SENTIMENT ANALYSIS FOR ARABIC SENTENCES USING NLP AND CLASSIFICATION TECHNIQUES

Inas Mohammed Milad[1] **Abdulrazag Mukhtar Atomi[2]**

Email: inasmilad2007@gmail.com **Email:** abdulrazag.atomi@gmail.com
Higher Institute of Science and Technology - Al-khomes[1]
Higher Institute of Science and Technology – Messallata[2]

Abstract

Sentiment analysis is one of most important domains of Natural Processing Language (NLP). The main aim of sentiment analysis is to summarize large text as simple opinions taxonomy in order to simplify extractions of people opinions about specific subject. Arabic sentiment analysis still faces many challenges due to complexity of Arabic language such as morphology, orthographic, and widespread of synonyms. The main aim of this research is to develop sentiment analysis system for Arabic language in order to analyze Arabic sentences that typed by Arabic people via twitter application as positive/neutral/negative opinions. The objectives of this research are: to analyze the Arabic texts using effective methods of sentiment detection and features extraction, and to classify the analysed Arabic texts as positive/neutral/negative opinions using effective classification method. There are many pre-processing and NLP methods adopted to address the first objective of this research. Tokenization, part of speech, and stemming methods are applied to detect the features in sentences, N-gram are conducted to extract the features, and Arabic WordNet tool are used to enrich the extracted features. In order to address the second objective of this research, Support Vector Machine (SVM) method is conducted to classify the extracted features as three polarities; positive, neutral, or negative. The proposed system of sentiment analysis was tested on dataset called SYR. This dataset consists of 2000 Tweets in Arabic language about the wars in Syria. The author uses 80% of tweets (first 1600 tweets) as training set while the rest 400 tweets (20%) are used as testing set for the classification system. The results of experimental tests show that the proposed system records 0.78 as total accuracy results of sentiment classification. The accuracy of proposed system consider as effective results comparing with the available systems of Arabic sentiment analysis for social dataset. The main contribution of this research is the development of sentiment analysis system for Arabic tweets with effective accuracy of sentiment classification. This indicates the effectiveness of the proposed methods to address the challenges of sentiment analysis of Arabic language.

Keywords: Sentiment analysis, Arabic language, Pre-processing, NLP, SVM

1.0 Introduction

Sentiment analysis is one of most important implications of Natural Processing Language (NLP) (Gebremeskel, 2011). Sentiment analysis can be defined as the texts summarization and predefining according to simple opinions taxonomy (Pang & Lee, 2008). There are many opinions taxonomies can be adopted to conduct the sentiment analysis such as; (1) positive/negative, (2) relevant/irrelevant, (3) agree/disagree, and (4) ranking from 1-100. Korayem et al. (2012) mentioned that taxonomy selection and classes could be identified depend on the implementation environment of sentiment analysis.

Wilson et al. (2005) argued that the sentiment analysis is usually obtained on various textual sources at different levels like documents, sentences, and word/feature. However, the core benefit of sentiment analysis is appeared when it obtained on large size of documents or large number of text sources. Nowadays, there has been an increasing interest in harvesting and analyzing the information written in from internet sources like online forums, twitter and Facebook (Refaee&Rieser, 2014). Social network allows the users from any country to type their opinions as textual paragraphs or sentences in real time without any constraints. This gives the organizations and companies the opportunity to analyze real users' opinions at the right time. On the other hand, the numbers of users who involve the social networks are large which represent reliable community of data collection and analysis.

The main aim of sentiment analysis is to simplify the understanding of texts directions at the right time and in less effort. Thus, there are many benefits could be gained from the sentiment analysis implementations in various domains such as business, Political, and social (Wright, 2009)

According to Siddiqui (2015) and World Stats (2013), there are about 65 million Arabic-speaking users online, or about 18.8% of the global Internet population. However, most of the systems built for sentiment analysis are tailored for the English and European languages. The sentiment analyses of Arabic text that gathered via social network is important due to many reasons such as the original Islamic sources are written in Arabic language, the Arabic region is hot political area, and there many revolutions were appeared in Arabic countries (Refaee&Rieser, 2014; Zaidan&Callison-Burch, 2013; Al-Sabbagh&Girju, 2012).

However, sentiment analysis of Arabic language still faces challenges due to many reasons such as the following (Siddiqui, 2015; Al-Marghilani et al.,2007; Al-Marghilani et a.,2008; El-Halees, 2008; Ghwanmeh, 2008):

- Orthographic with diacritics is less ambiguous and more phonetic in Arabic, certain combinations of characters can be written in different ways.
- Arabic has a very complex morphology recording as compare to English language. Arabic being derivational involves the derivation of verbs from two or three root words. Likewise, all adjectives and nouns are derivational as well.
- Arabic synonyms are widespread which increases the difficulty of identify the true meaning of Arabic concepts.

Hence, the main aim of this research is to develop sentiment analysis system in order to enhance the accuracy of analyze Arabic text that provided by Arabic people as positive or negative opinions.

2.0 Literature Review

This section presents the aspects and levels of sentiment analysis in addition to related works of sentiment analysis for Arabic and English languages. Thus, this section will gives between understanding of sentiment analysis directions and methods.

2.1 Aspects of Sentiment Analysis

According to Pang and Lee (2008), and Aldayel and Azmi (2015), the systemic processes of SA are relied on two main phases which are:

(i) Sentiment detection and features extraction: in this phase the text processed to analyze whether it includes objective or subjective contents and extract the main sentiment features in the text (Pang & Lee, 2008; Aldayel&Azmi, 2015). Mainly, the subjective contents in any text indicates that this text represent sentiment or opinion. For example, the text “جداجيدةجيسالسيارة” ("CIAZ car is very good") is considered as subjective. The subjectivity detection can be conducted in different levels of text such as sentence and documents levels. The text considered as objective if there no feelings, attitudes, sentiment are detected in the text. For example, the text “2015 السيارة جيس تم طرحها في السوق في يناير ” ("CIAZ car was launched in January 2015") isconsidered as objective text.

(ii) Sentiment classification: Once the text is classified as subjective, the classification phase can be conducted to classify the text based on its polarity. The most known classification in this phase is the taxonomy of positive, negative, or neutral (Pang & Lee, 2008; Aldayel&Azmi, 2015). The text considered as positive sentiment if its total features indicate positive opinion. The text considered as negative sentiment if its total features indicate negative opinion. The mixed sentiment text is include negative and positive features. The text considered as neutral sentiment if this text not include positive or negative features.

The text which is classified as Subjective is further worked on to classify as positive, negative, or neutral. Table 1 presents some examples of the sentiment classification of sentences.

Table 1: Examples of Sentiment Classification

sentence in Arabic	Sentence in English	Sentiment Classification
"الطعام في مطعم سيزارس لذيذ جدا"	CESSARS restaurant food is very delicious	Positive
والدتي هي أفضل طاهية	My mother is the best cook	Positive
التعليم في مدرسة النجاح سيء جدا	The education in Al-Nagah School is very poor	Negative
"يا " !و انا سأصبح مجنون مع هذا الغش"	Oh! I am going to be mad with this cheating	Negative
علي الرغم من أن الكلية توفر تعليم ممتاز لكن الأنشطة المتعلقة بالمناهج الإضافية جعلت بعض الطلاب يحيدون عن الدراسة	Though the college provides excellent education but the extra-curricular activities led some students to deviate from studies	Mixed (Positive and Negative)
قد تغلق المدرسة	The school may shut down	Neutral

2.2 Sentiment Analysis Classification levels

Sentiment analysis operates at different classification level which includes aspect, sentence or document level.

- Document level: The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment (Pang, Lee and Vaithyanathan, 2002; Turney, 2002). For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product. This task is commonly known as document-level sentiment classification. This level of analysis assumes that each document expresses opinions on a single entity (e.g., a single product). Thus, it is not applicable to documents which evaluate or compare multiple entities.

- Sentence level: The task at this level goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion. This level of analysis is closely related to subjectivity classification (Wiebe, Bruce and O'Hara, 1999), which distinguishes sentences (called objective sentences) that express factual information from sentences (called subjective sentences) that express subjective views and opinions. However, we should note that subjectivity is not equivalent to sentiment as many objective sentences can imply opinions, e.g., "We bought the car last month and the windshield wiper has fallen off." Researchers have also analyzed clauses (Wilson et al., 2004), but the clause level is still not enough, e.g., "Apple is doing very well in this lousy economy."

- Feature and Aspect level: Both the document level and the sentence level analyses do not discover what exactly people liked and did not like. Aspect level performs finer-grained analysis. Aspect level was earlier called feature level (feature-based opinion mining and summarization) (Hu and Liu, 2004). Instead of looking at language constructs (documents, paragraphs, sentences, clauses or phrases), aspect level directly looks at the opinion itself. It is based on the idea that an opinion consists of a sentiment (positive or negative) and a target (opinion).

2.3 Related Works of Sentiment Analysis for English

Opinion detection has mostly been performed on the document level (Wiebe et al., 1999; Wiebe, 2000). Another focus that can be discerned is in terms of genre, on movie and product reviews (Dave et al., 2003; Hu and Liu, 2004; Turney, 2002). However, there are a number of sentence- and phrase-level classifiers (Wiebe et al., 1999; Morinaga et al., 2002; Yu and Hatzivassiloglou, 2003; Kim and Hovy, 2004). Yu and Hatzivassiloglou (2003) propose a three-stage approach that performs subjectivity analysis first on the document level, then on the sentence level. In a final step, they classify the sentences into positive, negative, or neutral opinions. If sentiment analysis is performed on the sentence level, it is generally with regard to a target concept (Morinaga et al., 2002; Kim and Hovy, 2004), i.e., the system has asits goal to identify sentiment towards a concept such as "cellular phone" or one of its attributes such as "size". Standard approaches to subjectivity include rule-based approaches (Morinaga et al., 2002), supervised classifiers such as Naive Bayes (Yu and Hatzivassiloglou, 2003) or statistical approaches using linguistic features as well as metadata (Dave et al., 2003). Sentiment is often determined based on lexical resources such as wordnet (Dave et al., 2003; Yi et al., 2003; Kim and Hovy, 2004), sentiment lexicons (Yi et al., 2003), or bootstrapped word lists based on seeds (Kim and Hovy, 2004). Since sentiment is often determined for given target concept, a subtask of extracting opinion targets has been investigated separately using support vector machines (Kessler and Nicolov, 2009) or conditional random fields (Jakob and Gurevych, 2010). More recently, the focus on movie and product reviews is becoming less prominent, with work on blogs and twitter data (Davidov et al., 2010) becoming available. Since tweets are rather short, often additional information such as twitter hashtags and smileys (Davidov et al., 2010) or from label propagation are used. More recent developments with regard to machine learning techniques use graphbased methods, which allow a more global view of the problem as well as joint inference with sentence cohesion, agreement between speakers, or discourse relations (Somasundaran et al., 2009).

As Twitter becomes more popular, sentiment analysis on Twitter data becomes more attractive. (Go et al., 2009; Barbosa and Feng, 2010; Davidiv et al., 2010) all follow the machine learning based approach for sentiment classification of tweets. Specifically, (Davidiv et al., 2010) propose to classify tweets into multiple sentiment types using hashtags and smileys as labels. In their approach, a supervised KNN-like classifier is used. Go et al., (2009) used NB and SVM methods to classify the emotions sentiment of Twitter messages as positive or negative emotions. The significant results of Go et al. Show that the SVM (81.9%) is more accurate than NB (79.9%) when using Unigram and POS as preprocessing methods.While the NB (82.7%) is more accurate than SVM (81.6%) when using bigrampreprocessing method. In contrast, (Barbosa and Feng, 2010) propose a two-step approach to classify the sentiments of tweets using SVM classifiers with abstract features. The training data is collected from the outputs of three existing Twitter sentiment classification web sites. Table 2.1 summarizes the related works to sentiments of tweets. Table 2 summarizes the related works of sentiment analysis for English Tweets.

Table 2: Summary of Related Works of Sentiment Analysis for English Tweets

Author	Research Aim	Methods	Findings
Davidiv et al. (2010)	classify tweets into multiple sentiment types using hashtags and smileys as labels	KNN	The accuracy of classifying is about 82%
Barbosa and Feng (2010)	classify the sentiments of tweets with abstract features	SVM	The accuracy of classifying is about 81%.
Go et al. (2009)	classify the emotions sentiment of Twitter messages as positive or negative emotions	SVM and NB suppoting uni and bigram.	The besr accuracy of SVM is with unigram (81.9%); The best accuracy of NB is with bigram (82.7).

2.5 Related Works of Sentiment analysis for Arabic

Siddiqui (2015) mentioned that there are clear limitation in the works of sentiment analysis for Arabiclanguage and the most known developed system for Arabic sentiment analysis is called SAMAR which developed by Abdul-Mageed et al.(2013). Since 2016, only three known studies have been founded for sentiment analysis based on Arabic language. All of these works were conducted the sentiment analysis at sentences level and classify the subjectivity as positive or negative opinions.

Abbasi et al. (2008) perform a sentiment analysis of English and Arabic Web forums, with an overarching goal of identifying hostility in computer-mediated communication. They make use of not only syntactic but also stylistic features. Syntactic features include word, ngrams, POS tag, and word roots. Stylistic features include character n-grams among other types of information. Thus, Abbasi et al. deal with the morphological richness of Arabic by indirect means in the form a n-grams as well as by reducing words to their roots. Abbasi et al. use an entropy-weighted genetic algorithm (EWGA) as a feature selection technique on (1) an English benchmark movie review database taken from the IMDb movie review archive (http://www.imdb.com) and (2) a testbed of messages from two major extremist forums (a U.S. American one in English and a Middle Eastern one in Arabic). Their EWGA uses information gain as a heuristic to weight the various sentiment attributes. Abbasi et al. find that stylistic features on their own were outperformed by syntactic features(i.e., word n-grams, punctuation, word roots), but when triangulated with syntactic features, stylistic features helped gain higher classification accuracy of approximately 5%. A number of stylistic features were found to be specifically helpful, including the total number of characters, use of digits and emphasizing symbols, and vocabulary richness. The root extraction is handled by a clustering algorithm, which compares words against a list of roots. The number of roots is manually set to 50. From the publication, it is unclear how well this algorithm performs on identifying roots, especially since only the most frequent roots are recognized.

Another work is performed by Abdul-Mageed et al. (2011) on building an SSA system that exploits newswire data from the Penn Arabic Treebank . The proposed system uses the gold-labeled morphological features and a polarity lexicon from the news domain. This system reaches an F-score of 71.54 for subjectivity detection and 95.52 for sentiment classification.

Furthermore, Abdul-Mageed et al. (2013) developed a system called SAMAR for sentiment analysis based on Arabic language. The proposed system was developed based on two main methods which POS in order to detect the sentiment of are given sentences and SVM for subjectivity classification. The proposed system was tested on four data sets of web forums such as Wikipedia and tweeter. For sentiment detection the system record 90.81% as best result and for sentiment classification the system record 78.65% as best result. Table 3 summarizes the related works of sentiment analysis for Arabic.

Table 3: Summary of Related Works of Sentiment Analysis for Arabic

Author	Research Aim	Methods	Dataset	Findings
Abbasi et al. (2008)	perform a sentiment detection of Arabic Sentences	ngrams, POS tag, and word roots	Arabic Web forums	The accuracy of sentiment detection is 88%
Abdul-Mageed et al. (2011)	building an SSA system for Arabic sentiment analysis	gold-labeled morphological features and a polarity lexicon	news	71.54 % for subjectivity detection. 95.52% for sentiment classification.
Abdul-Mageed et al. (2013)	developed a system called SAMAR for sentiment analysis based on Arabic language	POS and SVM	four data sets of web forums such as Wikipedia	For sentiment detection the system record 90.81% as best result. for sentiment classification the system record 78.65%

3 Research Methods

The methods of proposed sentiment analysis system can be categorized as two main phases which are (1) methods for sentiment detection and extraction, and (2) methods for sentiment classification. As figure 1 illustrates, the proposed methods of this research can be categorized as three stages which are preprocessing stage for sentiment detection, stage of features extraction, and classification stage.

In the preprocessing phase, there are many methods are used to analyze the main features of Arabic sentences; tokenization and stemming. Furthermore, POS is used to identify the features based on the sentence pattern. In feature extraction phase, N-gram is used to list the extract the identified features as 1-gram, 2-gram... n-gram. The Arabic WordNet tool is applied to enrich the extracted features by find more features or synonyms of the same meaning. In the feature classification phase, SVM method is adopted to classify the sentence according as positive, neutral, or negative based on the related features to this sentence. Lastly, the classification accuracy could be measured using F-Score formulas through compare the system classification results with the true sentiment of each tweeter sentence or message.

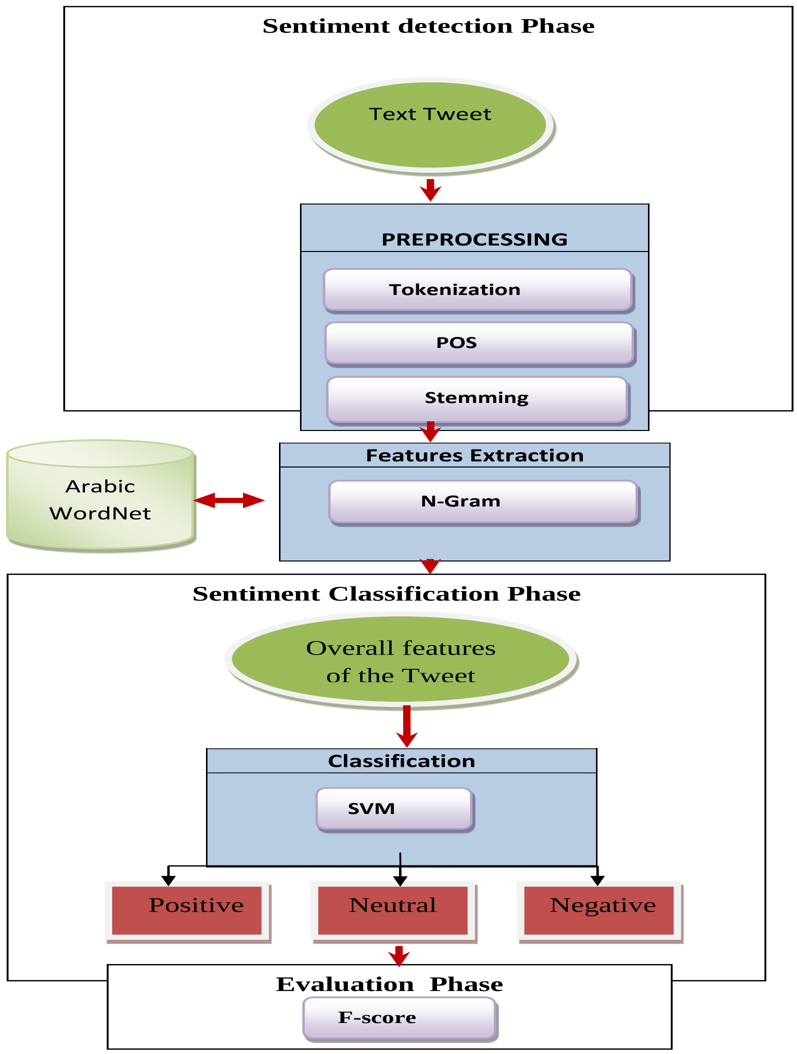

Figure 1: Research Design

The selected methods take in consideration many points based on the challenges that face the sentiment analysis for Arabic sentences, these points are as the following:

i. **Preprocessing methods for sentiment detection**: there are many preprocessing methods in this stages: (1) **Tokenization:** pre-processing of Arabic data set in terms of performing tokenization brings in new challenges resulting in satisfying different shapes of letters found at the middle, end, and beginning (i.e. changes in shape of a letter). (2) **Stemming:** the stemming or word root is effective methods to address the morphology challenge of Arabic words. (3) **Part of Speech (POS)**: POS is effective method to address the Orthographic and lack of capitalization challenges through identifies the concepts entity based on the sentences pattern.

ii. **Natural processing Language (NPL) methods for features extraction:** N-gram method is effective method to address the morphology challenge of Arabic words through extract the various meaning or synonyms of give word or words. This requires connection with efficient lexicon such as WordNet. Arabic WordNet (AWN) is effective lexicon to extract the various meaning or synonyms of words that listed using N-gram.

iii. **Sentiment classification**: The sentiment classification could be conducted effectively using Support Vector Machine (SVM). For sentiment analysisif texts, the SVM considered as effective method for sentiment classification.

The Dataset in this research is collected from Salameh et al., (2015) work. The collected dataset is Syria Dataset (SYR) which consists of 2000 Tweets in Arabic language about the wars in Syria. Salameh et al., (2015) were created SYR in 2014 based on the tweets that gathered between Arabic people about Syrian wars. Some of these tweets were originally written by users using Arabic language, while the other tweets were written in English language before translated to Arabic. Salameh et al., (2015) were classifying the sentiment of tweets in SYR manually as Table 4 presents. The SYR was downloaded from the authors' formal website (http://saifmohammad.com/WebPages/ArabicSA.html).

Table4 Manual classification of SYR Tweets

Polarity	**Number of Tweets**
Positive Tweets	1350
Negative Tweets	448
Neutral Tweets	202
Total	2000

3.1 Preprocessing Methods

The pre-processing phase consists of many related methods that integrated together in order to identify the main features of tweets. These methods are as the following:

A. Tokenization

Tokenization is the process that divides sentences into the text sequence and then the text sequence into single tokens. This step considered as the first pre-processing steps to identify features as single tokens. The given query by users is segmented as single words. Usually, the word segmentation is done though identify the whitespace before and after any word. In addition, if there any special characters in the query like question mark or punctuation then it will be tokenized as single word. Suppose the following example in Table 5 to clarify the tokenization process.

Table 5: Tokenization Example

Before tokenization: شافي العجمي يتبراء من تنظيم داعش بعد تمويلهم من قبله بجمع التبرعات لقتل أطفال سوريا وشيوخها حسبي الله عليك يا شافي Shafi al-AjmirejectISIS after their funding from the collection of donations to kill children, Syria and the elderly, by God, you Shafi

After Tokenization:	
شافي العجمي يتبراء من تنظيم داعش بعد تمويلهم من قبله	بجمع التبرعات لقتل أطفال سوريا وشيوخها حسبي الله عليك يا شافي

B. POS

The results of tokenization process represent the input of POS process. This work adopts the library of "Stanford Log-linear Part-Of-Speech Tagger" in order to conduct the POS process. Stanford POS Tagger is a piece of software that reads text in some language like Arabic, and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'.

Basically, the Arabic POS conducted based on two types of tokens; (1) noun types, and (2) verb types. Any token is considered as 'NN' if this token is tagged as noun word such as illusive noun, noun of time, instrumental noun, or adverb. For example, in the tweet " شافي العجمي يتبراء من تنظيم داعش بعد تمويلهم من قبله بجمع التبرعات لقتل أطفال سوريا وشيوخها حسبي الله عليك يا شافي", there are many noun tokens such as "شافي", "اطفال", "سوريا" , "داعش" . all of these token are tagged as "NN" type.

On the other hand, any token is considered as 'VB' if this token is tagged as verb. The verb can be classified according to many rules such as; if it has vowels, passive or active, if it is merely or has extra letter and the number of letters, if verbs have special case. For example, in the tweet " شافي العجمي يتبراء من تنظيم داعش بعد تمويلهم من قبله بجمع التبرعات لقتل أطفال سوريا وشيوخها حسبي الله عليك يا شافي", there are many verb tokens such as "يتبراء", "لقتل", "بجمع".all of these token are tagged as "VB" type.

C. Stemming

After POS processes, the Stemming process is conducted. The stemming is the process of extract the main word root in order to identify the true meaning of feature. For example "تمويلهم" word is stemmed as "تمويل".

In stemming process, the meaning of features could be changed. For example, the word "روعه" which means "magnificent", and stem it using any traditional Arabic light stemmer, the result will be the term "روع", which means "terrorized". While the first term is very positive, the second is very negative. Therefore, the lemmatization process is conducted in this step to avoid the stemming of many features. Lemmatization process identifies the list of features that cannot be stemmed due to critical meaning of these features. The features lemmatization are listed in internal file, and this list was extracted from Dictionary Based Tool (DBT) that developed by El-Beltagy and Rafea (2016). DBT contains set of Arabic words or features that not recommended to be stemmed.

Furthermore, in this step, the some characters were normalization in order to avoid the challenge of changing character shape. In this step, letters "إ" , "أ" and "آ" are replaced with "ا" while the letter "ة" is replaced with "ه", and the letter "ى" is replaced with "ي".

3.2 Methods of Features Extraction

After pre-processing phase, the main features could be extracted and enriched using the following two methods:

A. N-gram

The N-gram algorithm focuses on listed the extracted features from each tweet as word by word, two words by two words and so on. The main aim of N-gram is to generate all possible features intersections to increase the possibility of analysis the tweet sentiment. In this research, the N-gram lists are generated according to number of identified features in tweet i.e. 8-gram. Table 6 presents example of N-Gram combination list based on five features in tweet:

Table 6: 5-Gram Example

الجيش الحر يستهدف حاجز السلام The free army targeted Al-Salam block				
1-Gram	2-Gram	3-Gram	4-Gram	5-Gram
A **الجيش** **الحر** **يستهدف** **حاجز** **السلام**	**الجيش الحر** **الحريستهدف** **يستهدف حاجز** **حاجز السلام**	**الجيش الحر يستهدف** **الحر يستهدف حاجز** **يستهدف حاجز السلام**	**الجيش الحر يستهدف حاجز** **الحر يستهدف حاجز السلام**	**الجيش الحر يستهدف حاجز السلام**

B. WordNet

The output of N-gram represents the inputs of Arabic WordNet (AWN). The main aim of the AWN is to enrich the extracted features from tweets by finding additional features (synonyms) in order to enhance the accuracy of sentiment analysis using wide possible synonyms. Table 7 shows example of additional synonyms that can added by AWN to enrich the extracted features from tweets.

Table 7: Example from AWN

Word	**Added Synonyms**
فرق	فارق, فرق, فريق, فراق, فرقه, فرقان, تفارق, تفرق, تفرقه, تفريق,
حرر	حراري, حرير, حرر, حر, حريه, محرر, متحرر, تحرير, تحرري, تحريري
بطل	البطل, باطل, بطاله, بطل, بطال, بطلان,

3.3 Features Classification

In this phase, the author uses 80% of dataset as training set while the rest tweets (20%) of are used as testing set for the classification system. The training set is used to learn the system the sentiment of extracted features while the testing is set to evaluate the accuracy of the proposed methods. For both training and testing sets, the extracted features from various methods (tokenization, POS, stemming, N-gram, and AWN) are transferred to numerical data in order to address the processes of sentiment classification using SVM method. The numerical data of all extracted features represent a vector of SVM. This vector contains the frequency of each extracted feature in every tweet.Table 8 presents examples of numerical vector of SVM for five tweets. The original (stored in dataset) sentiment of each tweet is connected with the produced vectors. The feature “اموي:Umaway” is appeared in the tweet #1 for one time, and this feature is not appear in the other four tweets. Thus, the “Umaway”will be denoted as positive feature because the first tweet considered as positive polarity. “سوريا : Syria” feature appeared for one time in each tweet. Therefore, “Syria” will be denoted as negative feature because it computed four times as negative and one time as positive. Same processes conducted on “NNP” feature which denoted as negative feature because it appear 11 times as negative and 4 times as positive.

Table 8 Example of SVM Vector

Tweet	أموي	سوريا	بسم	الله	الرحمن	الرحييم	فرقة	تحرير	الشام	ابطال	اللواء	VBD	NNP	NN	Original Sentiment

#	(Umaway)	(Syria)													in Dataset
1	1	1	1	1	1	1	1	1	1	1	1	1	4	2	Positive
2	0	1	0	0	0	0	0	0	0	0	0	4	3	3	Negative
3	0	1	0	0	0	0	0	0	0	0	0	1	5	0	Negative
4	0	1	0	0	0	0	0	0	0	0	0	1	2	0	Negative
5	0	1	0	0	0	0	0	0	0	0	0	2	1	0	Negative

Regarding to training set, the full supervised training is conducted. The frequency of each feature in all tweets is calculated. In the above example, NNP feature that collected from POS method was appeared 15 times (4 as positive and 11 as negatives). Thus the system training considered this feature as negative feature in the total. Another example, the "فرقة" feature was appeared one time in the vector, and this feature appeared as positive. Thus, this feature is considered as positive feature in when appear in any tweet.

Technically, the training set was applied on 1600 of 2000 tweets. The produced vector of training processes consists of 11,262 features that extracted from various feature extraction methods (tokenization, POS, stemming, N-gram, and AWN). SVM was trained to classify the features into three classes; positive, negative, and neutral (Figure 2).

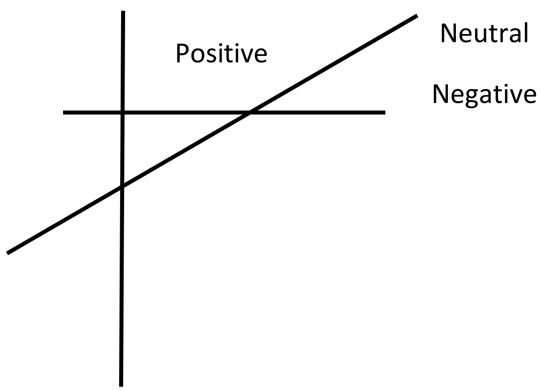

Figure 2: Classification Spaces of SVM

The feature classification of the features in the produced vector are represented by term vectors of the form d = (ti, tj, ..., tp) where each feature (tk) identifies a content term assigned to sample of tweets (d) as is done in the popular vector representation for in-formation retrieval. Typically, each jth in d is represented as a vector of weight of the content terms selected, where V is the set of features that occur at least once in at least one d, and wkj represents how much term tk contributes to the semantics of dj. Each element wkj is calculated as a combination of the statistics tf(tk ,dj) and idf (tk). This weighting scheme starts with the frequency of a term in a given tweets tf (tk, dj), and multiplies this by the "inverse document frequency" idf(tk) of the term in the corpus. The idf of a term is lower the more tweets appears in. The idea is that the more tweets a feature appears in, the less likely it is to be a good measure for distinguishing one tweets from another. The tf-idf formula for a term tk is as follows:

$$w_{kj} = tf(t_k, d_j) \times idf(t_k)$$

Where tf(tk ,dj) is equal to when term tk is not assigned to dj ,and equal to (tk ,dj) for the assigned terms. The idf term is calculated as follows:

$$idf(t_k) = \log\left(\frac{|T_r|}{df(t_k)}\right)$$

Where (|Tr |) is the total number of features in the training tweets and df (tk) is every feature frequency in each tweets. The idf (tk) is the weighting heuristic says that a feature tk is an important indexing of sentiment class (positive, negative, or neutral). Each feature indexing was conducted as the following:

1. If idf (tk) of feature is mostly appeared in the positive tweets, then this feature classified as positive feature.
2. If idf (tk) of feature is mostly appeared in the negative tweets, then this feature classified as negative feature.
3. If idf (tk) of feature is approximately same appeared in the negative and positive tweets, then this feature classified as neutral feature.

Regarding to testing set, the classification test using SVM was applied on 20% of dataset. The classification processes can be describes as the following:

1. For each feature in the tweet, idf (tk) is computed.
2. The mean of all features weights in the tweet is calculated. Here, there are three possible means; (1) positive in overall, negative in overall, or neutral.
3. The tweet is classified based on the calculated mean in step 2.

3.4 Evaluation

According to Khan et al. (2015), the testing and evaluation of the proposed system need to be conducted carefully to show the effectiveness of the proposed system. The evaluation of the sentiment analysis system can be performed based on the accuracy of system output (sentiment classification). The accuracy of sentiment classification reflects the performance of the methods that utilized and integrated to construct the proposed system.

In this research, the accuracy of proposed system was calculated using Precision, Recall and Accuracy formulas. The measurement equations are as follows (Rushdi et al. 2013):

- Precision = TP/ (TP + FP)
- Recall = TP/ (TP + FN)
- Accuracy = (TP + TN)/ (TP + TN + FP + FN)

Where:

- TP- True Positive, all the tweets which were classified correctly as positive
- TN- True Negative, all the tweets which were correctly classified as negative
- FP- False Positive, all the tweets which were incorrectly classified as positive
- FN- False Negative, all the tweets which were incorrectly classified as negative.

The F-measure of proposed system was calculated based on systemic evaluation, and the evaluation processes are as the following:

a. Conduct the final classification of the three sentiments (positive, negative, and natural) based on the proposed methods of sentiment analysis. .
b. Extract the matrix of sentiment matching (TP, FP, and FN).
c. Compute the precision of each sentiment (positive, negative, and natural).
d. Compute the recall of each sentiment (positive, negative, and natural).
e. Compute the accuracy of each sentiment (positive, negative, and natural).
f. Compare each classified sentiment of each tweet with the original sentiment that stored already in the dataset.

g. Compute the overall accuracy of all sentiments.

4 Results and Discussion

In the experiments, the author uses 80% of tweets as training set while the rest (20%) were used as testing set for the classification system. Thus, the first 1600 tweets in SYR dataset were used in the training processes, and the last 400 tweets were used in the testing processes.

The training set was applied on 1600 of 2000 tweets. The produced vector of training processes consists of 11,262 features that extracted from various feature extraction methods (tokenization, POS, stemming, N-gram, and AWN). SVM was trained to classify the features into three classes; positive, negative, and neutral.

Regarding to testing set, the classification test using SVM was applied on 400 of 2000 tweets. The classification processes can be describes as the following:

1. For each feature in the tweet, the features polarities are identified based on training processes.
2. The mean of all features weights in the tweet is calculated. Here, there are three possible means; (1) positive in overall, negative in overall, or neutral.
3. The tweet is classified based on the calculated mean in step 2.

The results of the testing set (sentiment classification of each tweet) were compared with the original sentiment that stored in SYR dataset in order to measure the classification accuracy using F-score formulas.

According to sentiment classification results of proposed system comparing with the original sentiment that stored in SYR dataset, Table 9 shows the matrix of the true and false classification of each sentiment (negative, neutral, and positive).

Table 9: Matrix of sentiment classification

	Negative	**Neutral**	**Positive**
Negative	253	18	28
Neutral	10	16	7
Positive	22	2	44

Based on the matrix in above Table 9, the classification accuracy was computed based on three formulas; precision, recall, and F measurement. Table 10 presents the accuracy of sentiment analysis and classification of all tested tweets. The accuracy of negative classification of proposed system was recorded 0.866, the accuracy of neutral classification was recorded 0.464, and the accuracy of positive classification was recorded 0.599. The overall accuracy of the proposed system is 0.783.

Table 10: Accuracy of Sentiment Classification of Tested Tweets

Sentiment Class	**Precision**	**Recall**	**F Measure**
Negative	0.888	0.846	0.866
Neutral	0.444	0.485	0.464
Positive	0.557	0.647	0.599
Overall Accuracy	*TP+TN / (TP+TN+FP+FN)= 313/400*		**0.783**

The accuracy result of the proposed system was compared with SAMAR system which developed by Abdul-Mageed et al. (2013) work.

1. TGRD: this dataset consists of 3015 Arabic tweets that were collected in May 2010. TGRD tweets are mixed between the formal and modern Arabic language. 1446 tweets of TGRD are represented by modern Arabic language, while the entire tweets are represented by the formal Arabic language. The comparison between the accuracy results of proposed system based on SYR and the accuracy results of SAMAR based on TGRD considered as the most suitable comparison due to matching between SYR and TGRD specifications. Both of these datasets are collected from Arabic tweets.
2. MONT: this dataset consists of 3097 Arabic sentences that collected from web forums in 2010.

The accuracy of proposed system using SYR dataset recorded good accuracy comparing with the SAMAR system using various datasets. SAMAR system records accuracy results as 0.66 using MONT dataset, 0.67 using TGRD dataset, and 0.73 using DAR dataset. On the other hand, the proposed system using SYR dataset records 0.783 as accuracy result.

The discussion standpoint between SAMAR and the proposed system is based on the utilized methods to solve the challenges of sentiment analysis for Arabic language. SAMAR was developed based on two methods which are; (1) features detection using POS, and (2) sentiment classification using SVM. SAMAR focused on the orthographic challenge in Arabic language. POS is effective to address the orthographic of Arabic language rather than other challenges such as morphology and synonyms widespread. The proposed system utilized several methods to address three main challenges (refer to problem statement in chapter 1) of sentiment analysis for Arabic, which are the following:

- Morphology challenge: this challenge addressed by the proposed system through utilize the stemming (word root), tokenization methods to cover the different shapes of letters found at the middle, end, and beginning of text.

- Orthographic challenge: POS is method utilized to address the orthographic and lack of capitalization challenges through identifies the concepts entity based on the sentences pattern.

- Variety of Synonyms**:** N-gram method and AWN tool utilized to extract the various meaning or synonyms of Arabic features.

The utilized methods in the proposed system to address various challenges in Arabic language are clarifying the reason of record better accuracy results than SAMAR system. Therefore, the accuracy result of the proposed system using SYR dataset considered as effective results due to its performance over SAMAR and the solved challenges that face the sentiment analysis for Arabic language.

5 Conclusion

Arabic sentiment analysis still face many challenges due to complexity of Arabic language such as morphology, Orthographic, and widespread of synonyms. This work enhances the accuracy of sentiment analysis for Arabic language through three main phases. In the preprocessing phase, there are many methods are used to analyze the main features of Arabic sentences; tokenization and stemming. Furthermore, POS is used to identify the features based on the sentence pattern. In feature extraction phase, N-gram is used to list the extract the identified features as 1-gram, 2-gram... n-gram. The Arabic WordNet tool is applied to enrich the extracted features by find more features or synonyms of the same meaning. In the feature classification phase, SVM method is adopted to classify the sentence according as positive, neutral, or negative based on the related features to this sentence. The classification accuracy measured using F-Score formulas through compare the system classification results with the true sentiment of each tweeter sentence or message. The results of experimental tests show that the proposed methods are effective to improve the accuracy of sentiment analysis for Arabic sentences.

References

1. Ababou, N.& Mazroui, A. 2016. A hybrid Arabic POS tagging for simple and compound morphosyntactic tags. *International Journal of Speech Technology*, *19*(2), 289-302.
2. Abbasi, A., Chen, H., & Salem, A. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Transactions on Information Systems (TOIS)*, *26*(3), 12.
3. Abdul-Mageed, M., Diab, M. T., & Korayem, M. 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*,*2*, 587-591.
4. Abdul-Mageed, M., Diab, M.& Kübler, S. 2013. SAMAR: Subjectivity and sentiment analysis for Arabic social media. *Computer Speech & Language*,*28*(1), 20-37.
5. Aldayel, H. K., & Azmi, A. M. 2015. Arabic tweets sentiment analysis–a hybrid scheme. *Journal of Information Science*, 0165551515610513.
6. Al-Kabi, M. N., Gigieh, A. H., Alsmadi, I. M., Wahsheh, H. A., & Haidar, M. M. 2014. Opinion mining and analysis for arabic language. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 5(5).
7. Al-Maimani, M. R., Al Naamany, A., & Bakar, A. Z. A. 2011. Arabic information retrieval: techniques, tools and challenges. In *GCC Conference and Exhibition (GCC)*, *2011 IEEE*, 541-544.
8. Al-Marghilani A., Zedan H., Ayesh A. 2007. A General Framework For Multilingual Text Mining Using Self-Organizing Maps. The *25th IASTED Int. Multi-Conf: Artificial Intelligence and Applications (AIA'07)*,520-525.
9. Al-Marghilani A., Zedan H., Ayesh A.2008. Text Mining Based on the Self-Organizing Map Method for Arabic-English Documents. *Proc. of the 19th Midwest Artificial Intelligence and Cognitive Science Conf. (MAICS 2008)*, 174-181.
10. Almutawa, F., & Izwaini, S. 2015. Machine Translation in the Arab World: Saudi Arabia as a Case Study.
11. Al-Sabbagh, R. and Girju, R. 2012. Yadac: Yet another dialectal arabic corpus. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12),* Istanbul, Turkey. European Language Resources Association (ELRA).
12. Barbosa.L & Junlan F. 2010. Robust Sentiment Detection on Twitter from Biased and Noisy Data. Coling 2010.
13. Bdour, N., & Gharaibeh, K. 2013. Development of Yes/No Arabic Question Answering System. *arXiv preprint arXiv*, 1302(5675).
14. Beesley, K. R. 2001. Finite-state morphological analysis and generation of Arabic at Xerox Research: Status and plans in 2001. In *ACL Workshop on Arabic Language Processing: Status and Perspective*,*1*, 1-8.
15. Biggins, S., Mohammed, S., Oakley, S., Stringer, L., Stevenson, M. & Priess, J. 2012. University_of_Sheffield: Two Approaches to Semantic Text Similarity. *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2. Proceedings of the Sixth International Workshop on Semantic Evaluation*, 655-661.
16. Brants, T.& Franz, A. 2006. Web IT 5-gram Version 1. www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13 [Accessed on 23 February 2016].
17. Cambria, E., & Hussain, A. 2012. *Sentic Computing: Techniques, Tools, and Applications.* Vol. 2. Springer Science & Business Media.
18. Cambria, E., Schuller, B., Xia, Y.& Havasi, C. 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, *28*(2), 15-21.
19. Chen, B., Zhu, L., Kifer, D.& Lee, D. 2010. What Is an Opinion About? Exploring Political Standpoints Using Opinion Scoring Model. *In AAAI*.
20. Cortes, C.& Vapnik, V. 1995. Support-vector networks. *Machine learning*, *20*(3), 273-297.

21. Das, S. R., & Chen, M. Y. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science,53*(9), 1375-1388.
22. Dave, Kushal, Steve Lawrence, & David M. 2003. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. in *Proceedings of International Conference on World Wide Web (WWW- 2003).*
23. Davidov, D., Oren, T., & Ari, R. 2010. Enhanced sentiment learning using twitter hashtags and smileys. *In Proceedings of Coling*.
24. Elgibali, A. 2005. *Investigating Arabic: Current parameters in analysis and learning* (Vol. 42). Brill.
25. El-Halees A. 2008. A Comparative Study on Arabic Text Classification. *Egyptian Computer Science Journal,20*,2.
26. Federici, S., Montemagni, S. & Pirrelli, V. 1996. Shallow parsing and text chunking: a view on underspecification in syntax. *Cognitive Science Research Paper-University of Sussex (CSRP)*, 35-44.
27. Gebremeskel, G. 2011. Sentiment Analysis of Twitter posts about news.Sentiment Analysis. Feb.
28. Ghwanmeh S. 2005. Applying Clustering of Hierarchical K-means-like Algorithm on Arabic Language. *Int. Journal of Information Technology,3*,3.
29. Go, A, Bhayani, R. & Huang, L. 2009. Twitter Sentiment Classification using Distant Supervision.
30. Gupta, N.K., Rahim, M.G. & Riccardi, G. 2007. U.S. Patent No. 7,197,460. Washington, DC: U.S. Patent and Trademark Office.
31. Hatzivassiloglou, V. & McKeown, K. R. 1997. Predicting the semantic orientation of adjectives. *In Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, 174-181.
32. Hu, M. & Liu, B. 2004. KDD-2004-Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. *In KDD-2004-Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*
33. Internet world stats. 2013.http://www.internetworldstats.com/stats7.htm.
34. Jakob, N. & Gurevych, I. 2010. Extracting opinion targets in a single-and cross-domain setting with conditional random fields. *In Proceedings of the 2010 conference on empirical methods in natural language processing*,1035-1045.
35. Joachims, T. 1999. *Making large scale SVM learning practical*. n.p: Universität Dortmund.
36. Joshi, M., Das, D., Gimpel, K., & Smith, N. A. 2010. Movie reviews and revenues: An experiment in text regression. In Human Language Technologies: *The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 293-296.
37. Kessler, J. S. & Nicolas N. 2009. Targeting sentiment expressions through supervised ranking of linguistic configurations. *In Proceedings of the Third International AAAI Conference on Weblogs and Social Media (ICWSM-2009).*
38. Khan, A. Z., Atique, M., & Thakare, V. M. 2015. Combining lexicon-based and learning-based methods for Twitter sentiment analysis. *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE)*, 89.
39. Kim, S. M.& Hovy, E. 2004. Determining the sentiment of opinions. *In Proceedings of Interntional Conference on Computational Linguistics(COLING-2004).*
40. Kim, S. M. & Hovy, E. 2006. Identifying and analyzing judgment opinions. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics.* Association for Computational Linguistics. pp. 200-207.
41. Kim, Y. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv,*1408.5882.
42. Korayem, M., Crandall, D., & Abdul-Mageed, M. 2012. Subjectivity and sentiment analysis of arabic: A survey. *In Advanced Machine Learning Technologies and Applications*, 128-139.
43. Lin, C.Y. & Och, F.J. 2004. Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. *In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, 605.
44. Lin, H.T., & Lin, C.J. 2003. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. *Neural Computation*, 1-32.
45. Lin, J.Y., Cheng, C.T. & Chau, K.W. 2006. Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal,51*(4), 599-612.

46. Liu, B. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies, 5*(1), 1-167.
47. Liu, J., Cao, Y., Lin, C. Y., Huang, Y., & Zhou, M. 2007. Low-Quality Product Review Detection in Opinion Summarization. *In EMNLP-CoNLL,7*, 334-342.
48. Maamouri, M., Bies, A., Buckwalter, T., & Mekki, W. 2004. The penn arabic treebank: Building a large-scale annotated arabic corpus. *In NEMLAR conference on Arabic language resources and tools,27*, 466-467.
49. McGlohon, M., Glance, N. S. & Reiter, Z. 2010. Star Quality: Aggregating Reviews to Rank Products and Merchants. In *ICWSM*.
50. Mesfar, S. 2010. Towards a cascade of morpho-syntactic tools for arabic natural language processing. *In International Conference on Intelligent Text Processing and Computational Linguistics*, 150-162.
51. Mihalcea, R., Corley, C. & Strapparava, C. 2006. Corpus-based and Knowledge-based Measures of Text Semantic Similarity. *In Proceedings of AAAI'06 the 21st national conference on Artificial intelligence,1*, 775-780.
52. Miller, G.A. 1995. WordNet: A lexical Database for English. *Communications ACM,38*(11), 39.
53. Min, J.H. & Lee, Y.C. 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert systems with applications, 28*(4), 603-614.
54. Mohler, M. & Mihalcea, R. 2009. Text-to-Text Semantic Similarity for Automatic Short Answer Grading. *In Proc. 12th Conf. European Chap. ACL,* Athens, Greece, pp. 567-575.
55. Monem, A. A., Shaalan, K., Rafea, A., & Baraka, H. 2008. Generating Arabic text in multilingual speech-to-speech machine translation framework. *Machine Translation, 22*(4), 205-258.
56. Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T. 2002. Mining product reputations on the web. *In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining,* 341-349.
57. Motik, B., Sergio Antonio B. N., Pablo C. G.& González-Cadenas,C. 2011. Natural Language Question Answering System and Method Based On Deep Semantics. U.S. Patent Application 13/171,391, filed June 28, 2011.
58. Nasukawa, T., & Yi, J. 2003. Sentiment analysis: Capturing favorability using natural language processing. *In Proceedings of the 2nd international conference on Knowledge capture*, 70-77.
59. Nitish, A., Asooja, K., & Buitelaar, P. 2012. DERIandUPM: Pushing corpus based relatedness to similarity: Shared task system description. *In Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation,* 643-647.
60. Ortony, A., Clore, G. L. & Collins, A. 1990. *The cognitive structure of emotions*. Cambridge university press.
61. Pang & Lee, L. 2004. A Sentiment Education:Sentiment Analysis using Subjectivity Summarization Based on Minimum Cuts. *In Proceeding 42nd Ann. Meeting of the Assoc. for Computional Linguistics*, 271-278.
62. Pang, B., & Lee, L. 2008. Opinion mining and sentiment analysis.*Foundations and trends in information retrieval, 2*(1-2), 1-135.
63. Pang, B., Lee, L., & Vaithyanathan, S. 2002. Thumbs up?: Sentiment classification using machine learning techniques. *In Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 10*, 79-86.
64. Parikh.R.& Matin, M. 2009. Sentiment Analysis of User-Generated Twitter Updates using Various Classification Techniques.
65. Poibeau, T. & Kosseim, L. 2001. Proper Name Extraction from Non-Journalistic Texts. *In Proceeding Computational Linguistics in the Netherlands*, 144–157.
66. Refaee, E., & Rieser, V. 2014. An Arabic Twitter Corpus for Subjectivity and Sentiment Analysis. *In LREC,* 2268-2273.
67. Rushdi-Saleh, M., Martín-Valdivia, M. T., Ureña-López, L. A., & Perea-Ortega, J. M. 2011. OCA: Opinion corpus for Arabic. *Journal of the American Society for Information Science and Technology, 62*(10), 2045-2054.
68. Sakunkoo, P., & Sakunkoo, N. 2009. Analysis of Social Influence in Online Book Reviews. *In ICWSM*.
69. Salameh et al., 2015. Arabic Sentiment Analysis and Cross-lingual Sentiment Resources. Retrieved from: http://saifmohammad.com/WebPages/ArabicSA.html.

70. Salton G., Buckley C. 1998. Term weighting approaches in automatic text retrieval. *Proc. of information processing & management,24,*(5), 513-523.
71. Schwartz, L., Aikawa, T. & Pahud, M. 2004. Dynamic language learning tools. *In InSTIL/ICALL Symposium 2004.*
72. Sha, F. & Pereira, F. 2003. Shallow parsing with conditional random fields. *In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology,1*, pp. 134-141.
73. Shaalan, K. 2014. A survey of arabic named entity recognition and classification. *Computational Linguistics, 40*(2), 469-510.
74. Shoukry, A., & Rafea, A. 2012. Sentence-level Arabic sentiment analysis. *In Collaboration Technologies and Systems (CTS), 2012 International Conference on*, 546-550. IEEE.
75. Siddiqui, S. 2015. Sentiment Analysis for Arabic. Master Thesis.
76. Somasundaran, S., Galileo, N., Lise, G.& Janyce W. 2009. Opinion graphs for polarity and discourse classification. *In Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing.*
77. Stevenson, R. A., Mikels, J. A. & James, T. W. 2007. Characterization of the affective norms for English words by discrete emotional categories. *Behavior research methods, 39*(4), 1020-1024.
78. Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM, 10*, 178-185.
79. Turney, P. D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *In Proceedings of then association for computational linguistics* 417-424.
80. Wan, S. & Angryk, R.A. 2007. Measuring Semantic Similarity using WordNet-based Context Vectors. *In Proceeding IEEE International Conference on Systems, Man and Cybernetics (ISIC),* 908-913.
81. Wiebe, J. 2000. Learning subjective adjectives from corpora. *In AAAI/IAAI,* 735-740.
82. Wiebe, J. M., Bruce, R. F. & O'Hara, T. P. 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*,246-253).
83. Wiebe, J. 1990. Identifying subjective characters in narrative. *In Proceedings of the International Conference on Computational Linguistics (COLING-1990).*
84. Wiebe, J. 2000. Learning subjective adjectives from corpora. *In Proceedings of National Conf. on Artificial Intelligence (AAAI-2000).*
85. Wilson, T., Wiebe, J., & Hoffmann, P. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. *In Proceedings of the conference on human language technology and empirical methods in natural language processing*, 347-354.
86. Wilson, T., Wiebe, J., & Hoffmann, P. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. *In Proceedings of the conference on human language technology and empirical methods in natural language processing,* 347-354.
87. Wilson, T., Wiebe, J., & Hwa, R. 2004. Just how mad are you? Finding strong and weak opinion clauses. *In aaai 4,* 761-769.
88. Wright, A. 2009. Our sentiments, exactly. Commun. *ACM, 52*(4), 14-15.
89. Ye, Q., Zhang, Z. & Law, R. 2009. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications, 36*(3), 6527-6535.
90. Yi, J., Tetsuya, N., Razvan, B.& Wayne, N. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. *In Proceedings of IEEE International Conference on Data Mining (ICDM-2003).*
91. Yu, H.& Vasileios, H. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. *In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2003).*
92. Zaidan, O. F. & Callison-Burch, C. 2013. Arabic dialect identification. *Computational Linguistics.*