, and J. Perea-Ortega. 2011. Oca: Opinion corpus for arabic. *J. Am. Soc. Inf. Sci. Technol.*, 62(10):2045–2054, October.

K. Ryding. 2006. *A Reference Grammar of Modern Standard Arabic*. Reference Grammars. Cambridge University Press, New York.

D. Sanjiv and M. Chen. 2001. Yahoo! for amazon: extraction market sentiment from stock message boards. In *Proceedings of the 8th Asia Pacific Finance Association Annual Conference*.

A. Shoukry and A. Rafea. 2012. Sentence-level arabic sentiment analysis. In *Collaboration Technologies and Systems (CTS), 2012 International Conference on*, pages 546–550.

StudyQuran. 2004. Project root list online. 2004. http://www.studyquran.co.uk/prlonline.htm.

K. Taghva, R. Elkhoury, and J. Coombs. 2005. Arabic stemming without a root dictionary. In *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume I - Volume 01*, ITCC '05, pages 152–157, Washington, DC, USA. IEEE Computer Society.

Tashaphyne. 2010. Arabic light stemmer, 0.2. 2010. http://tashaphyne.sourceforge.net/.

P. Turney and M. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Trans. Inf. Syst.*, 21(4):315–346, October.

J. Wiebe, R. Bruce, and T. O'Hara. 1999. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, pages 246–253, Stroudsburg, PA, USA. Association for Computational Linguistics.

H. Yu and V. Hatzivassiloglou. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, EMNLP '03, pages 129–136, Stroudsburg, PA, USA. Association for Computational Linguistics.