Opinion Mining and Analysis for Arabic Language

Mohammed N. Al-Kabi

Faculty of Sciences & IT
Zarqa University
Zarqa, Jordan

Amal H. Gigieh

Faculty of Sciences & IT
Ajloun- College
AL-Balqa' Applied University
Ajloun, Jordan

Izzat M. Alsmadi

Computer Science Department
Boise State University
Boise, ID 83725, USA

Heider A. Wahsheh

Computer Science Department
College of Computer Science
King Khalid University
Abha, Saudi Arabia

Mohamad M. Haidar

Web development Department
Brandtologie Company

Abstract—**Social media constitutes a major component of Web 2.0 and includes social networks, blogs, forum discussions, micro-blogs, etc. Users of social media generate a huge volume of reviews and comments on a daily basis. These reviews and comments reflect the opinions of users about different issues, such as: products, news, entertainments, or sports. Therefore different establishments may need to analyze these reviews and comments. For examples: It is essential for companies to know the pros and cons of their products or services in the eyes of customers. Governments may want to know the attitude of people towards certain decisions, services, etc. Although the manual analysis of textual reviews and comments can be more accurate than the automatic methods, nonetheless, it is time consuming, expensive, and can be subjective. Furthermore, the huge amount of data contained in social networks can make it impractical to perform analysis manually. This paper focuses on evaluating Arabic social content. Currently, Middle East is an area rich of major political and social reforms. The social media can be a rich source of information to evaluate such contexts. In this research we developed an opinion mining and analysis tool to collect different forms of Arabic language (i.e. Standard or MSA, and colloquial). The tool accepts comments and opinions as input and generates polarity based outputs related to the comments. Additionally the tool can determine the comment or review is: (subjective or objective), (positive or negative), and (strong or weak). The evaluation of the performance of the developed tool showed that it yields more accurate results when it is applied on domain-based Arabic reviews relative to general-based Arabic reviews.**

Keywords—*Sentiment Analysis; Arabic Sentiment Analysis; Opinion mining; Opinion Subjectivity; Opinion Polarity*

I. INTRODUCTION

The use of Internet is changed significantly through decades, where in the beginning it was restricted to the connection of four major U.S. universities and a number of government agencies in 1969. Since then the number of servers start to increase rapidly. In 1989 a major event occurs when Tim Berners-Lee casts the term World Wide Web (WWW) which is based on hypertext, and changed the way of communication through the Internet. In 2004 the term Web 2.0 is used and a number of services and tools are released, to make the WWW more cooperative and sharable. Therefore the key component of Web 2.0 is social media, which helps to serve different societies all around the world.

Web 2.0 is offering products and services that are different from its predecessor WWW. The number of Web 2.0 users increase on a daily base, where it is impossible for a single user to learn and use all these products and services. Web 2.0 helps to let it users to be more collaborative. YouTube, Facebook, Twitter, etc. are examples of Web 2.0 services. Arabic language is spoken by many people in many countries. Arabs constitute around 5% of World population and around 3.8% of Internet users [1]. Last ten years witnessed an explosive increase in the volumes of social media content which is broadcasted and shared, and this content is related to different daily activities. This may include data which use Modern Standard Arabic (MSA) and colloquial Arabic. The colloquial Arabic is greatly varied, and is classified into five main regional forms in the Middle East [2]:

1) Arabian Peninsula Arabic (Khaliji Arabic): Includes Gulf, Baharna, Najdi, Omani, Hejazi, Shihhi, Dhofari, and Yemeni Arabic slangs.

2) Mesopotamian Arabic: Includes Iraqi and North Mesopotamian Arabics.

3) Syro-Palestinian Arabic: Includes Levantine, Judeo, Mediterranean Sea or Cypriot, and Bedawi Arabic.

4) Egyptian Arabic: includes Chadic and Sudanese Arabic including: (Nubi, Juba and Darfuri Arabics), Sa'idi and Egyptian Arabic.

5) Maghrebi Arabic: includes the Arabic Vernaculars used in North African coast of the Mediterranean Sea such as: Moroccan, Tunisian, Arabic, Arabic, Arabic, and Saharan Arabics.

The Arabs who used Arabian Peninsula Arabic (i.e. Khaliji Arabic) could not understand the accent used in Maghrebi Arabic, so in this case both have to use MSA to communicate with and understand each other.

Social media data include: News stories, opinions, current status, different activities, and comments and reviews about these items. Opinions are essential to people and before the Internet era when somebody needs an opinion he/she asks his/her family, relative or a friend. Customer opinions are essential to companies; therefore they used to conduct surveys in different forms before the evolution of the Internet to evaluate people's opinions on an issue or event.

Opinions are then very important. Whenever we need to make a decision we want to hear others' opinions. This is not only true for individuals who may use advices from the others, but it is true for organizations and governments. Many tools were built and developed to analyze English opinions. The interest in opinion analysis and mining has grown due to different reasons. On one side it is due to the rapid evolution of the World Wide Web (WWW), which changed the view and the use of the Internet. It has changed the web into a collaborative framework where technological and social trends come together. On the other side, the huge use of the services has been accompanied with an increase in freely available online reviews and opinions about different topics, subjects or entities [3].

Opinion mining/sentiment analysis is an emerging field of study and a very active research area since the year 2003. It is concerned with the analysis of people's sentiments, opinions, attitudes, evaluations, and emotions expressed in one of the known natural languages towards entities such as: persons, products, services, companies, events, issues, or topics. Studies in this field are conducted as part of computer science studies. However, it is conducted in management and social sciences, since only a few numbers of these studies are important to the business and to the society [4]. Sentiment analysis and opinion mining were first explored in 2003 by [5, 6]. Although these two terms (Sentiment analysis and Opinion mining) are not exactly the same, but they used interchangeably by a number of authors, where the meaning of term opinion is broader than the meaning of the term sentiment.

Web-based social network services such as: Twitter, Facebook, and Google+ enable different users with common interests or real-life connections to connect with each other through those virtual networks to share their opinions, ideas, and information. These Web services are applied in different domains such as: Government, Business, Dating, Education, Finance, Medical/health, Social and political applications [7].

According to the leading free provider of Internet Web metrics, Alexa (www.alexa.com) [8], social network sites such as: Facebook were ranked second globally at the time of conducting this study [9]. Moreover, YouTube is ranked third,

and Twitter ranked tenth. Those social networks in the top ten showed that such websites and services are widely used by humans all over the world. In the Arab countries these Web metrics are similar to those presented on the global level. In Egypt the largest Arab country for example, Facebook is ranked first, YouTube ranked third, etc. The same thing can be said about other countries in the region.

Most opinion analysis and mining methods have been developed for English text and are difficult to generalize to other natural languages such as: Arabic which is highly inflectional. The number of studies in this field which are conducted on Arabic text whether it is expressed in MSA or colloquial Arabic is limited when it's compared to the number of studies conducted on English sentiments and opinions. Arabic is one of the Semitic languages which is written from right to left, and written in a cursive way. Furthermore Arabic language has 28 consonants, and has no upper and lower case consonants as in English.

Arabic is a challenging language for a number of reasons: It has a very complex morphology relative to the morphology of other languages such as: English. Arabic language is a highly inflectional and derivational language which makes monophonically analysis a very complex and difficult task [10]. Furthermore Arabic opinions are highly subjective to context domains, where you may face words that have different polarity categories in different contexts. Arabic Internet users mostly used colloquial Arabic rather than using MSA, where colloquial Arabic resources are scarce. The percentage of spelling mistakes within these Arabic opinions is high, and this represents an additional challenge.

These few lines would not be sufficient to list the differences between Arabic and English languages. Therefore it is impossible to apply most of the opinion analysis and mining methods which are proposed and implemented on English sentiments and opinions directly on Arabic sentiments and opinions. A few numbers of these studies is related to Arabic opinions/sentiments analysis, and are using the analysis methods developed mainly, but not directly for English language. Therefore, such studies are using machine translation (MT) to automatically translate Arabic sentiments and opinions to English, in order to be able to use those analytical methods which are designed mainly for English opinions/sentiments. For example Bautinet et al. study [11] and Rushdi-Saleh et al. [12] study conclude that this approach is an attractive one. The use of MT will lead to degradation of the accuracy of final results of the opinion analysis and mining, as a result of the incapability of MT systems nowadays to accurately translate from one natural language into another, as accurately as professional human translators. Our intuition or idea is that such translation is not necessary and is not effective and does not yield more accurate results than methods that used directly to mine opinions and sentiments, without using machine translation.

In this research, we have developed a tool to analyze different Arabic opinions whether they are written in colloquial Arabic or Modern Standard Arabic (MSA) or both. This was an ambitious goal to develop a tool to deal with both standard and colloquial Arabic. In comparison with previous

tools such as those mentioned in Rushdi-Saleh et al. [12] study which is restricted to MSA, Almas and Ahmad [13] study which is restricted to MSA Arabic financial terms, and El-Halees [14] study which uses MSA, our tool can be hence be more comprehensive.

In this study different opinions written in MSA or/and colloquial Arabic are classified into a predefined set of categories based on their contents. Classifying those different opinions is not a straight forward process, since the essential lexical resources are not there, especially those related to colloquial Arabic. Implicitly this study includes a manual building of two general purpose lexicons to discern the polarity of an opinion expression, whether the opinion uses MSA or/and colloquial Arabic. Furthermore, another sixteen domain-specific lexicons were built manually. Those domain-specific lexicons were built to decide automatically the polarity of a sentiment expression within the following eight domains: Technology, Books, Education, Movies, Places, Politics, Products, and Society. So the total number of lexicons built is 18, where nine of these polarity lexicons are dedicated to positive polarity, and the other nine lexicons are dedicated to negative polarity. An opinion is considered neutral, when its tokens are divided equally between positive and negative lexicons. The tool is capable to determine whether Arabic social media reviews are (subjective or objective), (positive or negative), and (strong or weak).

The rest of this paper is organized as follows. Section 2 overviews related work. Section 3 describes the methodology followed with examples showing exactly how our tool works. Section 4 exhibits the algorithms implemented in our opinion mining tool. Section 5 presents the results of the experimental analysis and evaluation. Finally in section 5, conclusions and possible future work are discussed.

II. RELATED WORKS

A review to previous studies conducted in this field shows that researchers proposed and used several approaches which provide variant solutions to automatic sentiment analysis and opinion mining. This section exhibits few numbers of these studies about this field, with an emphasis on studies related to automatic analysis of Arabic sentiments and reviews.

Sentiment analysis systems can be divided according to the scope of the input; therefore we have document-level (where the classification of opinions depend on the whole document), sentence-level, or phrasal-level which analyzes part of the sentence. Sentence-level sentiment analysis classify sentiments after segmenting the document into several sentences and compute the polarity of each sentence, while document-level sentiment analysis systems do not segment sentiment's document into several sentences. Pang et al. [15] used a document level polarity categorization to classify opinions. El-Halees [14] study evaluated three different methods to identify the polarity of documents. Yi et al. [16], Kim et al. [17], Elhawary and Elfeky [18], and Abdul-Mageed et al. [19] on the other hand dealt with sentence-level polarity categorization attempts to classify positive and negative sentiments for each analyzed sentence. Phrase-level sentiment analysis is conducted by Wilson et al. [20], where they determine first whether the expression is neutral or has a

polarity. Afterward if the expression under consideration is not neutral, the contextual polarity is determined.

Elhawary and Elfeky [18] study is similar to our study, since it discussed the lack of a standard Arabic dataset for business reviews and sentiments. For Arabic, the Internet lacks websites similar to www.yelp.com which has many English business reviews. Therefore their study started by collecting Arabic business reviews, and dedicating 80% of the collected business reviews to train their classifier which is used to identify review's documents. They constructed a number of Arabic lexicons used to analyze different Arabic reviews and sentiments. The polarity of each Arabic business review whether it is: positive, negative, neutral or mixed is judged based on the built lexicons.

A manually annotated corpus of Modern Standard Arabic (MSA) and a polarity lexicon are developed by [19]. The authors developed high performance automatic Subjectivity and Sentiment Analysis (SSA) system which is based on manually annotated MSA corpus.

Different methods were used by El-Halees [14] to determine the polarity of a number of Arabic documents. The polarity of the whole Arabic documents is determined first using lexicon-based method, where the output from the first method (lexicon-based) is considered as a training set for maximum entropy method, which is used to classify these documents. Author used KNN method in her study to classify collected Arabic documents.

Sentiment analysis can be divided according to the type of output or the desired classification. Traditionally, sentiment analysis indicates whether a review or comment is positive, negative or neutral. Wilson et al. [21], Abbasi et al. [22], Elhawary and Elfeky [18] studies depend on lexicons containing positive and negative words/phrases ranked by their score, and classify opinions into positive, negative, neutral or mixed. In other classification category the opinions were determined as strong or weak. There are few studies proposing a feature weighting schemes that can enhance classification accuracy. Paltoglou et al. [23] study assigns weights to features and applies weighting functions scale linearly related to the number of times a term occurs in a document. This was a significant factor to increase the accuracy of sentiment classification.

One of the earlier approaches adopted in a number of studies is based on translating the source Arabic document (opinions) into English and then use the same applicable techniques to analyze the resulted English sentiments. Almas and Ahmad in [13] used machine translation systems to translate the source document or review from (Arabic, Italian, French, Chinese, Korean, German, Japanese, and Spanish) to English language before passing them to an English based sentiment analysis system. The problem of this approach was the loss of nuance after translating the source to English. Rushdi-Saleh et al. [12] used different machine learning algorithms to classify the polarity of Arabic reviews extracted from specialized Web pages related to movies and films. Inui et al. [24] study adopts translating opinions from English to Japanese, followed by sentiment analysis. They applied sentiment-oriented sentence filtering method to mitigate many

translation errors that occur as a side-effect of translation to reduce the influence of translation errors in multilingual document-level review.

The use of machine translation followed by sentiment analysis is not restricted to Arabic comments and reviews, but it includes other languages. As a sample of these studies the Banea et al. [25] is presented in this section, which used machine translation to translate Romanian and Spanish reviews and comments to English, and then apply the sentiment analysis tools on the translated materials. Moreover, they improve their study and conduct another study Banea et al. [26], where in their study they added Arabic, French, and German reviews besides Romanian and Spanish reviews and comments used in their previous study.

Some studies in this field are domain based studies. Domain features should be collected for the domain under consideration, as exhibited in the study of Balahur et al. [27], where the term is used to describe special product classes. Afterward the polarity (i.e. positive or negative) is determined for each of the features' attributes using an annotated corpus. Other researchers select domain specific features plus the topic of the opinion as a clue. Choi et al. [28] study presents a framework for sentiment analysis, focus on the sentiment clue that is related to a sentiment topic (defined as a primary subject of sentiment expression in a sentence), such as: company, person or event. They use a domain-specific sentiment classifier for each domain with the newly aggregated clues (e.g. a subject or the topic of the opinion) based on a proposed semi-supervised method. Yi et al. [17], Kim et al. [18], Choi et al. [28] extract opinion about a subject focus on the sentiment clue that is related to a sentiment topic. This is defined as a primary subject of sentiment expression in a sentence such as: company, person or event.

Ortiz et al. [29] study views and evaluate a domain independent sentiment analysis system against a multiple-domain opinion corpus. The results showed that high accuracy can be achieved by relying entirely on high quality, manually acquired and linguistic knowledge.

Al-Subaihin et al. [30] study exhibits a design for a sentiment analysis tool for Modern Arabic which segments the reviews into sentences, then collect sentimental meaning of words in each sentence based on sentiment lexicons. The tool can get the pattern of word's role in the sentence and use that pattern to match from a set of the acquired annotated patterns that map the sentence to get the polarity. The whole polarity is deducted from the sentiments of sentences. Their tool focused on Modern Standard Arabic (MSA) only while in this paper we tried to enable the tool to deal with both (Modern Standard Arabic (MSA) and Colloquial Arabic).

Al-Kabi et al. [31] conducted a study to compare two free online sentiment analysis tools: SocialMention and Twendz using Arabic and English comments and reviews. To conduct their study they constructed three polarity dictionaries: English polarity dictionary, Arabic polarity dictionary, and Emoticon polarity dictionary. They conclude that SocialMention is more effective than Twendz. Another study compares two free online sentiment analysis tools (SocialMention and SentiStrength) that support Arabic language is conducted by Khasawneh et al. [32] and based on 1,000 Arabic comments and reviews collected from Twitter and Facebook. They conclude that SentiStrength tool is more effective than SocialMention.

Al-Kabi et al. [33] collected 4625 Arabic reviews and comments from Yahoo!-Maktoob Website. The collected reviews and comments are classified manually into four domains (Arts, Politics, Science and Technology, and Social). They analyze different aspects of the collected dataset such as the reviews' length, the numbers of likes/dislikes, the polarity distribution and the languages used.

III. THE METHODOLOGY

This section presents the used approach to automatically analyze large volumes of Arabic user's reviews using both Modern Standard Arabic (MSA) and colloquial Arabic, where the analysis includes adopting classification algorithms to determine: Subjectivity, Polarity, and Intensity.

We first developed a basic lexicon-based tool for Arabic opinion mining. This tool can process Arabic opinions collected from different social media resources, regardless of their domain. Therefore this proposed tool uses word/phrasal sentiment features to handle Arabic textual opinions whether they are using MSA or colloquial Arabic or both. The following steps were followed to identify subjectivity, polarity, and intensity.

A. Opinion Analysis Schema

Sentiment analysis is concerned with analyzing the attitude of the opinion holder (i.e. the person who presented the opinion) or in other words analyzing the subjective opinions text (i.e. text containing opinions, emotions or sentiments). This study presents an automatic tool to analyze Arabic opinions regardless of the Arabic language style used whether it is MSA or colloquial Arabic or both. The tool is capable to determine the subjectivity, polarity and intensity of the evaluated Arabic opinions, where specific syntactical features are used to determine the strength of the opinion. The schematic overview of our approach is exhibited in figure 1.

This study is based of the following five phases:

1) Dataset collection.
2) Text normalization.
3) Specific features extraction from the text opinions that were collected.
4) Creating 18 Lexicons.
5) Using classification algorithms to classify opinions into several categories using the built lexicons.

Fig. 1.   Opinion Analysis Schema

B. Dataset collection

This study started by collecting Arabic reviews from 72 social media websites. The total number of the collected Arabic reviews was 1,080. These reviews use either colloquial Arabic or MSA, or both.

1) Dataset Characteristics

This section exhibits few characteristics of the collected Arabic reviews:

Some reviews consist of only one word, e.g. good,"جيد". Chat language is used to express some reviews such as:" 7ilwi awi" which means in English "very sweet".

Latin letters and English phonetics (transliteration) are used to express Arabic phrases such as "jamiljidannnnnnnnnnn" which means in English "very nice". An appropriate method has to handle repeated Latin and Arabic letters.

Some of the collected Arabic reviews use elongation, through the use of dash-like "kashida" character to stretch the Arabic word, i.e., حـلـــو جـــــــدا which means very sweet. This extension or Kashida should be handled by the developed system as well.

Most of the collected Arabic reviews use a mixture of colloquial Arabic and MSA, such as (very nice, "وايد جميل"), where the Arabic word (very, "وايد") is a colloquial Arabic word, and the Arabic word (nice, "جميل") is an MSA word.

Many of the collected Arabic reviews contain spelling errors, such as (very beautiful, "جمي جدا، جميل جدا").

Some of the textual reviews are mixtures of Arabic and English. It usual to find reviews that consist of Arabic and English, i.e." حبيت التصميم it was very nice", which means I like the design it was very nice.

Some of reviews weren't related to the topic of the review, so it is considered a spam review or irregular.

There is no exact or specific style or pattern the users have to follow to write their reviews. Therefore we are dealing with fully unstructured Arabic text.

Around 90% of the Arabic reviews in the dataset were opinions or subjective text, and around 10% were objective text (facts). The above characteristics represent a summary of the Arabic opinions' analysis problems that should be usually handled by any proposed automatic solution or handling system.

C. Preprocessing Steps

The opinion mining tool performs a number of preprocessing steps to normalize and prepare the opinions for processing. First, the tool removes punctuations and non-Arabic letters. Specified Arabic letters are normalized such as: (أ, إِ, and آ) are converted to (bare Alif, ١), and (Yaa', "ئ, ى") replaced by (Yaa', ي) and, (taa', haa', "ة") replaced by (haa',ه). These steps are shown below:

1) Remove digits, punctuations, special symbols and non-letters.

2) Normalization:

   a) Convert (Alif, أ,آ,أ,ى,آ) to (bare Alif, ١).

   b) Convert (Yaa', "ي, ئ") to (Yaa', ي).

   c) Convert (Taa' Marbuuta , "ة", final haa', "ه") to (final haa', "ه").

   d) Convert (Waaw "ؤ","و") to (Waaw, "و").

3) Filtering non-Arabic text

4) Tokenization

D. Taxonomy of Opinion Analysis

Here is the taxonomy for the major concepts and steps used to analyze different Arabic reviews. Table I presents different main taxonomies generated by the tool.

TABLE I.   MAIN CLASSIFICATION CATEGORIES

Classification Category	Description
Subjectivity Classes	Opinion/Fact
Polarity Classes	Positive/Negative/Neutral Sentiments
Strength (Intensity) Classes	Strong/Weak/Normal Opinion Text
Domain Classes	Determine on what domain is the opinion

The tool uses the features shown in table II to generate the taxonomies shown in table I. Features shown in table II are extracted manually from the collected Arabic reviews.

TABLE II.   FEATURES CATEGORIES

Feature Category	Description
Domain Features	All words or bag of words which can distinguish domains from each other.
Polarity Features	All words/phrases yield (positive or negative) sentiment in opinion text.
Negation Features	All words that preclude the word or sentence.

Table III exhibits the main techniques adopted in this study to classify different Arabic reviews.

TABLE III.   OPINION ANALYSIS TECHNIQUES

Classification Category	Description
Machine Learning	Naïve Bayes Technique.
Similarity Score	Word/Phrase Matching, frequency term counts, weight score.
Normalization and Tokenization	Prepare Arabic opinions before analysis.

This tool can handle Arabic general opinions collected from different social media recourses, and try to categorize them into specific domains. Table IV shows the domains of different Arabic reviews covered in this study.

TABLE IV.   CLASSIFICATION DOMAINS

Classification Category	Description
General Domain	Independent Base Domain.
Specific-Domain Arabic Opinion	Technology, Books, Education, Movies, Places, Politics, Products, and Society.
Web Media Corpus	Social media web pages e.g. (Facebook, blogs, online news, forums).

Our tool is based on more than one lexicon to classify different Arabic opinions. These lexicons contain the extracted features included in the dataset collection, where the content of each lexicon is shown in table V.

TABLE V.   LEXICON CATEGORIES

Category	Description
Polarity Lexicon	Contains the Positive and Negative Sentiment's features.
Domain Lexicon	Contains the features that discriminate specific domain from the others.
Strength Lexicon	Polarity lexicons with weight for each entry.
Negation Lexicon	Contains the negation words.

E. Feature Extraction

Opinion features are extracted manually. After collecting opinions' dataset, these features are used to construct different lexicons used in the analysis and classification steps. Figure 2 shows the essential steps to extract different types of features which are used in this study.
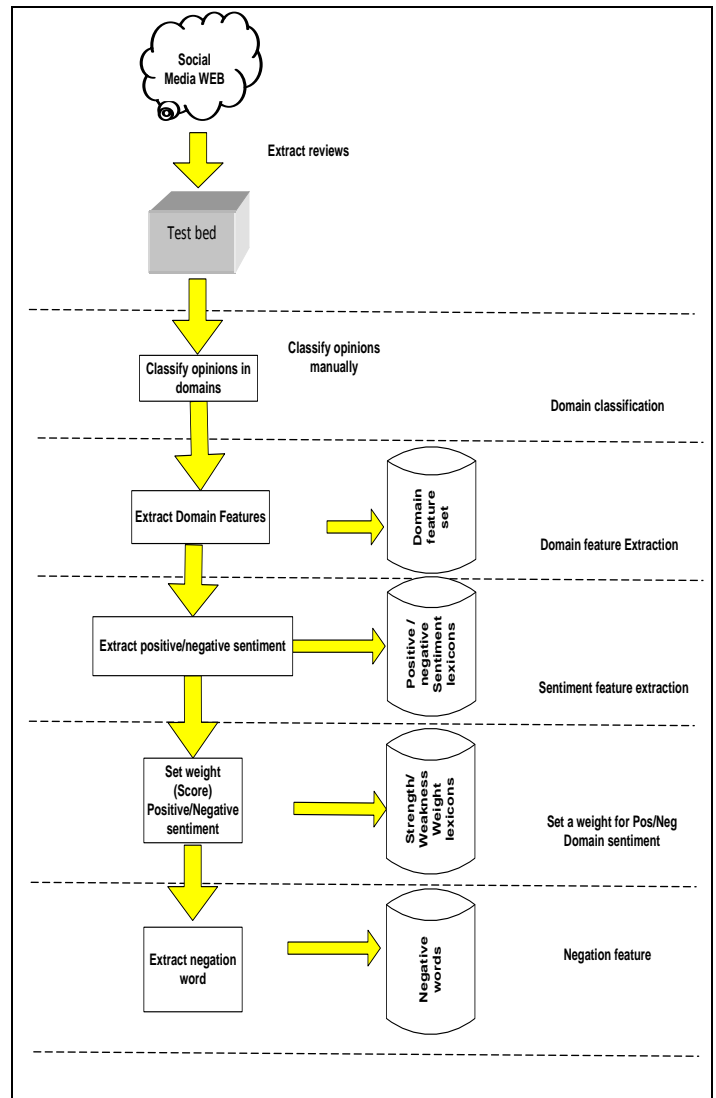
Fig. 2.   Outline of Feature Extraction

1) Domain Features

Domain features are used as clues to determine the domain to which the opinion may belong to be used by classification algorithms [34, 36]. These features are collected from the training dataset after classifying them manually into domains, to select the features that can discriminate one domain from another. In other words to use them as inputs (training data) to the classifier, to determine the instance reviews related to any domain automatically (domain adaptation).Our dataset is classified into eight domains: Technology, Books, Education, Movies, Places, Politics, Products, and Society.

To prepare the domain sentiment lexicons, we extract the domain features from the opinions text after classifying the dataset manually into the different domains.

2) Polarity Features

Polarity features are divided into positive and negative (sentiments). These features are extracted from the collected Arabic reviews to build the polarity lexicons. Arabic polarity features are Arabic words or phrases that express the positivity or the negativity of the user attitude related to a specific topic. These features are considered from syntactical point of view such as: "adjectives", "verbs", "nouns", and "adverbs". They may also come as a mixture of a "group of words".

As mentioned before the main challenge to researchers in Arabic opinion analysis field is the lack of necessary resources, especially the lack of polarity sentiment lexicons. Therefore we have to create these lexicons which contain the positive and negative features already extracted manually from Arabic reviews.

3) Negation Features

Arabic negation words represent all the words that negate Arabic words and sentences. Arabic negation keywords such as: (no, "لا") and (not, "لم") convert the sentiment polarity state to an opposite state.

4) Examples

Two polarity examples are shown in this subsection. The first example shows how to extract a positive polarity feature, while the second example shows how to extract a negative polarity feature.

Example 1: Consider the following sample in the next excerpt, of the collected Arabic reviews (Movies domain) with its English translation:

Arabic Comment	فيلم رائع جدا وممتع استمتعوا بمشاهدة هذا الفيلم و هو يعتبر نقلة في صناعة هوليود الفيلم خليط من أفلام زي ماتريكس و سبيدرمان
English Translation	"A wonderful and enjoyable movie. Enjoy watching this movie, it is considered as a shift in the Hollywood industry. It is a mix between movies like Matrix and Spiderman "

Table VI exhibits the manually extracted domain features from the above Arabic excerpt. This sample is taken from the domain of movies. Domain features are used to determine the domain of each Arabic review within the dataset. To be more specific, a classifier (e.g. NB) is used to determine the domain of each Arabic review and comment.

TABLE VI.   MANUALLY EXTRACTED DOMAIN FEATURES

أفلام	الفيلم	فيلم	بمشاهدة	هوليود
Films (Movies)	The film (movie)	Film(movie)	watching	Hollywood

Table VII exhibits a sample of positive polarity features extracted from the above Arabic sample review. These features are stored in the polarity lexicon to be used later by the tool to determine the polarity of different Arabic reviews and comments.

TABLE VII.   A SAMPLE OF POSITIVE POLARITY FEATURES

Arabic Sentiment Features	ممتع	رائع
English Translation	Enjoyable	Wonderful
Polarity	Pos	Pos

Example 2: Consider the following sample in the next excerpt, of the collected Arabic reviews (Movies domain) with its English translation:

Arabic Comment	فيلم طويل وبطيء وممل
English Translation	"The movie was long, slow and boring "

In this case the manually extracted domain features are restricted to one feature (Movie, "فيلم"), so the NB classifier is based on this feature to determine the domain of the above excerpt. Table VIII exhibits a sample of negative polarity features extracted from the above Arabic sample excerpt.

TABLE VIII.   A SAMPLE OF NEGATIVE POLARITY FEATURES

Arabic Sentiment Features	بطيء	ممل	طويل
English Translation	Slow	Boring	Long
Polarity	Neg	Neg	Neg

Moreover, two other examples are presented in this study about political reviews in example 3 and 4. These two examples show how the tool can determine the polarity of any review, and how to determine whether the review is a fact or an opinion.

Example 3:

Consider the following sample in the next excerpt, of the collected Arabic reviews (Politics domain) with its English translations:

Arabic Comment	انا ضد هذا القرار السياسي، قرار صعب يشكل ازمه للمواطنين.
English Translation	I am against this political decision, it is a hard decision It may cause a crisis for citizens

Table IX exhibits manually extracted polar features from the above Arabic political excerpt.

TABLE IX.   MANUALLY EXTRACTED FEATURES FROM AN ARABIC POLITICAL REVIEW

Arabic Sentiment Features	ازمه	صعب	ضد
English Translation	Crisis	Hard	Against
Polarity	Neg	Neg	Neg

Therefore the tool considers the above Arabic political excerpt as a negative point of view.

Next is another example that shows the review under consideration as a fact, where this review is free from any sentiment. This review expresses a fact and not an opinion.

Example 4:

Consider the following Arabic excerpt which is considered by the tool as a fact, since it is free from any polarity feature.

Fact:

Arabic Comment	هذا القرار جاء لتنفيذ احكام القانون فقط
English Translation	This decision was made to implement the provisions of the law only.

F. The Lexicons

Our tool depends on the lexicons already built, where each lexicon is composed of the manually extracted features to be used by the lexicon-based tool. So in the following subsections we will present a brief summary about each of these lexicons.

1) Domain Lexicons

Domain features were extracted manually to be used by the classification process to identify automatically the domain of each evaluated Arabic review.

2) General Polarity Lexicons

Two lexicons were created to classify opinions. The first lexicon is for positive sentiment features which contain 2,404 positive features or sentiments. The second lexicon is for negative features or sentiments which contain 5,521 ones. These positive and negative features/terms sentiments were collected from the training dataset and there is a part added by translating an English sentiment lexicon presented in [36].

3) Domain-based Sentiment Lexicons

Two lexicons were built for every domain. One for Arabic positive opinions and the other for Arabic negative opinions.

4) Score (Weight) Lexicon

Polarity lexicons used in this study have a weighting score for each Arabic term/feature in these lexicons. Those weights were proposed by the authors of this study. The values of the weighting scores are in range of1 to 10, for both positive and negative features/terms, where 1 indicates that the feature/term is the weakest possible positive or negative feature/term, and 10 indicate that the feature/term is the strongest possible positive or negative feature/term.

G. The Classification Categories Set

This study is based on the syntactical features using the sentiment term frequencies to identify the subjectivity and the polarity of different Arabic reviews and comments. The weight scores of sentiment features are used to determine the polarity and the strength of each Arabic review and comment under consideration.

The sentiment features used in this study are terms extracted manually from the collected Arabic comments and reviews which correspond to documents in this field. Where *TF* (term frequency) refers to the number of times a specific term T_i occurs in D (Arabic comment/review). The weight of each sentiment feature is determined manually.

The tool depends on the frequency of positive and negative features/terms to identify the polarity of evaluated Arabic review. The evaluated Arabic review is considered positive when the frequency of positive terms/features in it exceeds the frequency of negative terms/features in the same Arabic review.

The tool considers the evaluated Arabic review as negative when the frequency of negative features/terms in it exceeds the frequency of positive terms/features in the same Arabic review. The evaluated Arabic review is considered by the tool as neutral if the frequency of positive terms/features in the evaluated Arabic review is equal to the frequency of negative features/terms. The scores in polarity lexicons are used by the tool to determine the strength on each inputted Arabic review.

The following paragraphs show the pseudo code used in our tool to identify different taxonomies of Arabic reviews. Our tool considers any evaluated Arabic review free from any of the terms in the built polarity lexicons as a fact.

Review's Opinion Determination:

If (*No. of Positive terms in a review* > 0) or (*No. of Negative terms in a review* > 0) *then*
    Review is opinion
Review's Fact Determination:

If (*No. of Positive terms in a sentiment* = 0) && (*No. of Negative terms in a sentiment* = 0) then
    Review is Fact.
Let

$P = \{Pos, Neg, NU, U\}$, where *Pos*: Positive review, *Neg*: Negative review, *NU*: Neutral review, and *U*: undetermined.

$$\sum_{i=1}^{n} Pos - TF_i : \text{Total number of positive terms in the}$$
evaluated review.

$$\sum_{i=1}^{n} Neg - TF_i : \text{Total number of negative terms in the}$$
evaluated review.

Then the polarity is determined as shown in the following pseudo code:

Review's Positive Polarity Determination:

$$If \left(\sum_{i=1}^{n} Pos - TF_i \right) > \left(\sum_{i=1}^{n} Neg - TF_i \right) then$$
$$P \leftarrow Pos$$

$$P_{pos\%} = \frac{\sum_{i=1}^{n} Pos - TF_i}{\left(\sum_{i=1}^{n} Pos - TF_i \right) + \left(\sum_{i=1}^{n} Neg - TF_i \right)} \times 100$$

Review's Negative Polarity Determination:

$$If \left(\sum_{i=1}^{n} Pos - TF_i \right) < \left(\sum_{i=1}^{n} Neg - TF_i \right) then$$
$$P \leftarrow Neg$$

$$P_{neg\%} = \frac{\sum\limits_{i=1}^{n} Neg - TF_i}{\left(\sum\limits_{i=1}^{n} Pos - TF_i\right) + \left(\sum\limits_{i=1}^{n} Neg - TF_i\right)} \times 100$$

Review's Neutral Polarity Determination:

$$If \left(\sum\limits_{i=1}^{n} Pos - TF_i = \sum\limits_{i=1}^{n} Neg - TF_i = 0\right) then$$

$$P \leftarrow NU$$

Review's Undetermined Polarity Determination:

$$If \left(\sum\limits_{i=1}^{n} Pos - TF_i = \sum\limits_{i=1}^{n} Neg - TF_i \neq 0\right) then$$

$$P \leftarrow U$$

The tool second step is to compute the (strength/intensity) of each evaluated Arabic review with its polarity. The computation of the strength/intensity is based on emotions closeness to (sentiments/opinions) as shown in [3] and [37]. The following pseudo code shows different types of strength/intensity and the formulas used to compute them:

Let

$I = \{ SP, SN, WP, WN, NU, U \}$, *where SP: strong positive, SN: strong negative WP: weak positive, WN: weak negative, NU: neutral, U: undetermined.*

Max_Pos_Score: Max of the set of Pos-Weights of the positive sentiments.
Max_Neg_Score: Max of the set of Neg-Weights of the negative sentiments.

Then the polarity is determined as shown in the following pseudo code:

Review's Strong Positive Polarity Determination:
If ($Max\_Pos\_Score > Max\_Neg\_Score$) *and*
($Max\_Pos\_Score \geq 5$) *then*
$I \leftarrow SP$

Review's Weak Positive Polarity Determination:
If ($Max\_Pos\_Score > Max\_Neg\_Score$) *and*
($Max\_Pos\_Score < 5$) *then*
$I \leftarrow WP$

Review's Strong Negative Polarity Determination:
If ($Max\_Pos\_Score < Max\_Neg\_Score$) *and*
($Max\_Neg\_Score \geq 5$) *then*
$I \leftarrow SN$

Review's Weak Negative Polarity Determination:
If ($Max\_Pos\_Score < Max\_Neg\_Score$) *and*
($Max\_Neg\_Score < 5$) *then*
$I \leftarrow WN$

Review's Neutral Polarity Determination:
If ($Max\_Pos\_Score = Max\_Neg\_Score \neq 0$) *then*
$I \leftarrow NU$

Review's Undetermined Polarity Determination:
If ($Max\_Pos\_Score = 0$) *and* ($Max\_Neg\_Score = 0$) *then*
$I \leftarrow U$

Examples:

Consider the following Arabic excerpts from the places domain with its English translations. This example includes four Arabic reviews considered by the tool as: Positive, Negative, Undetermined, or Neutral.

Example 5:

Positive Review:

Arabic Comment	الخدمه في هذا الفندق اكثر من رائعه، انبسطت كثير على التصاميم داخل المبنى، جد جميل هالمكان ولو انه بعيد.
English Translation	The service in this hotel is more than wonderful; I really enjoy many designs inside the building, very beautiful place even that far.

So the algorithm identifies the above Arabic review from the places domain as a positive review since it has three positive terms and one negative term as shown in table X.

TABLE X.   MANUALLY EXTRACTED FEATURES FROM A POSITIVE ARABIC REVIEW RELATED TO PLACES

Arabic Sentiment Feature	بعيد	جميل	انبسطت	رائعه
English Translation	Far	Beautiful	Enjoyed	Wonderful
Polarity	Neg	Pos	Pos	Pos

Another example is provided in this section to show how our tool identifies an evaluated Arabic review as undetermined.

Example 6:

The Arabic comment shown in this example is selected from places domain. The tool identifies an Arabic comment/review as undetermined when the number of positive polarity features is equal to the number of negative polarity features.

Undetermined:

Arabic Comment	خدمتهم سيئه ولكن المكان جميل
English Translation	Their services are bad, but the place is beautiful

Table XI shows the essential two extracted Arabic features with their polarities and English translations. This table shows equality in the number of positive and negative Arabic features/terms extracted from the above Arabic comment. Our tool labeled the above Arabic comment as undetermined when the frequencies of opposite and negative polarities are equal.

TABLE XI.   MANUALLY EXTRACTED FEATURES FROM AN UNDETERMINED ARABIC REVIEW RELATED TO PLACES

Arabic Sentiment Feature	جميل	سيئه
English Translation	Beautiful	Bad
Polarity	Pos	Neg

The Arabic comment shown in example 7 is considered by the tool as a fact and not an opinion (subjectivity classification). Moreover, our tool identifies the same comment as neutral and not as a positive or negative (polarity classification).

Example 7:

The Arabic comment presented in this example is identified by the tool as a fact within subjectivity category, and as neutral within polarity category since it is free from any extracted feature.

Fact/Neutral:

| Arabic Comment	هذا الفندق يقع في شمال عمان
English Translation	This hotel is located north of Amman.

A number of Arabic reviews related to books domain are presented below to show how we can determine the strength and polarity using weight scores for the terms in the review.

Example 8:

Consider the following Arabic comment which belongs to domain of books. This review is characterized by using only positive features/terms, so it will be identified by the tool as a positive Arabic comment. This tool will search for the highest score (weight) in such cases, where the highest weight of the extracted features from the Arabic comment presented in this example is 10. If the value of the highest weight of features extracted from the Arabic comment exceeds 5 the tool will consider the strength of comment under consideration as strong, otherwise the strength of the comment will be considered weak.

In such cases, the tool will consider the above Arabic comment as a strong positive, since the highest weight shown in table XII is 10, which implicitly means that this comment is strong, and since all comment features are positive, it considered a strong positive comment.

This method to identify the strength of each Arabic review and comments is suggested by the authors of this study.
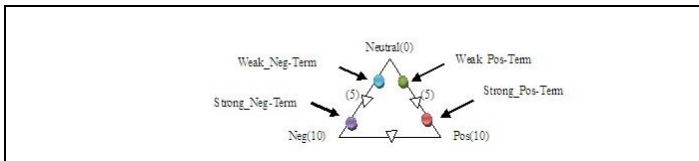
Strong Positive:

Arabic Comment	كتاب رائع ، محتواه جيد والمواضيع فيه مميزه
English Translation	A wonderful book, the content is good and the topics are distinctive

Three positive features are extracted from the above Arabic excerpt with their strength weights are shown in table XII.

TABLE XII. MANUALLY EXTRACTED FEATURES WITH THEIR STRENGTH WEIGHTS FROM AN ARABIC REVIEW RELATED TO BOOKS

Arabic Sentiment Feature	مميزه	جيد	رائع
English Translation	Distinctive	Good	Wonderful
Polarity (Weight)	Pos (w=10)	Pos (w=4)	Pos (w=10)

Let us consider the following Arabic review from the domain of books which considered by our tool as a weak positive review.

Example 9:

Consider the following Arabic review from the domain of books which considered as a weak positive by our tool, since it has only one weak positive feature/term so it will considered positive and it intensity will be considered weak since the weight of this feature/term is 4 which is less than 5, so it will be identified as weak.

Weak Positive:

Arabic Comment	كتاب جيد الى حد ما.
English Translation	A good book to some extent.

Table XIII exhibits the extracted feature from the above Arabic review about a certain book.

TABLE XIII. MANUALLY EXTRACTED FEATURE WITH IT STRENGTH WEIGHT FROM THE ABOVE EXAMPLE 7

Arabic Sentiment Feature	جيد
English Translation	Good
Polarity (Weight)	Pos (w=4)

Example 10:

Consider another Arabic review related to books domain, and considered by the tool as strong negative review. The following Arabic comment has only negative extracted features, so it will be considered a negative comment by the tool. In this case the tool has six negative weight, so the tool output is based on the highest score (weight) in such cases. Therefore in this example there are 3 weight values (4, 7, and 10), and since 10 is the highest our tool will identify the Arabic comment presented in this example as a strong negative.

Strong Negative:

Arabic Comment	كتاب ممل ولا يوجد ترابط بن المواضيع ، انا ما حبيته وما عجبني ،لا يستحق كل هالحكي لغته ضعيفه ،معلوماته سطحيه
English Translation	Boring book with no association between its topics. I do not like it, it is not worth mentioning, since its language is weak with superficial information.

TABLE XIV. MANUALLY EXTRACTED FEATURES WITH IT STRENGTH WEIGHTS FOR THE ABOVE EXAMPLE

Arabic Sentiment Features	سطحيه	ضعيفه	لا يستحق	ما عجبني	ما حبيته	ممل
English Translation	Superficial	Weak	Not worth	I dislike it	I do not like it	Boring
Polarity (Weight)	Neg (w = 4)	Neg (w=7)	Neg (w=7)	Neg (w=10)	Neg (w=10)	Neg (w = 7)

The tool uses the extracted features in table XIV to identify the above Arabic review as strong negative.

Consider another Arabic review related to books domain, and considered by the tool as a weak negative review.

Example 11:

It is usual in Arabic and English to face sentences which have words that used before or after the extracted features/terms and leads to reduce the weights of some words, so in the following Arabic comment the user uses the Arabic colloquial word (Somewhat, "شوي ") after the MSA word (Difficult, "صعبه") and this leads to a change in the strength of phrase, where the feature (Difficult, "صعبه") is saved in negative polarity lexicon and given a weight of 8, but these two terms are stored in negative polarity lexicon and the phrase (Somewhat difficult) given a weight equals to 4. This weight is considered by the tool as a weak negative review.

Weak Negative:

Arabic Comment	كتاب مفاهيمه صعبه شوي وتحتاج الى توضيح يا ريت الكاتب راعى مستويات القراء ، بس الصراحه محتاج تعديلات.
English Translation	A book with somewhat difficult concepts, and need to be clarified. It would be better if the writer took into account the levels of readers, but frankly it needs a revision.

The polarity and strength weight shown in table XV are used by the tool to identify the above review as weak negative review.

TABLE XV. MANUALLY EXTRACTED FEATURE WITH IT STRENGTH WEIGHT

Arabic Sentiment Feature	صعبه شوي
English Translation	Somewhat difficult
Polarity (Weight)	Neg (w = 4)

Example 12:

This example shows how the tool identifies an Arabic review as undetermined. Strength determining algorithm labeled Arabic review as undetermined when the values of high strength weights are equal as shown in the following sample review from books domain. This example is based on Algorithm 3 which is presented in section 4.3 of this study.

Undetermined:

Arabic Comment	كتاب مميز في طرح المواضيع ولكنه لا يعرض المواضيع بشكل مترابط وهذه سيئه في الكتاب.
English Translation	Distinguished book in presenting topics, but does not present topics coherently, and this is a disadvantage of the book.

The tool uses the polarity and strength in table 16 to identify the above review as undetermined.

TABLE XVI. MANUALLY EXTRACTED FEATURES WITH IT STRENGTH WEIGHTS FOR THE ABOVE EXAMPLE

Arabic Sentiment	سيئه	مميز
English Translation	Disadvantage	Distinguished
Polarity (Weight)	Neg (w = 10)	Pos (w = 10)

IV. ALGORITHMS

This section presents the pseudo code of the algorithms adopted in the opinion mining tool. The tool enables its users to input a single Arabic review/comment or a group of Arabic reviews/comments to identify their subjectivity (fact/opinion), polarity (Pos/ Neg/ Neut), and strength.

A. Subjectivity Algorithm

The following algorithm 1 is adopted by the tool to identify different Arabic reviews evaluated as facts or opinions.

Algorithm 1: CNSA-MSA-SAT "determining the subjectivity"
Input:
    R = Review/Document Text
    T = the set of the Opinion tokens
    PD = the set of Positive Sentiment Dictionary
    ND = the set of Negative Sentiment Dictionary
Output*:*
    R_1={F,O} where F is a Fact and O is an Opinion

Initialization:
$Pos\text{-}TF =0$ where Pos-TF: term frequency for Positive Sentiments
$Neg\text{-}TF= 0$ where Neg-TF: term frequency for Negative Sentiments

Begin
  1: For each ti $\in$ T do
  2:   Find ti $\in$ PD where ti $\in$ T
  3:   If (ti $\in$ PD) then
  4:   Pos-TF $\leftarrow$ Pos-TF + 1
  5:   End if
  6:   Find ti $\in$ ND where ti $\in$ T
  7:   If ti $\in$ ND then
  8:   Neg-TF $\leftarrow$ Neg-TF + 1
  9:   End if
  10: end for
  11: If (Pos-TF > 1) or (Neg-TF > 1) then
  12: $R_1 \leftarrow O$
  13:   Display R_1
  14: Else
  15: $R_1 \leftarrow F$
  16:   Display R_1
  17: end if
End

B. Polarity Algorithm

The following algorithm 2 is adopted by the tool to determine the polarity of evaluated Arabic reviews regardless whether they are using MSA or colloquial Arabic. Each evaluated Arabic review is considered either as positive, negative, neutral or undetermined review. This algorithm is used in [38] to build our tool that presented in this study.

Algorithm 2: CNSA-MSA-SAT" determining the polarity"
Input:
R: Review/Document Text
T: the set of the Opinion tokens
PD: the set of Positive Sentiment Dictionary
ND: the set of Negative Sentiment Dictionary
Output:
P={Pos, Neg, NU, U}, where Pos: Positive, Neg: Negative, NU: Neutral, U:Undetermined

---

$$P_{pos\%} = \frac{\sum_{i=1}^{n} Pos - TF_i}{\left(\sum_{i=1}^{n} Pos - TF_i\right) + \left(\sum_{i=1}^{n} Neg - TF_i\right)} \times 100$$

$$P_{neg\%} = \frac{\sum_{i=1}^{n} Neg - TF_i}{\left(\sum_{i=1}^{n} Pos - TF_i\right) + \left(\sum_{i=1}^{n} Neg - TF_i\right)} \times 100$$

Initialization:

Pos-TF = 0, where Pos-TF is the term frequency for positive sentiments

Neg-TF = 0, where Neg-TF is the term frequency for negative sentiments

Begin

1:   *For each $t_i \in T$ do*
2:    *Search for t_i in PD where $t_i \in T$*
3:    *If $t_i \in PD$ then*
4:     *Pos-TF ← Pos-TF +1*
5:    *Else*
6:     *Search for t_i in ND where $t_i \in T$*
7:     *If $t_i \in ND$ then*
8:      *Neg-TF ← Neg-TF + 1*
9:   *End For*
10: *If (Pos-TF ≥ 2) and (Pos-TF > Neg-TF) then*
11:   *P ← Pos*
12:   *Return $P_{pos\%}$*
13: *End If*
14: *If (Neg-TF ≥ 2) and (Pos-TF < Neg-TF) then*
15:   *P ← Neg*
16:   *Return $P_{neg\%}$*
17: *End If*
18: *If (Pos-TF = Neg-TF) and (Pos-TF = 0) then*
19:   *P ← NU*
20:   *Display P*
21: *End If*
22: *If (Pos-TF = Neg-TF) and (Pos-TF ≠ 0) then*
23:   *P ← U*
24:   *Display P*
25: *End If*

End

C. Strength/Intensity Algorithm

This section presents algorithm 3 that used to determine the Strength/Intensity of evaluated Arabic reviews this tool. Each evaluated Arabic review is considered either as: strong positive, strong negative, weak positive, weak negative, neutral or an undetermined review.

Algorithm 3: CNSA-MSA-SAT" determining the intensity and the polarity depend on the weight "

Input:

    R = Review/Document Text
    T = the set of the Review tokens
    PD = the set of Positive Sentiment Dictionary
    ND = the set of Negative Sentiment Dictionary
    Pos-Weight = The value of Positive Weight
    Neg-Weight = The value of Positive Weight

Output:

    I = {SP,SN,WP,WN,NT,U} where SP: strong positive, SN: strong negative WP: weak positive, WN: weak negative, NT: neutral, U: undetermined

Max_Pos_score = Max of the set of Pos-Weight of the positive sentiments

Max_Neg_Score = Max of the set of Neg-Weight of the negative sentiments

Begin:

1:   *For each ti Є T do*
2:    *If ti Є PD where ti Є T*
3:    *Find Pos-Weight*
4:   *Else*
5:    *Pos-Weight = 0*
6:   *End For*
7:   *Find Max_Pos_score*
8:   *For each ti Є T do*
9:    *If ti Є ND where ti Є T*
10:   *Find Neg-Weight*
11: *Else*
12:   *Neg-Weight = 0*
13: *End for*
14: *Find Max_neg_Score*
15: *If (Max_Pos_Score > Max_Neg_Score and Max_Pos_Score ≥ 5) then*
16:   *I ← SP*
17:   *Return I*
18: *End If*
19: *If (Max_Neg_Score > Max_Pos_Score and Max_Neg_Score ≥ 5) then*
20:   *I ← SN*
21:   *Return I*
22: *End If*
23: *If (Max_Pos_Score > Max_Neg_Score and Max_Pos_Score < 5) then*
24:   *I ← WP*
25:   *Return I*
26: *End If*
27: *If (Max_Neg_Score > Max_Pos_Score and Max_Neg_Score < 5) then*
28:   *I←WN*
29:   *Return I*
30: *End If*
31: *If (Max_Pos_Score = 0 and Max_Neg_Score = 0) then*
32:   *I←NU*
33:   *Return I*
34: *End If*
35: *If (Max_Neg_Score = Max_Pos_Score and Max_Pos_Score ≠ 0) then*
36:   *I ← U*
37:   *Return I*
38: *End If*

End

V. EXPERIMENTAL RESULTS

This section presents and discusses experimental results. The conducted tests aim to evaluate the effectiveness of the developed opinion mining tool to identify domains, subjectivity, polarity and strength of evaluated Arabic reviews. The results in this section are presented in the following three subsections. The first subsection presents the results of the tests related to subjectivity classification, the second subsection presents the results related polarity, and the third subsection presents results related to the identification of the intensity of each evaluated Arabic review to the tool.

In the experiments for all classifiers, we used 66% of the dataset as a Training Dataset and 34% as a Testing Dataset.

We used the following four metrics to evaluate the quality of the tool in terms of opinion decision:

Accuracy: Is the degree of closeness that a measured value represents the correct value.

The Accuracy is defined by the formula (5.1):

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad .... (5.1)$$

Where *TP* is a true positive rate, *FP* is a false positive rate, *TN* is a true negative rate, and *FN* is a false negative rate [36].

Error: Is the degree of closeness that a measured value represents the incorrect value [39].

The formulas of the other two performance metrics (Recall and Precision) are shown next.

The Recall is defined by formula (5.2) [40]:

$$Recall_i = \frac{TP}{TP + FN} \quad .... (5.2)$$

The Precision is defined by the formula (5.3) [40]:

$$Precision_i = \frac{TP}{TP + FP} \quad .... (5.3)$$

where *TP* is the number of documents correctly classified as belonging to a class *i* ("true positive"), *FP* is the number of documents falsely classified as belonging to a class *i* ("false positive") and *FN* is the number of documents falsely classified as not belonging to a class *i* ("false negative") [37].

A. Subjectivity Results

This subsection presents the results of the tests conducted on the tool to evaluate its effectiveness to identify Arabic facts and opinions. A Naive Bayes Classifier proves it is more effective than others classification algorithms such as Decision Tree, K-NN, SVM to identify Arabic facts and opinions. Therefore it's adopted and used. The overall accuracy shown in table XVII is 93.9%. Furthermore, table XVII presents recall and precision values according to 5.2 and 5.3 formulas.

TABLE XVII. NAIVE BAYES SUBJECTIVITY RESULTS

Class	Accuracy	Error	Precision	Recall
Opinion	-	-	0.96	0.96
Fact	-	-	0.85	0.85
Dataset	93.9%	6.01%	0.93	0.93

B. Polarity Evaluation Result

This subsection presents an evaluation to accuracy of the tool to identify the polarity of each evaluated Arabic review. A K-NN Classifier proves it is more effective than others classification algorithms such as Decision Tree, Naive Bayes, and SVM to identify the polarities of different Arabic reviews. The overall accuracy shown in table XVIII is 90%. Table XVIII presents recall and precision values according to 5.2 and 5.3 formulas.

TABLE XVIII. K-NN POLARITY RESULTS

Class	Accuracy	Error	Precision	Recall
Positive	-	-	0.8	1
Negative	-	-	1	0.3
Neutral	-	-	1	1
Dataset	90%	10%	0.9	0.9

Table XVIII shows that the effectiveness of tool to identify neutral Arabic reviews is optimum

C. Intensity Evaluation Result

This subsection presents the results of the tests conducted on the tool to evaluate its effectiveness to identify the intensity of different Arabic reviews. Once again Naive Bayes Classifier proves it is more effective than others classification algorithms such as Decision Tree, K-NN, SVM to identify the strength of the evaluated Arabic review. Therefore Naive Bayes is adopted and used. The overall accuracy shown in table XIX is 96.6%. Table XIX presents recall and precision values according to 5.2 and 5.3 formulas.

TABLE XIX. NAÏVE BAYES INTENSITY RESULTS

Class	Accuracy	Error	Precision	Recall
Dataset	96.9%	3.1%	0.95	0.97

VI. CONCLUSION AND FUTURE WORK

This study presents a basic tool which can be used to analyze Arabic reviews and comments regardless of the type of the Arabic language (MSA or Colloquial) they used. In order to evaluate the proposed tool, we need a standard dataset to test its effectiveness. We found that there is no standard dataset to be used. Therefore we collected Arabic reviews and comments. The collected Arabic reviews use only MSA and the first four Arabic Vernaculars presented in the section 1: Arabian Peninsula Arabic (Khaliji Arabic), Mesopotamian Arabic, Syro-Palestinian Arabic, and Egyptian Arabic. The proposed tool presented in this study is a lexicon-based tool. The collection of Arabic comments and review phase is followed by lexicon creation phase. The lexicons used in this study are manually created, since they have manually extracted features, terms, and phrases from the collected reviews and comments. The tool is capable to identify the polarity, subjectivity, and strength/intensity of each evaluated Arabic review and comment. This study is based on 18 lexicons which are built manually. Two general purpose lexicons were built to be used to identify polarity, and 16 domain-specific lexicons were built to be used to identify the polarity with eight different domains: Technology, Books, Education, Movies, Places, Politics, Products, and Society. The last phase of this study includes an evaluation to the effectiveness of the built tool. The evaluation of this tool yields: a 93.9% accuracy to classify the evaluated Arabic comments and reviews into their proper domains, a 90% accuracy to identify the real polarity of the evaluated Arabic comments and reviews, and a 96.9% accuracy to identify the strength/intensity of the evaluated Arabic comments and reviews. Tests on the tool reveal the reasons behind errors.

The main reasons behind these errors are summarized by the use of spam reviews, spelling mistakes, and short comment length (One word).

We plan to enhance and extend this study by using a larger dataset which has more Arabic comments and reviews written in a wider range of Arabic vernaculars. This tool is incapable to deal with emoticons, chat language, Arabizi, so we plan to enhance this tool to be able to deal with these inputs. Future plans include adopting semantic techniques to identify polarity, subjectivity, and strength/intensity. Our future plans include investigating the automatic creation of Arabic lexicons to be used by sentiment analysis tool.

REFERENCES

[1] Arabic Speaking Internet Users and Population Statistics. http://www.internetworldstats.com/stats19.htm (2012, accessed January 2013).

[2] Varieties of Arabic. http://en.wikipedia.org/wiki/Varieties_of_Arabic (2013, accessed May 2013).

[3] B. Liu, "Sentiment Analysis and Subjectivity," In Indurkhya N and Damerau F J (Eds) Handbook of Natural Language Processing, Chapman and Hall/CRC, Second Edition, 2010.

[4] B. Liu. "Sentiment Analysis and Opinion Mining. California," US: Morgan & Claypool Publishers, 2012.

[5] T. Nasukawa, y. Yi, "Sentiment analysis: capturing favorability using natural language processin," In: Proceedings of the 2nd international conference on Knowledge capture(K-CAP '03), New York, NY, USA. pp. 70-77, 2003.

[6] D. Kushal, S. Lawrence,D. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews, " In: Proceedings of International Conference on World Wide Web (WWW-2003); 2003.

[7] Social networking service - Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Social_networking_service (2013, accessed February 2013).

[8] Alexa - The Web Information Company. http://www.alexa.com (2013, accessed February 2013).

[9] Facebook Statistics and Metrics by Countries - Social bakers. http://www.socialbakers.com/facebook-statistics (2013, accessed February 2013).

[10] B. Hammo, H. Abu-Salem,S. Lytinen, "QARAB: A Question Answering System to Support the Arabic Language," In: Annual Meeting of the ACL Proceedings of the ACL-02 Workshop on Computational Approaches to Semitic, pp. 1-11. 2001.

[11] M. Bautin, L. Vijayarenu,S. Skiena, "International Sentiment Analysis for News and Blogs," In: Proceedings of the International Conference on Weblogs and Social Media, Seattle, Washington, 2008.

[12] M. Rushdi-Saleh, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López,J. M. Perea-Ortega, "Bilingual Experiments with an Arabic-English Corpus for Opinion Mining. Language," In: Proceedings of Recent Advances in Natural Language Processing, Hissar, Bulgaria. pp. 740-745, 2011.

[13] Y. Almas,K. Ahmad, "A note on extracting 'sentiments' in financial news in English, Arabic & Urdu," In: The Second Workshop on Computation, al Approaches to Arabic Script-based Languages, Linguistic Society of America 2007 Linguistic Institute, Stanford University, Stanford, California., Linguistic Society of America, pp. 1–12, 2007.

[14] A. El-Halees, "Arabic Opinion Mining Using Combined Classification Approach," In: Proceedings of the International Arab Conference on Information Technology, Zarqa, Jordan, 2011.

[15] B. Pang, L. Lee, S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," In: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, Philadelphia, Penn. pp. 79-86, 2002.

[16] J. Yi, T. Nasukawa, R. Bunescu, W.Niblack, "Sentiment Analyzer: Extracting Sentiments about a given Topic using Natural Language Processing Techniques," In: Proceedings of the Third IEEE International Conference on Data Mining IEEE Computer Society, pp. 427-434, 2003.

[17] S-K. Kim, E. Hovy, "Determining the sentiment of opinions," In: Proceedings of the 20th international conference on computational linguistics (COLING 2004), Geneva, Switzerland. pp. 1367–1373, 2004.

[18] M. Elhawary,M. Elfeky, "Mining Arabic Business Reviews," In: Proceedings of the2010 IEEE International Conference on Data Mining Workshops; pp. 1108-1113, 2010.

[19] M. Abdul-Mageed, M. Diab, M. Korayem, "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, Portland, Oregon, USA. pp. 587-591, 2011.

[20] T. Wilson, J. Wiebe, P.Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," In: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing(HLT '05), Stroudsburg, PA, USA, pp. 347-354, 2005.

[21] T. Wilson, J. Wiebe,R. Hwa, "Recognizing Strong and Weak Opinion Clauses," Computational Intelligence Journal, vol. 22, pp. 73-99, 2006.

[22] A. Abbasi, H. Chen,A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums," ACM Transactions on Information Systems, vol. 26, no. 3, pp. 1-25, 2008.

[23] G. Paltoglou, M. Thelwall, "A study of Information Retrieval weighting schemes for sentiment analysis," In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL '10) Association for Computational Linguistics, Stroudsburg, PA, USA. pp. 1386-1395, 2010.

[24] T. Inui, M. Yamamoto "Applying Sentiment-oriented Sentence Filtering to Multilingual Review Classification," In: Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), IJCNLP, Chiang Mai, Thailand, pp. 51–58, 2011.

[25] C. Banea, R. Mihalcea, J. Wiebe, S. Hassan, "Multilingual subjectivity analysis using machine translation," In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 127–135, 2008.

[26] C. Banea,R. Mihalcea, J. Wiebe, "Multilingual subjectivity: Are more languages better?," In: Proceedings of the 23rd International Conference on Computational Linguistics, pp.28–36, 2010.

[27] A. Balahur, A. Montoyo, "A feature dependent method for opinion mining and classification," In: Proceedings of the IEEE international conference on Natural Language Processing and Knowledge Engineering(NLP-KE '08), Beijing, China, pp. 1-7, 2008.

[28] Y. Choi, Y. Kim,S-H. Myaeng, "Domain-specific Sentiment Analysis using Contextual Feature Generation," In: Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement(TSA'09), Hong Kong – China, pp. 37-44, 2009.

[29] A. Ortiz, C. Hernández,R. García, "Domain-neutral, Linguistically-motivated Sentiment Analysis: A Performance Evaluation, Evaluación de unsistema de análisis de sentimiento basado en conocimiento,"Machine Learning, pp. 361-369, 2007.

[30] A. Al-Subaihin, H. Al-Khalifa,A. Al-Salman, "A Proposed Sentiment Analysis Tool for Modern Arabic Using Human-Based Computing," In: Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services(iiWAS '11), New York, NY, USA. pp. 543-546, 2011.

[31] M. Al-Kabi, N. Al-Qudah, I. Alsmadi, M. Dabour M., H. Wahsheh, (2013). "Arabic / English Sentiment Analysis: An Empirical Study," The 4th International Conference on Information and Communication Systems (ICICS 2013), Irbid, Jordan, pp. 1-6, 2013.

[32] R. Khasawneh, H. Wahsheh, M. AL-Kabi, I. Alsmadi,"Sentiment Analysis of Arabic Social Media Content: A Comparative Study," The 8th International Conference for Internet Technology and Secured Transactions (ICITST-2013), London, UK., pp. 101-106, 2013.

[33] M. AL-Kabi, N. Abdulla, M. Al-Ayyoub,"An Analytical study of Arabic Sentiments: Maktoob Case Study," The 8th International Conference for

Internet Technology and Secured Transactions (ICITST-2013), London, UK, pp. 89-94, 2013.

[34] Y. Fang, S. Parthasarathy,F. Schwartz,"Using Clustering to Boost Text Classification," In: Proceedings Of the IEEE International Conference on Data Mining, California, USA, pp. 123-127, 2001.

[35] J. Han,M. Kamber,"Data Mining: Concepts and Techniques, second edition," Morgan Kaufmann Publishers, CA, San Francisco, 2006.

[36] Opinion Mining, Sentiment Analysis, and Opinion Spam Detection. http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html (2013, accessed January 2013).

[37] A. Chaudhuri, "Emotion and Reason in Consumer Behavior. Amsterdam," Elsevier Butterworth-Heinemann, 2006.

[38] M. Al-Kabi, A. Gigieh, I. Alsmadi, H. Wahsheh,M. Haidar,"An Opinion Analysis Tool for Colloquial and Standard Arabic," In: Proceedings of the fourth International Conference on Information and Communication Systems (ICICS 2013), Irbid, Jordan, pp. 1-5, 2013.

[39] I. Witten,E.Frank, "Data Mining: Practica Machine Learning Tools and Techniques," Morgan Kaufmann Series in Data Management Systems, second edition, Morgan Kaufmann (MK), 2005.

[40] G. Paltoglou,M. Thelwall,"Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 3, no. 4, pp. 1-19, 2012.

AUTHORS PROFILE

Mohammed Naji Al-Kabi obtained his Ph.D. degree in Mathematics from the University of Lodz/Poland (2001), his master's degree in Computer Science from the University of Baghdad/Iraq (1989), and his bachelor degree in statistics from the University of Baghdad/Iraq (1981). Mohammed Naji AL-Kabi is an assistant Professor in the Faculty of Sciences and IT, at Zarqa University. Prior to joining Zarqa University, he worked many years at Yarmouk University in Jordan, Nahrain University and Mustanserya University in Iraq. He also worked as a part time lecturer at Jordan University of Science and Technology, and Sunderland University. AL-Kabi's research interests include Information Retrieval, Web search engines, Data Mining, Social media, Natural Language Processing and Software Engineering. He is the author of more than 66 peer reviewed articles in these topics. His teaching interests focus on information retrieval, Web programming, data mining, DBMS (ORACLE & MS Access).

Amal H. Gigieh got a Master Degree in Computer Information System CIS (2012) from Yarmouk University-Jordan, the bachelor degree in Information Technology IT (2003) from the AL-Balqa' Applied University-Jordan, worked as a computer lab supervisor and a computer practical courses lecturer in AL-Balqa' applied university. Gigieh also worked as online exams manager in the same univesity. Her research in Opinion Analysis and Mining.

Izzat Alsmadi. An associate professor in software engineering. Born in Jordan 1972, Izzat Alsmadi has his master and PhD in software engineering from North Dakota State University (NDSU), Fargo, USA in the years 2006 and 2008 respectively. His main areas of research include: software engineering, testing, metrics, and information retrieval.

Heider Wahsheh. Born in Jordan, in August 1987, he obtained his Master degree in Computer Information Systems (CIS) from Yarmouk University, Jordan, 2012.Since 2013 Mr. Wahsheh starts working as a lecturer in the college of Computer Science at King Khalid University, Saudi Arabia. His research interests include: Information Retrieval, Data Mining, and Mobile Agent Systems.

Mohamad Haidar. Born in Jordan, in February 23 1991. Mr. Haidar obtained his bachelor degree in Computer Information Systems (CIS) from Yarmouk University, Jordan in 2012. Now, he works in Brandtologie company (Amman – Jordan) as a Web developer.