The Islamic University–Gaza
Research and Postgraduate Affairs
Faculty of Information Technology
Master of Information Technology

الجامعـــــــة الإســـــلاميـة – غـزة
شئون البحث العلمي والدراسات العليا
كـليــــة تكنولوجيا المعلومات
ماجستير تكنولوجيا المعلومات

Feature Based Approach in Arabic Opinion Mining Using Ontology

التنقيب عن الآراء العربية باستخدام الأنتولوجيا بالاعتماد على المستوى

Ahmed M. I. AlAsmar

Supervised by

Alaa M. Al Halees
Prof.Dr. of Computer Science

**A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Information Technology**

July/2016

إقـــــرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان:

Feature Based Approach in Arabic Opinion Mining Using Ontology

التنقيب عن الآراء العربية باستخدام الأنتولوجيا بالاعتماد على المستوى

أقر بأن ما اشتملت عليه هذه الرسالة إنما هو نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيثما ورد، وأن هذه الرسالة ككل أو أي جزء منها لم يقدم من قبل الاخرين لنيل درجة أو لقب علمي أو بحثي لدى أي مؤسسة تعليمية أو بحثية أخرى.

Declaration

I understand the nature of plagiarism, and I am aware of the University's policy on this.

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted by others elsewhere for any other degree or qualification.

Student's name:	أحمد محمد الأسمر	اسم الطالب:
Signature:		التوقيع:
Date:	07/06/2016	التاريخ:

Abstract

With the rapid increase in the volume of Arabic reviews that use applications such as online review sites, blogs, forums, social networking, and so forth, comes at an increasing demand for Arabic opinion mining techniques. In Arabic language, researchs in this area is progressing at a very slow pace compared to that being carried out in English and other languages.

In this thesis, we highlight two problems for Arabic opinion mining technique: firstly, when analyzing review having different features with diverse opinion strengths. It considers all features extracted from the reviews to be equally important in failing to determine the proper polarity of the review and makes the review's sentiment classification less accurate. Secondly, the opinion summary for each feature doesn't consider the sub-features that represented it and makes the feature-based summary is incomplete. This research presents a technique using ontology that work at feature level classification to classify Arabic user generated reviews by identifying the important features from the review based on level of these features on the ontology tree and to generate an opinion summary for each feature in the whole corpus by identifiying the opinion of its sub-feature terms in the ontology. To evaluate our work, we use public datasets which are hotels and books datasets. We use accuracy, recall, precision, f-measure metrics to evaluate the performance and compare the results with other supervised or unsupervised techniques. Also, subjective evaluation is used in our method to demonstrate the effectiveness of feature and opinion extraction process and summarization. We show that our method improves the performance compared with other opinion mining classification techniques, obtaining 78.83% f-measure in hotel domain and 79.18% in book domain. Furthermore, subjective evaluation shows the effectiveness of our method by obtaining an average f-measure of 84.62% in hotel domain and 86.31% in book domain.

Keywords: Arabic Opinion Mining, Sentiment Classification, Feature level opinion mining, Ontology, Opinion Mining Summarization.

Abstract in Arabic

الملخص

مع الزيادة السريعة في حجم الآراء(التعليقات) العربية التي تستخدم تطبيقات مثل مواقع التعليقات على الأنترنت أو المدونات أو المنتديات أو شبكات التواصل الاجتماعي و غيره، فان هناك حاجة ماسة لإيجاد تقنيات للتنقيب عن الآراء العربية. والأبحاث في هذا المجال باللغة العربية يسير بوتيرة بطيئة جدا مقارنة مع اللغة الإنجليزية وغيرها من اللغات.

في هذه الأطروحة، نسلط الضوء على مشكلتين في تقنيات التنقيب عن الأراء باللغة العربية: أولها, عند تحليل رأي يحتوي على مكونات متعددة ومتفاوتة في الأهمية. فهي تعتبر ان كافة المكونات المستخرجة من الرأي لها نفس القدر من الأهمية وذلك يؤدي الى عدم الدقة في تحديد القطبية السليمة للرأي وتجعل من تصنيف المشاعر و تحليل الأراء أقل دقة. ثانيا, الملخص الناتج لكل مكون لم يأخذ بعين الاعتبار المكونات المكونة لهذا المكون في الانتولوجيا مما يجعل الملخص العام للمكون غير كامل. يقدم هذا البحث تقنية تستخدم الانتولوجيا تعتمد على مستوى المكون لتصنيف الأراء التي يكتبها المستخدم العربي من خلال تحديد المكونات الهامة الموجودة فيه من خلال التعرف على مستوى هذا المكون في شجرة الانتولوجيا ويتم عمل ملخص لكل مكون وذلك بتحديد آراء المكونات المكونة لها من خلال الانتولوجيا. و من اجل تقييم عملنا فإننا نستخدم مجموعات البيانات العامة التي هي الفنادق والكتب. ولتقييم الأداء ومقارنة النتائج مع تقنيات أخرى فأننا نستخدم مقياس–F. ونستخدم كذلك التقييم الشخصي في طريقتنا لإثبات فعالية استخراج المكون والرأي المتعلق بها. وتبين الدراسة أن طريقتنا تحسن الأداء مقارنة مع طرق تصنيف الأراء الأخرى، والحصول على 78.83٪ على مقياس–F في مجال الفنادق و79.18٪ في مجال الكتب. وعلاوة على ذلك، يظهر التقييم الشخصي فعالية طريقتنا عن طريق الحصول على متوسط مقياس–F قدره من 84.62٪ في مجال الفنادق و86.31٪ في مجال الكتب.

كلمات مفتاحية: التنقيب عن الآراء باللغة العربية ، تصنيف المشاعر، التنقيب عن الآراء بالإعتماد على المستوى، الانتولوجيا، تلخيص الأراء.

Dedication

To my beloved father,

To my beloved mother,

To my dear wife,

To my sons,

To brothers and sisters,

To my best friends,

To those who gave me all kinds of support,

To all, I dedicate this work.

Acknowledgment

Praise is to Allah, the Almighty for having guided me at every stage of my life.

Many thanks and sincere gratefulness go to my supervisor Dr. Alaa El-Halees, without his valuable guide, assistance, and continuous follow-up; this research would never have been.

In addition, I would like to extend my thanks to the academic staff of the Faculty ofInformation Technology who helped me during my master 's study and taught me different courses.

I also extend my thanks to my parents for their support during my course studies and during my thesis work.

Last but not least,I am greatly indebted to my wife for her support during my course studies and during my thesis work.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ArSenL	Arabic Sentiment Lexicon
AWN	Arabic WordNet
EWGA	entropy weighted genetic algorithm
GIBC	General Inquirer Based Classifier
IE	Information Extraction.
K-NN	K-Nearest Neighbour
LB	Lexicon Base.
ML	Machine Learning.
MSA	modern Standard Arabic
NB	Naïve Bayes
OBPRM	Ontology Based Product Review Miner
OP	Overall Polarity
OWL	Web Ontology Language
POS	Part-of-Speech
PWN	Princeton Wordnet
RBC	Rule Based Classifier
SBC	Statistic Based Classifier
SO	Semantic Orientation
SQL	Structured Query Language
SVM	Support Vector Machine
Tf	term frequency
tf-idf	term frequency–inverse document frequency

Chapter 1

Introduction

Chapter 1
Introduction

This chapter is an introduction to the thesis, first it gives a brief description of opinion mining, opinion mining in Arabic language and then, description about opinion mining level and approaches. In addition, it states the thesis problem, the research objectives, the significance of the thesis, the scope and limitation of the thesis work, and the research methodology.

1.1 Opinion Mining

In the last decade, the number of Internet users has increased significantly. This increase can be seen as the result of the technologies that facilitated the widespread of the Internet, along with the various online services such as online review sites, forums, blogs, social networking sites, and others (Al Shboul, Al-Ayyouby, & Jararwehy, 2015). The exposure of people to these online services allowed them to express their feelings and emotions regarding the provided services or in reaction to some subject in their lives. Furthermore, organizations expolit the Internet to collect people’s opinions about almost all the subjects that concern them through easing the process of getting feedback or by collecting what people are feeling from the various public websites (Al Shboul et al., 2015). After the collection of the raw unstructured data containing these expressions, some processing must be needed to analyze the peoples' opinions. As a result, the field of opinion mining has emerged.

Opinion mining is concerned with analyzing the opinions of a particular matter expressed by users in the form of natural language that appear in a series of texts. The opinion mining process makes it possible to figure out whether a user’s opinion is positive, negative or neutral, and how score it is (Zhao & Li, 2009). In general, opinion mining aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document.

1.2 Opinion Mining in Arabic Language

Most researchs in the opinion mining field has focused on English texts, but little in other languages like Arabic, despite the fact that Arabic is one of the top ten languages used on the Internet, and is spoken by hundreds of millions of people (Rushdi et al., 2011). Choosing to work in opinion mining with the Arabic language is due to several challenges. Firstly, most reviewers express their opinion using slang instead of modern standard Arabic (MSA). Secondly, the complexity of the language in regards to both the morphology and the structure has created a lot of challenges which result in very limited tools currently available for the aim of opinion mining (Farra, Challita, Assi, & Hajj, 2010).

1.3 Opinion Mining Levels

Opinion Mining can be done at any one of the three levels: the document level, sentence level, or feature level (Kaur & Duhan, 2015).

1.3.1 Document level

Document level classifies the whole document as positive, negative or neutral and commonly known as document-level sentiment classification.

1.3.2 Sentence level

Sentence level classifies the sentences as positive, negative or neutral commonly known as sentence-level sentiment classification.

1.3.3 Feature level

Feature level classifies sentences/documents as positive, negative or neutral based on the aspects of those sentences/documents commonly known as aspect-level sentiment classification.

For reviews, we should discover what exactly people liked and did not like. Both the document level and sentence level analysis do not discover what exactly people liked or not (B. Liu, 2010). Therefore, this research concentrates at the feature-level, which is little work in this field in Arabic opinion mining.

1.4 Opinion Mining Approaches

Many approaches are used in opinion mining classification such as: machine learning approach, lexicon based approach, combined approach, ontology based opinion mining approach, and feature based summary approach.

1.4.1 Machine Learning Approach (ML)

ML approach is typically a supervised approach in which a set of data labeled with its class, such as positive, negative and neutral is represented by feature vectors. Then, these vectors are used by the classifier as a training data inferring that a combination of specific features yields a specific class using one of the supervised classification algorithm (Morsy, 2011). Examples of classification algorithms are Support Vector Machine (SVM), Naïve Bayesian Classifier, Maximum Entropy (Han, Kamber, & Pei, 2011).

1.4.2 Lexicon Based Approach (LB)

The LB approach is an unsupervised approach in which a sentiment lexicon is created with each word having its weight as a number indicating its class. Then, this lexicon is used to extract all sentiment words from the sentence and sum up their polarities to determine if the sentence has an overall positive or negative sentiment in addition to its intensity whether they hold strong or weak intensity (Morsy, 2011).

1.4.3 Combined Approach

In the combination approach, it uses both lexicon based and machine learning approach. In LB based approach, it takes unannotated documents and identifies all opinion words and phrases (using negations when needed). Then aggregate these words to give a sentiment (positive or negative) to the document. Then use the ML approach to classify as many documents as possible that remain from the LB approach (Kim, Ganesan, Sondhi, & Zhai, 2011) .

1.4.4 Feature-Based Summarization Approach

Feature-based summarization involves generating opinion summaries around a set of features. These features are usually arbitrary topics that are considered important in the text being summarized. In general, feature-based summarization is made up of three distinct steps: feature identification, sentiment prediction and summary generation.

1.4.5 Ontology Based Opinion Mining Approach

Ontologies provides a formal, structured knowledge representation with the advantage of being reusable. However, it works using domain ontologies exploiting the ontology as a taxonomy by using relations between concepts and also provides a common vocabulary for a domain. Using ontology in opinion mining improves the performance of feature based identification systems by providing the structuring of the features and extraction of features. (Pang & Lee, 2008).

In this research, we use ontology in opinion mining classification to get several advantages which are:

- **Structuring of features:** Ontologies are tools that provide a lot of semantic information. They help to define concepts, relationships, and entities that describe a domain with an unlimited number of terms**.**
- **Extraction of features:** Relationship between concepts and lexical information can be used to extract explicit features.
- **Determine the important features:** The concepts present on the top of the ontology have more importance than other concepts that exist in the bottom of the top one. For example, shown in figure (1.1) we show the ontology tree for camera domains with 4 levels. The first level is the root of the camera domain, and the second level has more important features concerning the camera domain like the picture "الصورة" and video "الفيديو" etc., And the third level has features less important in camera domain, such as the accessories "الاكسسوارات", compression "الضغط" etc., And so on.

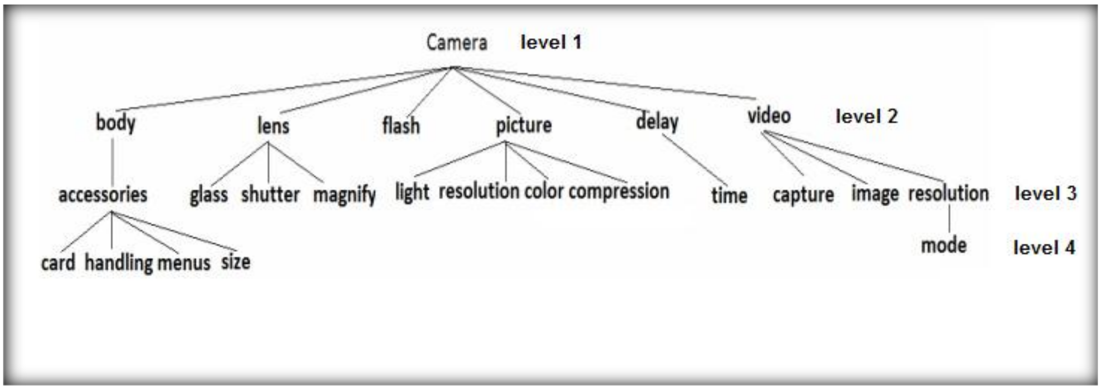

Figure (1.1): Sample ontology for camera domain Mukherjee and Joshi (2013).

- **Generate feature-based summary:** We identify the opinion of each feature in the whole corpus by identifying the opinion of its sub-class terms in the ontology.

1.5 Problem Statement

First problem, opinion mining techniques do not work accurately for review having different features with diverse opinion strengths, it considers all features extracted from the review to be equally important in failing to determine the proper polarity of the review. Second problem, the opinion summary generation for each feature doesn’t consider the sub-features that represent it in the ontology and makes the feature-based summary is incomplete. In this research, we use ontology structure to determine the important feature in the review and to generate an opinion summary for each feature.

1.6 Objectives

1.6.1 Main Objective

The main objective of our work is to provide technique that improve the performance of Arabic opinion mining classification technique by using ontology to determine the important features for the review having different features with diverse opinion strengths and exploit the important features extracted to determine the proper polarity of the review. Also, to generate opinion summary for each feature in the whole corpus by identifiying the opinion of its sub-feature/class terms in the ontology

1.6.2 Specific objectives:

1. Collect Arabic reviews for product domain.
2. Investigating the most suitable opinion mining preprocessing techniques such as (eg. Stop word removal, stemming, part of speech etc.).
3. Finding the best way to construct an ontology tree for the specific domain.
4. Determine the important features from the review using the ontology tree constructed.
5. Determine the opinions of the features using public lexicon.
6. Exploit extracted features and opinions to determine the overall polarity of review.
7. Summarization is done to generate feature-based summaries of document reviews.
8. Evaluate the performance using different performance metrics such as accuracy, precision, recall and f-measure.
9. Compare our method with supervised methods.
10. Subjective evaluation of our method using the f-measure metric to ensure the effectiveness of feature and opinion extraction process.

1.7 Signification

The significance of this research are:

- Improve the performance of Arabic opinion mining at feature level classification and generate complete feature-besed summary can be utilized for e-commerce and many businesses' benefit. It can be taken into account in product quality improvement by understand what the Arabic customer like and dislike in the product. Also, for the customer who wants to buy a product would like to know the opinions about features in specific product from the existing users
- Feature summary can be saving efforts and time by helping the manufactures to find which features will be improved in the product that customer dislike it.
- Mining Arabic opinions from these vast amounts of reviews becomes urgent, and has attracted a lot of attentions from many Arabic researchers.
- We are one of the few researchers who have worked at feature level in Arabic opinion mining.

1.8 Scope and Limitations

- Corpus collected in domain dependent reviews such hotels and books.
- We work at feature level opinion mining.
- We exclude the review that doesn't have the Arabic words and features.
- We manually add features that do not exist in the ontology.
- We use global lexicon to get the opinion of extracted features.
- We manually add opinions in our lexicons that do not exist in the global lexicons.

1.9 Methodology

To achieve our main objective to improve the accuracy for sentiment classification we need to follow the following steps, as seen in figure (1.2), which has the following steps:

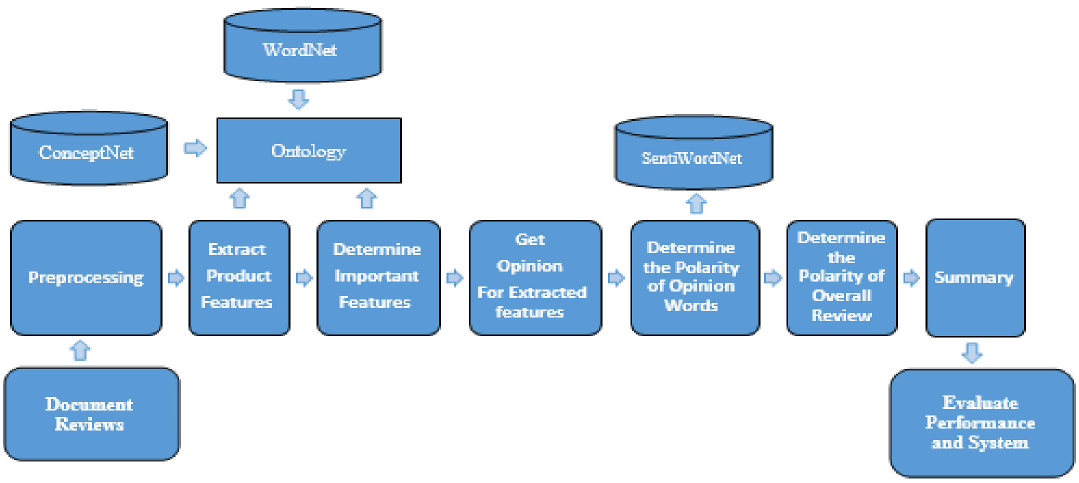

Figure (1.2): Flow diagram for our methodology.

Step 1 - Document reviews: The datasets selected for our methodology are hotel and book datasets from internet sites.

Step 2 - Preprocessing: The second step is text preprocessing. Preprocessing includes sentence splitting, tokenization, POS tagging and stemming techniques.

Step 3 - Construct Ontology Tree: Building the ontology tree is the main step in our method to extract features and determine the important features.

Step 4 - Extract Product Features: Ontology created is used to extract the product features.

Step 5 - Determine Important Features: Identifying which feature is important using the ontology constructed is the main contribution of our research.

Step 6 - Get Opinion of Extracted Features: Get opinion related to the features extracted.

Step 7 - Determine the Polarity of Opinion Words: Identify the polarity of opinion words are positive or negative.

Step 8 - Determine the Overall Polarity of the review: Determine the overall polarity (OP) of the reviews using a formula.

Step 9 - Summarization: Generating feature-based summaries of document reviews.

Step 10 - Evaluate the System: We compare our methods with other opinion mining techniques.

Step 11 - Subjective evaluation: Because our method is subjective**,** it requires subjective evaluation of our method**.**

These steps will be described in more detail in chapter four.

1.10 Overview of Thesis

This thesis consists of five main chapters, which are structured around the objectives of the research. The main points discussed throughout the chapters are listed below:

Chapter 1 - Introduction: Will present the introduction, research problem, objectives, scope, significance and our methodology.

Chapter 2 - Theoretical Background: This chapter presents details about the opinion mining definition, the opinion mining levels, the opinion mining approaches, the opinion mining classification based on supervised and unsupervised techniques, and the concept of ontology and ontology used in opinion mining.

Chapter 3 - Related Works: It presents other works related to my thesis.

Chapter 4 - Methodology: It includes the methodological steps and the architecture of our method. An explanation about each step used in our method.

Chapter 5 - Experiments and Results: It gives in detail the sets of experiments, and discusses the experimental results.

Chapter 6 - Conclusion and Future work: It discusses the final conclusions and presents possible future works.

Chapter 2

Theoretical Background

Chapter 2
Theoretical Background

This chapter provides some knowledge relevance to our work. We introduce the definition of opinion mining, opinion mining levels, opinion mining approaches, ontology concepts, databases, methods, and the preprocessing steps used in our method.

2.1 Opinion Mining

Opinion mining is the field of study that analyzes people's opinions, sentiments, appraisals, evaluations, attitudes, and emotions towards entities such as products, services, individuals, organizations, issues, events and their attributes. There are other names also used for opinion mining e.g., Sentiment analysis, opinion extraction, sentiment mining, subjectivity analysis, effect analysis, emotion analysis, review mining, etc. The main objective of opinion mining is to extract features and components of the object that have been commented in the review and to determine whether the review is positive, or negative (Almas & Ahmad, 2007).

Opinion mining can be useful in different ways. It can help the vendors in market to evaluate the success of a new product launch. also, to decide which versions of a product are familiar and identify which features lovable or unloved to the customers. Individual consumers also want to know the opinions of existing users of a product before buying it, and others opinions about political candidates before making a voting decision in a political election and others (Aggarwal & Zhai, 2012).

The field of opinion mining is recent and there are still a lot of challenges to be met such as the use of slang language, the fact that the reviews are entered by various people who vary in the way they expression or in the knowledge they use. Another challenge that we have attempted to initialize in this research is the strength of identified features in the review. That because, not all feature in the review have the same important in failing to determine the proper polarity of the review.

There are four levels on which opinion mining will be detailed in the following subsections.

2.1.1 Document Level

Document opinion analysis is about classifying the overall document that have sentiment words expressed by the authors. The task is to determine whether a document is positive, negative or neutral (Wawre & Deshmukh). Document level categorization attempts to classify sentiments in web forum postings, blogs, movie and news articles.

2.1.2 Sentence Level

Opinion mining at sentence level classifies each sentence to positive or a negative opinion. This is a more involved task than document-level classification. The first task is to identify whether the sentence is subjective (opinionated) or objective. The second task is to classify a subjective sentence and determine if it is positive, negative or neutral (Ding & Liu, 2010).

2.1.3 Word Level

The task of determine which word is positive or negative sentiment in certain domain. Words that represent a desirable state (e.g. Good) have a positive orientation, while words that represent an undesirable state have a negative orientation (e.g. Bad). To apply opinion mining, researchers have used a list of words and phrases for adjectives, adverbs, verbs, and nouns that called the opinion lexicon (Turney, 2002).

2.1.4 Feature or Aspect Level

Both the document level and the sentence level opinion mining do not discover what exactly people liked and did not like. Aspect level performs fine-grained analysis. Aspect level was recently called feature level (feature-based opinion mining and summarization) (Lu, 2013). Instead of looking at entire documents, paragraphs, sentences, clauses or phrases, feature level directly looks at the opinion itself. It is based on the idea that an opinion consists of a positive or negative sentiment and a target of opinion. The benefit of opinion targets helps us to understand the opinion mining problem better. In many applications, different features/aspects describe opinion target. Thus, the goal of this level is to discover features and opinion about their features. In our thesis, we work at the feature level opinion mining and make a summarization about each feature.

2.2 Preprocessing for Opinion Mining

Preprocessing stage is the important stage at feature level opinion mining classification, It includes sentence splitting, tokenizing strings of words, part of speech (POS) tagging technique and finally applying the suitable term of stemming. The following preprocessing process was used in the present work (Lazhar & Yamina, 2012).

2.2.1 Sentence Splitting

Sentence splitter is used to identify sentence terminating such as "." , "," and also to split the review into sentences.

2.2.2 Tokenization

In this process, a sequence of strings breaks into pieces such as words, phrases, symbols and other elements, called tokens, so that, text mining algorithms could be used. Arabic tokenization is complex due to the rich morphological features of Arabic (Salloum & Habash, 2011).

2.2.3 Part-of-Speech Tagging

Part of Speech (POS) is a category used in linguistics that is defined by a syntactic behavior of a word and plays important role in a sentence (Esuli & Sebastiani, 2006). POS tagging is often used in sentiment analysis, especially due to the fact that it can be used in word-sense disambiguation. A strong correlation between the presence of adjectives and subjectivity in sentences has also been discovered (Pang & Lee, 2008). (Turney, 2002) used POS tagging to construct conceptual sentence phrases, most of them including an adjective or adverb. Product features are usually nouns or noun phrases in the review (Ghorashi, Ibrahim, Noekhah, & Dastjerdi, 2012). Thus the POS tagging is important process to information extraction (Esuli & Sebastiani, 2006). We used the Stanford parser (Stanford NLP Group, 2013) to produce the part-of-speech tag for each word (whether the word is a noun, verb, adjective, etc.). The process also identifies the simple noun or verb groups (syntactic chunking). A POS tagger take a plain text document as input, and returns document where every word is associated with a tag that indicates the part of speech term as output. An example of POS has been discussed in (section 4.2). Table (2.1) illustrates the noun and the noun phrase pattern. Also, a definite linguistic filtering pattern is a noun phrase as the following patterns (Marcus, Marcinkiewicz, & Santorini, 1993).

Table (2.1): Describes the noun and noun phrases patterns (Marcus et al., 1993).

Noun and Noun phrases	
Noun	NN NNP (Proper noun) NNPS (Proper noun, plural) NNS (Plural)

2.2.4 Stemming

Stemming techniques can be used in text preprocessing to reduce different forms of word to one form (root or stem) (Hadni, Lachkar, & Ouatik, 2012). There are two different stemming techniques; root stemming and light stemming.

2.2.4.1 Arabic Root Stemming technique

Root stemming technique would reduce the Arabic words such as (مطاعم,طعام) to one stem (طعم).

2.2.4.2 Arabic Light Stemming technique

Light stemming, in contrast, removes common affixes from words without reducing them to their stems. For example, the light stemming approach, maps the word (الكتاب) to (كتاب).

2.3 Opinion Mining Approaches

Semantic orientation (SO) determination is a task of determine whether a document, sentence, or the feature has either positive or negative orientation. The following subsection presents the major approaches for opinion mining classification.

2.3.1 Lexicon Based Approach

The lexicon-based approach has concentrated on using adjectives as indicators the polarity of text. First, a list of adjectives and corresponding score values are compiled into a dictionary. Then, for any given text, all adjective words are extracted and annotated with their polarity, using the dictionary scores. The polarity scores are aggregated into a final score for the review. In our method we use the general Arabic Sentiment Lexicon called ArSenL (Qatar University, 2014) as will be detailed in the following point.

2.3.1.1 Arabic Sentiment Lexicon (ArSenL)

ArSenL is an Arabic sentiment lexicon containing a polarity score of opinion words. Words in ArSenLare divided in four categories: adjective, adverb, verb, and noun. It can be obtained from (Qatar University, 2014). ArSenL was built using WordNet3 (Princeton University, 2010), SentiWordNet3 (SentiWordNet, 2013) and Morphological Analyzer (jonsafar, 2013). It contains words with three scores as given below, that is:

1. Positive score.
2. Negative score.
3. Objective score.

For every word, positive, negative and neutral scores are having values between 0.0 and 1.0 and the addition of all the scores, that is, positive score, negative score, and objective score for a word, is 1. The objective score of 1.0 denotes that it is a neutral word and does not express any opinion.

2.3.2 Machine Learning (ML) Approach

Like human learns from the past experiences, a computer doesn't have experiences, but it learns from data, which represent some past experiences of the application domain. Machine learning defined as "field of study that gives computers the ability to learn without being explicitly programmed." The machine learning approach for opinion mining often rely on supervised classification methods. In this approach, labeled data is used to train classifier. In supervised machine learning, two datasets are used: train and test data. The training data contains a set of training sets. A test data is the unseen data to evaluate classifier accuracy. In classification, the most commonly features used in most methods are the following:

Boolean model: Which indicates the presence or absence of a word with Booleans one or zero respectively.

Term Frequency: Is the number that the term T occurs in the document D.

Term Frequency Inverse Document Frequency (TF-IDF): Is a common weight scheme that is more meaningful, where large weights are assigned to terms that are occurred frequently in relevant documents but rarely in the whole document collection (Hammad, 2013; Hotho, Nürnberger, & Paaß, 2005).

In our experiments we use the most supervised classifiers as we detailed below to evaluate the performance of our method. The classifiers, we used are: decision tree (DT), Naïve Bayes(NB), and K-Nearest Neighbor (K-NN) classifier.

2.3.2.1 Decision Tree (DT)

The decision tree is supervised machine learning, where it is an active method for make classifiers from data. It is also a flow-chart-like tree structure, where each node denotes a test on an attribute value, each branch represents an outcome of the test and tree leaves represent label classes. In addition, it is used in determining the best course of action, in situations having several possible alternatives with uncertain outcomes. A decision tree classifier is modeled in two stages: tree building and tree pruning. In tree building stage, the decision tree model is built by recursively splitting the training data set and assigning a class label to leaf by the most frequent class. Pruning a sub tree with branches if error is obtained. Algorithm below presents the psedocode structure of Decision Tree (Quinlan, 1986).

Algorithm (2.1): Basic Structure of Decision Tree algorithm (Quinlan, 1986).

Algorithm: Generate decision tree. Generate a decision tree from the training tuples of data partition D.

Input:

Data partition, D, which is a set of training tuples and their associated class labels; attribute list, the set of candidate attributes; Attribute selection method, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes. This criterion consists of a splitting attribute and, possibly, either a split point or splitting subset.

Output: A decision tree.

Method:

(1) Create a node N;

(2) If tuples in D are all of the same class, C then

(3) Return N as a leaf node labeled with the class C;

(4) If attribute list is empty then

(5) Return N as a leaf node labeled with the majority class in D; // majority voting

(6) Apply Attribute selection method (D, attribute list) to find the "best" splitting criterion;

(7) Label node N with splitting criterion;

(8) If splitting attribute is discrete-valued and multiway splits allowed then // not restricted to binary trees

(9) Attribute list attribute list splitting attribute; // remove splitting attribute

(10) For each outcome j of splitting criterion // partition the tuples and grow sub trees for each partition

(11) Let Dj be the set of data tuples in D satisfying outcome j; // a partition

(12) If Dj is empty then

(13) Attach a leaf labeled with the majority class in D to node N;

(14) Else attach the node returned by Generate decision tree (Dj, attribute list) to node N;

End for;

(15) Return N;

2.3.2.2 Naïve Bayes (NB)

The NB is important for several reasons. It is very easy to construct, and not needing any difficult iterative parameter estimation schemes. This means it may be readily applied to large data sets. It is easy to interpret, understand, it often does surprisingly well and can usually be relied on to be robust and to do quite well (Wu et al., 2008). The NB classifier, works as follows (Hammad, 2013):

- Let D be training set of tuples and their associated class labels. As usual, each tuple is represented by a n-dimensional attribute vector, **X = (X1, X2,…., Xn)**, n measurements made on the tuple from **n** attribute, respectively, **A1, A2…An**.
- Assume that there are m classes, **C1, C2…Cm**. Given a tuple, **X**, the classifier will predict that **X** belongs to the class having the highest probability, conditioned on **X**. That is, the NB classifier predicts that tuple **X** belongs to the class **Ci** if and only if

$$P(Ci|X) > P(Cj|X) for\ 1 \leq j \leq m, j \neq \mathrm{I} \qquad (2.1)$$

Thus we maximize P(Ci|X). The class Ci for which P(Ci|X) is the maximized, is called the maximum posteriori hypothesis. By Bayes' theorem (Equation (2.2))

$$P(Ci|X) = \frac{\mathrm{P(X|Ci)P(Ci)}}{\mathrm{P(X)}} \qquad (2.2)$$

- As P(X) is constant for all classes, only P(X|Ci) P(Ci) needs maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equal.
- Based on the assumption that attributes are conditionally independent (no dependence relation between attributes), P(X|Ci) using Equation (2.3).

$$P(X|C) = \prod_{k=1}^{n} P(Xk|Ci) \tag{2.3}$$

Equation (2.3) reduces the computation cost, only counts the class distribution. If **Ak** is categorical, **P(Xk|Ci)** is the number of tuples in **Ci** having value **Xk** for **Ak** divided by **|Ci, D|** (number of tuples of Ci in D). And if **Ak** is continuous-valued, **P(Xk|Ci)** is usually computed based on a Gaussian distribution with a mean **μ** and standard deviation **σ** and **P(Xk|Ci)** is

$$P(X|C) = g(Xk, \mu ci, \sigma ci) \tag{2.4}$$

$$g(Xk, \mu ci, \sigma ci) = \frac{1}{\sqrt[2]{2\pi\sigma}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.5}$$

Where μ is the mean and $\boldsymbol{\sigma 2}$ is the variance. If an attribute value doesn't occur with every class value, the probability will be zero, and a posteriori probability will also be zero.

NB classifier is fast, accurate, simple, and easy to implement, thus chosen to be one of the classifiers in this case. It is based on a simplistic assumption in real life and is only valid to multiply probabilities when the events are independent. Despite its naïve nature, NB classifier actually works well on actual data sets (Han et al., 2011).

2.3.2.3 K-Nearest Neighbor (K-NN)

K-nearest neighbor finds a group of k objects in the training set that are closest to the test object, and bases the assignment of a label on the predominance of a particular class in this neighborhood. To classify an unlabeled object, the distance of this object to the labeled objects is computed, its k-nearest neighbors are identified, and the class labels of these nearest neighbors are then used to determine the class label of the object. Once the k-nearest neighbor list is obtained, the test object is classified based on the majority class of its nearest neighbors:

$$MajorityVoting: y = argmax \sum_{(Xi,Yi)\in Dz} I(V = Yi) \tag{2.6}$$

where V is a class label, Yi is the class label for the i^{th} nearest neighbors, and I (·) is an indicator function that returns the value one if its argument is true and zero otherwise (Wu et al., 2008).

2.3.3 Combined Approach

In combined approach use both lexicon base and machine learning approach. The lexicon based approach uses opinion words and phrases to determine the semantic orientation of the whole document or sentence. Then, using theses words to classify the entire sentence in document and then classify the entire document. The next step, is to use machine learning approach. The documents that have been classified from the previous step will be used as a training set for the classifier. The goal in this step is to classify as many documents as possible that remain form lexicon based approach (El-Halees, 2011).

2.3.4 Ontology Based Opinion Mining Approach

Ontology, commonly referred to the concept of a domain (Gruber, 1993), aims to provide knowledge and concepts about specific domains that are understandable by both developers and computers. In particular, an ontology enumerates domain concepts and relationships among the concepts (Guarino, 1995), and provides a sound semantic ground of machine-understandable description of digital content. Ontology is popular in annotating documents with metadata, improving the performance of information extraction and reasoning, and making data interoperable between different applications (Baziz, Boughanem, Aussenac-Gilles, & Chrisment, 2005; Duo, Juan-Zi, & Bin, 2005; Fensel, 2002). Using ontology in opinion mining get Several advantages which are: structuring of the features and extraction of features (Pang & Lee, 2008). In order to build ontologies in our method, we use ConceptNet and WordNet databases that explained in the following point.

2.3.4.1 ConceptNet Database to Build Ontology:

ConceptNet is a large semantic network consisting of huge number of common sense concepts or features (Havasi, Speer, & Alonso, 2007; H. Liu & Singh, 2004). Common sense knowledge in ConceptNet is subscribed by ordinary people on the Internet. It is the largest machine usable commonsense resource consisting of more than 250,000 relations. It consists of nodes (concepts) connected by edges (relations between concepts). Some of the relationships between concepts in the ConceptNet are IsA, PartOf, HasA, and so forth (Havasi et al., 2007). ConceptNet database contains the four major properties, that is, start, rel, and weight. Here, start and end are the two concepts which are having a relation. For example, Hotel UsedFor sleep, Hotel IsA place, and so forth. The ConceptNet semantic graph represents the information from the OpenMind corpus as a directed graph, in which the nodes are concepts and the labeled edges are common sense relation that interconnect them. However, in our research, we use ConceptNet for the following reasons:

1. Ontology creation using ConceptNet is independent from the reviews.
2. The relational predicates in ConceptNet have an inherent structure, it is appropriate for building ontology tree.
3. ConceptNet has a closed class of clear relations with weighted.
4. The continual expansion of the knowledge resource through crowdsourcing incorporates new concept and enriches the ontology.

2.3.4.2 Arabic WordNet (AWN) Database:

The Arabic WordNet database structure consists from four principal column which are: item, word, form and link. Item column is conceptual entities, including synsets, ontology classes and instances. Item has a unique identifier and descriptive information such as a gloss. Items lexicalized in different languages are distinct. A word column is a word sense, where the word's citation form is associated with an item via its identifier. A form column is an entity that contains lexical information. The form column are the root form for the word, where applicable. A link relates two items, and has a type such as “equivalence,” “subsuming,” etc. Links link sense items, e.g., a Princeton Wordnet (PWN) synset to an AWN synset. This data model has been specified in XML as an interchange format, but is also implemented in a MySQL database hosted by one of the partners. Black et al. (2006) present the basic criteria for selecting synsets covered in AWN witch are:

- **Connectivity:** AWN should be as violently connected as possible by hyperonymy/ hyponymy chains, etc. Most of the synsets of AWN should correspond to English WN counterparts and the overall topology of both wordnets should be similar.
- **Relevance:** Frequent and salient concepts have priority. Criteria will include the frequency of lexical items (both in Arabic and English) and the frequency of Arabic roots in their respective reference corpora.
- **Generality:** Synsets on the highest levels of WN are preferred. These criteria suggest two ways for proceeding:
- **From English to Arabic:** Given an English synset, all corresponding Arabic variants (if any) will be selected.
- **From Arabic to English:** Given an Arabic word, all its senses have to be found, and for each of these senses the corresponding English synsets have to be selected.

2.3.5 Feature-Based Summarization Approach

With such a feature-based summary, a potential customer can facilely see how the people feel about the any product. If customers is very interested in a particular feature, they can drill down by following the link to see why the existing customers like it or what they complain about. For companies, it is possible to produce final summary from multiple trader sites for each of its products. Feature based summaries task involves three subtasks (Hu & Liu, 2004):

1. Identifying product features that customers have expressed their opinions on.

2. For each feature, identifying opinion words related to feature and determine the polarity of it (positive or negative).
3. Producing a feature-based summary of feature extracted.

2.4 RapidMiner tools:

RapidMiner tool used for text preprocessing and classification process (Yang, 2007). RapidMiner is a software developed by the company of the same name that provides an integrated environment for machine learning, data mining and text mining. It is used for developer and industrial applications as well as for research, education, training and rapid prototyping. RapidMiner provides more than 1,000 operators for all main machine learning procedures, including input and output, and data preprocessing and visualization. Also provides a large collection of machine learning algorithms for data preprocessing, classification, clustering, association rules, and visualization, which can be invoked through a common Graphical User Interface. In text mining classification, process Documents from files generates word vectors from a text collection stored in multiple files. It also provides multiple term weighting schemes, and term pruning options (Yang, 2007)..

Chapter 3

Literature Review

Chapter 3
Literature Review

Opinion mining classification approaches focus on lexicon-based approach, machine learning approach and combine classification approach. Most of the research works at the feature level depending on the ontology to extract features and some of them generate a summary of product features. In the following sections we introduce some of these researches which relate to our research.

3.1 Lexicon Based Approach

The lexicon based approach is the work discussed by Elhawary and Elfeky (2010). It is one of the earliest works of Arabic opinion mining whose goal is to mine Arabic business reviews. In this research, system comprises into two components: a reviews classifier that classifies any webpage whether subjective or objective, and a opinion mining that identifies the review text contain a sentiment positive, negative or neutral. The authors use a seed set of 1,600 words. The seed words are used with in an Arabic similarity graph built using a large web corpus using a labeling propagation mechanism to determine the polarity of neighboring terms.

The authors of Farra et al. (2010) use two approaches for identification sentence polarity which are grammatical syntactic approach and a semantic approach. In grammatical approach, they use the manual POS tagging to tag each word in the sentence. In semantic approach the class of sentence is determined using specific lexicon which contains a list of Arabic word roots storing with the it's polarity (positive, negative or neutral), which are extracted using an Arabic stemmer program. Then this root is checked against the stored dictionary. If the root is present, its polarity was extracted as positive, negative or neutral. Otherwise, the lexicon asks the user to identify the polarity of the word it has not learned yet and adds its root to the list of learned roots.

The shortcoming using the previous approaches are most of the sentiment lexicons is not publically available for Arabic language (manually created). Also, our approach is working at feature level classification that's differ from the previous approaches.

Furthermore, we incorporated the ontology in our method to improve the performance.

3.2 Machine Learning Approach

Rushdi et al. (2011) built an opinion corpus that contains 500 movie reviews collected from different web pages and blogs in Arabic, 250 of them labelled as positive reviews, and 250 of them labelled as negative reviews. Different experiments have been carried out on this corpus such as Support Vector Machine (SVM) and Naïve Bayes (NB). They utilized various N-grams models like (unigrams, bigrams and trigrams) and use tf–idf (term frequency–inverse document frequency) and tf (term frequency) as a weighting scheme. They used ten cross-validation to compare the performance of both learning algorithms, it is noticeable that SVM get accuracy of 91% that overcome the accuracy of NB classifier.

Abbasi, Chen, and Salem (2008) perform opinion mining for English and Arabic web forums. They use both syntactic and stylistic features for opinion classification. Syntactic features include word, n-grams, POS tag, n-grams. In stylistic features include the length of the review, the existence of special characters and repeat some of special words. Then, they use Support Vector Machine (SVM) classifier with entropy-weighted genetic algorithm (EWGA) as a feature selection technique in on an English movie review and on English and Arabic forums. They used information gained as a heuristic to weigh the various sentiment attributes. The experimental results using EWGA with SVM indicate high performance levels, with accuracy over 95% in the movie review and over 93% for both the U.S. and Middle Eastern forums.

Duwairi, Marji, Sha'ban, and Rushaidat (2014) used the supervised approach to analyze opinion mining in Arabic tweets. They collected about 350,000 to 25,000 label reviews as Positive, Negative, or Neutral. Three classifiers in RapidMiner were used to assess their work named: Naïve Bayes (NB), K-nearest classifier (K-NN) and Support Vector Machines (SVM). The best accuracy achieved by SVM was 71.68% when both stopword filter and stemming were disabled and 10- fold cross validation was used.

Most research in the supervised approach work at the document level and sentence level classification and they do not consider the features of the underlying product domain in the review. Our method works at feature level and produces summarization with effective performance in opinion mining class.

3.3 Combined Classification Approach

Combined classification approach of three methods is used by El-Halees (2011) to extract reviews automatically from the Arabic documents. In the first step, manually built lexicon is used to classify those reviews. The classified reviews are used as a training set for maximum entropy classifier which subsequently classifies other documents. In the final step, he use K-Nearest Neighbor (K-NN) method to classify documents which contain opinions doesn’t classified before. Using the three approaches lead to enhancing the effectiveness of classification to 80%.

Prabowo and Thelwall (2009) used rule-based classification and machine learning into a new method. In machine learning classifier they carried out 10-fold cross validation for each sample set. For each fold, the samples were split into train and test sample. For a training sample, the Rule Based Classifier (RBC) used a Rule Generator to generate a set of rules derived from the training sample and used this rule to classify the test sample. If the test sample was unclassified, the RBC passed the associated antecedents onto the Statistic Based Classifier (SBC), if the SBC could not classify the test sample; the SBC passed the associated antecedents onto the General Inquirer Based Classifier (GIBC). For test sample, a combined classification is carried out for example, if one classifier fails to classify a document, the classifier passes the document into the next classifier, until no other classifier exists or the document is classified. They noticeable from their experiments that use of more than one classifiers in a combined manner is better effectiveness in terms of F-measure than individual classifier.

The previous research used both unsupervised and supervised approach work at document level and sentence level classification and they don't consider the features of the underlying product domain in the review. Our approach is to work the at feature level classification in specific domains and we incorporate the ontology in our method to improve the performance.

3.4 Ontology Based Opinion Mining

Abdullah and Abeer (2016) proposed feature-based Arabic opinion mining using ontology. Their work utilizes the semantic of ontology and lexicon to identify the features and polarity related to it's feature. Their work consists of three stages: First stage, ontology and lexicon development. Second stage, ontology-based feature identification. Finally, configurable N-GRAM methods to identify the opinion of the feature. They used 890 reviews related for hotel domain with equal number of positive and negative number reviews. For evaluate their system, they manually tagged the reviews to compare them with their system. The best results are obtained with average accuracy is 95.5%.

Lazhar and Yamina (2012) focused on domain ontology. They use domain ontology to provide many of semantic information, structuring of features, extraction of explicit features, and for producing feature-based summary. Their method consists of the following phases: starting from sentence splitting, then opinion extracting, feature extraction, associating opinion to specific features and finally, classify the identified opinions into positive or negative classes using supervised classification techniques.

The work of Peñalver-Martinez et al. (2014) presented an innovative method using ontology to improve the feature-based opinion mining classification by employing the ontology in feature selection process. Once the features of the opinions have been identified, the score of the features in each user's opinion is calculated, n-gram words are used to determine the polarity of the features. Finally, SentiWordNet lexicon was used for weighting this polarity.

Yaakub, Li, Algarni, and Peng (2012) developed an ontology to do feature based opinion mining of customer's review on smart phones. The main objective of their work is to transfer reviews to structure table that includes several dimensions, such as, customers, products, time and locations. This system consist of three parts:

extraction, transformation, and loading process. In the extraction process part, they use part of speech tagging to preprocess the data in the source files. In the transformation part, the extracted nouns will be matched with the ontology to determine the features that customers want to review. Consequentially, the polarity of extracted opinion words will be calculated. A pair of e feature and polarity insert into fact comment table with related dimension key and ready to loaded into data warehouse. They evaluate the features and opinion extraction process in the reviews. The valuation result shown the better precision of 91.8% and the recall of 82.8%.

The work of Freitas and Vieira (2013) proposed method to determine polarity of Portuguese user produced reviews based on the features described in the domain ontology. This work is composed into four main steps. Initially, the algorithm receives as input a set of reviews which are pre-processed. After, explicit aspects are identified in the reviews using ontology. The polarity of opinion words relies on a lexicon of tagged positive, negative, and neutral of opinion words. Finally, opinion mining module tuples with object features and polarity are generated. They obtained of 0.62 f-measure for movie concepts in positive polarity.

Mukherjee and Joshi (2013) used ConceptNet database to automatically construct a domain-specific ontology tree for product reviews, without requiring any labeled training data, they use lexicon to determine the polarity of opinion words in the reviews. Then the feature with it’s opinion word polarity by using ontology are aggregated bottom-up, exploiting the ontological information.

The work of Agarwal, Mittal, Bansal, and Garg (2015) uses the same idea in Mukherjee and Joshi (2013). The ontology used in opinion mining model extracted from the ConceptNet and WordNet database for better coverage of the product features. They used ontology to determine the domain specific features which in turn produced the domain specific important features. Further, the polarity of the extracted features are determined using more than one lexicons which we developed by considering the context information of a word.

In Cadilhac, Benamara, and Aussenac-Gilles (2010), the use of a hierarchy of features improves the performance of features based identification systems.

However, works using domain ontology exploit the ontology as a taxonomy using only “is a” relations between concepts. The opinion words are extracted using rule based approach. By manually annotated explicit features as well as opinions related to the features, the evaluation result shown the f-measure is 77%.

Zhao and Li (2009) proposed ontology-based approach for opinion mining classification. The ontology describes the semantics of a domain and concepts with their relation. Firstly, the part of speech POS tool used for giving names to extract the noun words from the review such as Zoom, battery life, image quality, etc. In this work, they define two categories of features, frequent features and infrequent features. The experiment result shows the benefits of exploiting ontology structure to opinion mining. They obtained accuracy of 88.30% in positive reviews and 81.7% in negative reviews. Also they evaluate their system by randomly select 60 positive review documents and 60 negative reviews and compare features extracted manually with features extracted from their method and obtained accuracy of 76.90%.

Lau, Lai, Ma, and Li (2009) reported an automated analysis of the sentiments presented in online customer’s feedbacks that can facilitate both organizations’ business strategy development and individual consumers’ comparison shopping. He also proposed general system architecture of their Ontology Based Product Review Miner (OBPRM). The system describes a user first selects a product category and a specific product for opinion mining, based on the selected target product. The OBPRM system will use the Web services or APIs provided by e-Commerce sites and Internet Search Engines to retrieve the customer reviews for the particular product. Ontology extraction is carried offline and it must be performed before opinion sentiment polarity is conducted. The fuzzy domain ontology captures taxonomic information such as “iPhone” (product) “is-a” mobile phone (product category), and non-taxonomic relationship, such as “screen” (product feature) is “associated with” “iPhone” (product). In addition, opinion sentiment (e.g., “excellent”) of a product feature (e.g., “screen”) is also captured in the fuzzy domain ontology. They Evaluate their system based on a benchmark dataset and real consumer reviews collected from Amazon.com, their system shows performance improvement over the baseline.

In our approach, we have added the following tasks that do not exist in the previous works: firstly, feature level classification was used to classify Arabic user generated reviews by identifying the important features from the review based on level of these features on the ontology tree. Secondly, summarization of the reviews is done using ontology characteristics to determine which features have satisfaction and dissatisfaction the customers. Finally, we use the supervised approach with other unsupervised approaches to evaluate the performance of our method and make subjective evaluation to demonstrate the effectiveness of our method.

3.5 Opinion Summarization

Summarization is discussed in (Hu & Liu, 2004). They proposed a set of techniques for mining and summarizing product reviews based on natural language processing methods to provide a feature-based summary of a large number of customer reviews of a product sold online. The objectives of their work are: extract product features that have been commented on by customers. Then, identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative. Finally, generate the summarization result. For evaluate their system, they downloaded 100 reviews from five electronics products. The evaluation result of feature generation and opinion sentence extraction shown the effectiveness of their method.

The summary produced in Eirinaki, Pisal, and Singh (2012) depends on extracts of the most representative features of each reviewed item, and assigns opinion scores to them. This research presents an algorithm to identify the semantic orientation of specific components of the review that lead to a particular sentiment. Also, their algorithm integrated in an opinion search engine which presents results to a query along with their overall tone and they produce the summary of the sentiments of the most important features.

Efficient methods and techniques are used in (Htay & Lynn, 2013) to build effective summarization. Firstly, they use patterns knowledge to extract features that are nouns identified using POS tagging. Then, extract opinion words or phrases through adjective, adverb, verb, and noun, and determining the orientation. Finally, generating the summary. After that, they evaluate their system by comparing the results generated by the system with the results generated manually and obtained 79% f-measure for feature extraction process.

Our summarization task is different from the traditional opinion summarization, because we use the ontology characteristics to present a summary about the features of Arabic product reviews. Furthermore, our method uses a new way to evaluate feature and opinion extraction process by detecting the tuple (feature, opinion, polarity) and compare manual results with our system result.

Chapter 4

Research Methodology

Chapter 4
Research Methodology

In this chapter, we explain our methodology to classify Arabic opinion reviews which we followed in this research. The chapter organized into thirteen sections. Section 4.1, about overview of our methodology and short description about each steps. Section 4.2 about document reviews. Section 4.3, about preprocessing steps that we followed. Section 4.4, gives description of building ontology tree. Section 4.5, extract product features. Section 4.6, about determine important features. Section 4.7, get opinion of extracted features. Section 4.8, determine the polarity of opinion word. Section 4.9, about factors that effect on opinion words and able to increase the opinion performance. Section 4.10 calculate the overall polarity of review. Section 4.11, summarization. Section 4.12, evaluate the performance by comparing proposed method with other supervised and unsupervised techniques. Section 4.13, about subjective evaluation of our method.

4.1 Methodology Overview

We proposed to use a methodology, as shown in figure (4.1), the methodology is divided into five stages: Stage one preparation which contain document reviews and preprocessing steps, stages two ontology construction which contain feature extraction and determine important features, stage three opinion mining which contain determine the opinion of word and overall review, stage four summarization, stage five evaluation which contain objective and subjective evaluation; which has the following steps:

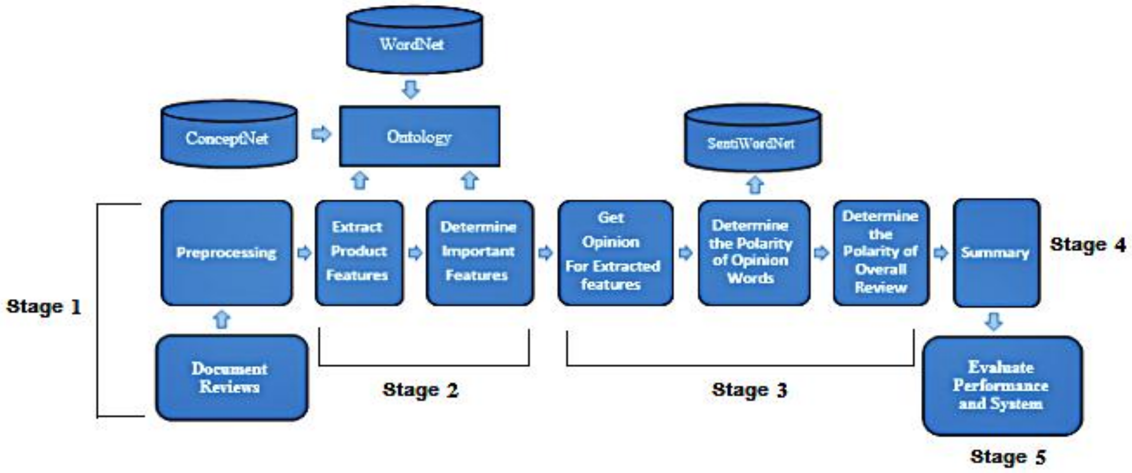

Figure (4.1): Flow diagram for our methodology.

4.2 Stage 1: Preparation

4.2.1 Document Reviews

To demonstrate the using of our proposed methods, we must choose domain have some features and available in website with Arabic language reviews. Therefore two datasets were selected from two different domains namely, hotel and book. First is a hotel review dataset from tripadvisor (2016) which contains reviews about hotels and its features such as room "غرفة", room service "خدمة الغرف" food "الطعام" etc. We collect corpus consists of 2000 reviews. Polarity of the review documents is classified as equal size positive and negative. Second corpus is book review dataset provided by Aly and Atiya (2013) which contains reviews about books and its features such as story "قصة", language "لغة", price "السعر" etc.; it consists of 2000 reviews of equal number of positive and negative.

Table (4.1): Number of positive and negative class with their source.

Domain	Number of Positive	Number of Negative	Source
Hotel	1000	1000	(tripadvisor 2016)
Book	1000	1000	(Aly & Atiya, 2013)

4.2.2 Preprocessing

Preprocessing is a necessary step for our method. There are some irrelevant and incorrect data, because that, we apply number of preprocessing techniques to get our objective, the steps we used are:

- Remove irrelevant data: Numeric words and non-Arabic words are not useful for our approach. Also, we exclude review doesn't have any feature.
- Sentence Splitting: Generally, split sentence depends on the use of delimiters such as "." , ",".
- Tokenization: Break up the review into tokens. The simplest meaningful token is a word which we used in our method.
- Arabic POS tagger: We used Stanford POS-Tagger from (Stanford NLP Group, 2013) to produce the part-of-speech tag for each word (whether the word is noun, verb, adjective, etc.). The Stanford POS tagger tool use the rule-based approach that consists of developing a rules knowledge base established by linguists in order to define precisely how and where to assign the various POS tags (El Hadj, Al-Sughayeir, & Al-Ansari, 2009). For example, "انه فندق جميل جدا", "It's a very beautiful hotel", POS tagged sentence is as follows: "انه/VBD فندق/NN جميل/JJ/جداNN".
- Stemming: This step is an important preprocessing step for input document reviews. Root stemming means to reduce words to their roots. In our method, we use root stemming technique for the following reasons: firstly, to produce better matching of features in the ontology. Secondly, the Arabic Sentiment lexicon contains only root words. For their reasons, we adopted Khoja stemming tool (fariscs, 2012) in our method rather than light stemming techniques.

4.3 Stage 2: Ontology

4.3.1 Building Automatic Domain Specific Ontology Tree

In our methodology, building ontology in our method is a needed step to extract product features in the reviews, to determine which extracted features is important and to generate feature-based summary.

In order to build an ontology, we use ConceptNet (Rob Speer, 2016) as a knowledge resource to automatically construct independent domain-specific ontology tree for product reviews. ConceptNet relations have an inherent structure which helps in the construction of an ontology tree from the resource. The sample ontology for “hotel” domain using ConceptNet database demonstrated in figure (4.2). In the next step, we expand our ontology by merging with each node in the ontology with synonyms words of Arabic Language using WordNet (Princeton University, 2010) database, the useful of using WordNet to better coverage of domain specific features in Arabic language. The Flow diagram for creation ontology demonstrated in figure (4.3).

Pseudocode was proposed in algorithm (3.1) to construct automatic ontology tree. This algorithm is a recursive function used to build automatic ontology tree. It takes the domain name (root) and number of levels of the ontology tree as input parameters. The get_features function using the SQL query to return a list of features from ConceptNet database that subclass of the root name parameter as seen in Table (4.2). The get_synonyms function also uses a SQL query to return Arabic synonyms words from WordNet database for the feature parameter. Finally, the output of function return ontology tree for specific domain.

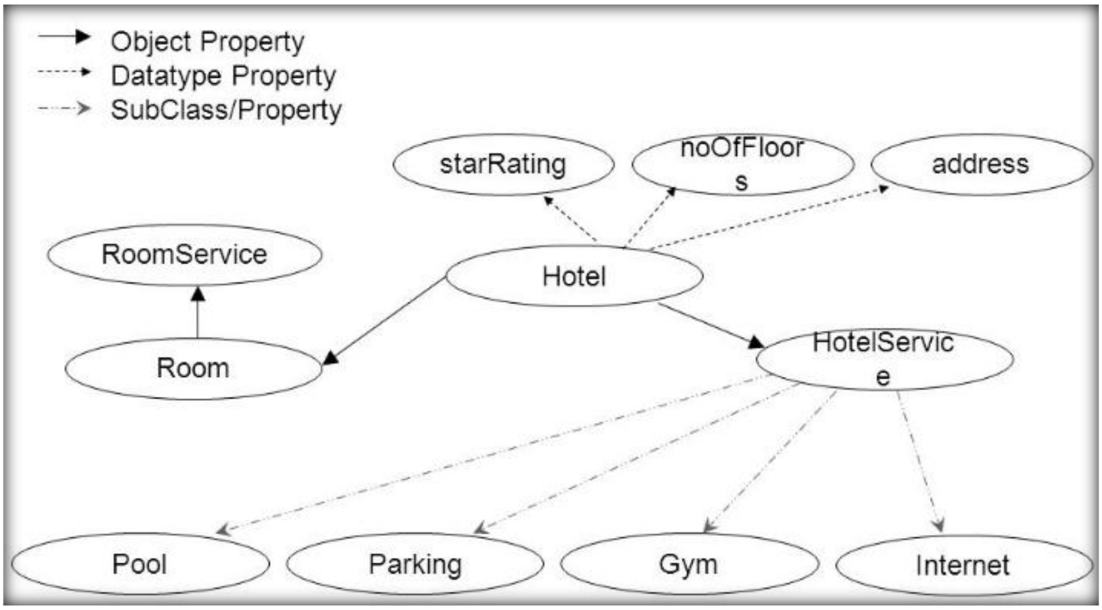

Figure (4.2): Sample Ontology for Hotel Domain.

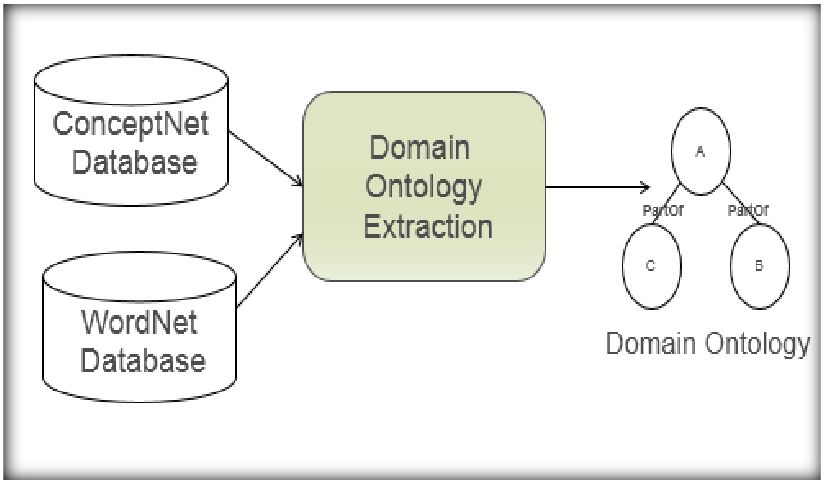

Figure (4.3): Flow diagram for building ontology tree.

Table (4.2): Concepts with relations in the ConceptNet database for hotel domain.

Start Concept	Relation	End Concept	Weight
Hotel	UsedFor	Sleep	1
Hotel	HasA	room service	1
hotel room	part of	Hotel	1

Algorithm (3.1): Algorithm for Creation Ontology Tree.

Algorithm: Function Build Ontology Tree
Function createOntologyTree (p_rootName, p_levelNo) **Input**: p_rootName parameter. // name root of ontology. p_levelNo parameter. // number of ontology tree level. **Output**: Ontology tree represents the concepts and their synonyms. Root $\longleftarrow$ p_rootName //A root node of tree that created recursively. No_of_level $\longleftarrow$ p_levelNo //The number of levels to deep into ontology searching for appropriate meaning for the root concept. If No_of_level = 0 then Return root; List_features=get_features(Root). //return the features that subclass from root For each feature ϵ List_features do Root.Add (feature); // append a node to the root List_synonyms = Get_synonyms(feature);//return the synonyms for these feature For each synonym ϵ List_ synonyms do Root. addSibling (synonym); // add Sibling a node to the root Return createOntologyTree (feature, No_of_level-1); *** **Function** get_features (root_parameter, ConceptNetDatabase) return list **Input**: root_parameter. // root name or node in the ontology tree. ConceptNetDatabase parameter. // database have two concepts with their relation. **Output**: return features that sub-class from root feature. V_list list; // variable list of nodes type; Select start into V_list from ConceptNetDatabase where end = root and rel = 'PartOf'

```
and weight = 1

Union

Select end into V_list from ConceptNetDatabase where start = root and rel = 'hasA';

Union

Select start into V_list from ConceptNetDatabase where end = root and rel =
'AtLocation'

etc.

return V_list;

}
```

****

Function get_synonyms (feature, WordNetDatabase) return list {

Input: feature parameter. // feature name or node in the ontology tree.

WordNetDatabase parameter. // database have synonym words for any word.

Output: return Arabic synonyms that related to feature parameter.

```
  V_list list; // variable list of nodes type.

Select Ar_Synonyms into V_list from WordNetDatabase where word = feature;

return V_list;

}
```

4.3.2 Extract Product Features

In our methodology, constructed ontology from previous step is used to extract product features. Feature is a term about which an opinion is expressed. To identify the feature term, all the noun terms are extracted from review. We used the Stanford Parser tool (Stanford NLP Group, 2013) to parse each review and to produce the POS tag for each word (whether the word is a noun, verb, adjective, etc.). This tool use the rule-based approach that consists of developing a rules knowledge base established by linguists in order to define precisely how and where to assign the various POS tags (El Hadj et al., 2009). For example, POS tagged review is as follows: "انه/VBD فندق/NN جميل/JJ/جداNN". Here, "فندق","جدا" is anoun terms may be relevant to features. After the extraction of noun terms, these are matched with the domain specific ontology constructed to eliminate all the irrelevant features like "جدا".

4.3.3 Determine Important Product Features

The main contribution of this research is to determine important features about which any opinion is expressed and identify which features is important than other features. with the help of automatically constructed ontology. The feature importance is captured by the height or level of a feature node in the ontology tree. For example, in the review:

"فندق جيد من حيث الاقامة ولكن توجد لدي ملاحظة من ناحية التلفاز ان نوعه قديم بالنسبة لمستوى الفندق".

"The hotel is good for staying but I have a remark about its TV. It's kind of an old version and doesn't match with the hotel"

Using baseline dictionary, the overall polarity of the review is neutral as respect with " جيد","قديم" but upon checking in the review will see the feature "الاقامة", "staying" is not the same important compared with feature "التلفاز","television" which means the overall polarity of the review is positive. For this reason, using ontology tree help us to determine the polarity of the review by determine the important features. As shown in Figure (4.4), The level in the hotel ontology presents the important features for example, the feature "الإقامة", "staying" was placed in level 2, but feature "تلفزيون", "TV" was placed in level 1.

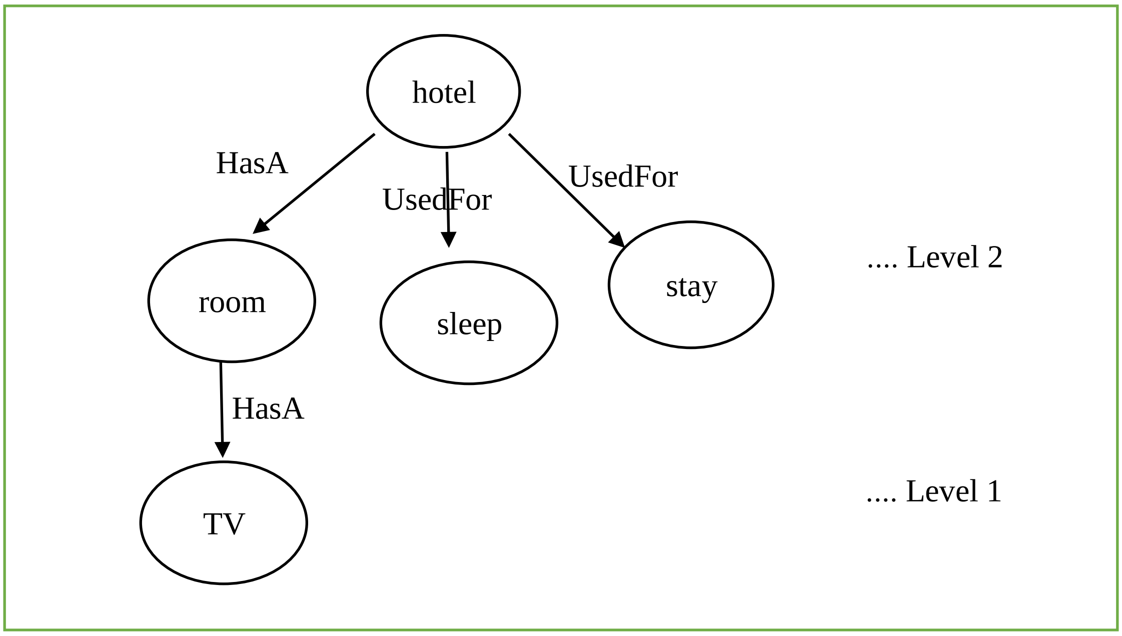

Figure 4.4): Sample Ontology tree with 2 levels from ConceptNet to explain previous review.

4.4 Stage 3: Opinion Mining

4.4.1 Get Opinion of Extracted Feature

After identifying the features in the review and determine the important features. The next step is to get opinion word related to specific feature. Determine polarity of the feature can get by identify the opinion word related to it's feature. Opinion words may be adjectives, verb and noun, for example, “ممتاز”, “excellent”,“أحب”, “love”,“أفضل”, “prefer”,“ممتاز”, “excellent”, “ليس جيدا”, “not good” are considered opinion words. To decide the opinion word related to specific feature, we used the following rules appropriate to deal with Arabic slang language and effective in opinion detection for specific feature. Figure (4.5) and (4.6) shown some example of rules used in our method. The first rule in figure (4.5) used to check the noun token that followed by opinion word of adjective category. For example,“البطارية ضعيفة”. On the other hand, the second rule in figure (4.6) used to check opinion word of the verb category, followed by first noun category example “أحب لغة الكتاب”.

```
First Rule
If (Token.category=="NN")
If (TokenNext1.category==JJ | | TokenNext1.category==DTJJ)
Then
candidateFeature = Token;
candidateOpinion= TokenNext1;
```

Figure 4.5): Rule used to detect the adjective opinion word related to feature.

```
Second Rule:
If (Token.category==" VBD ")
If (TokenNext1.category=="NN")
If (TokenNext2.category=="DTNN")
candidateFeature = TokenNext1;
candidateOpinion= Token;
```

Figure (4.6): Rule used to detect the verb opinion word related to feature**.**

During this step, we distinguish the following cases:

- If candidate features exist in features list extracted from previous steps, and ArSenL lexicon contains candidate opinion, then it is easy to extract the tuple (feature, opinion, polarity).
- If candidate feature doesn’t exist in the features list extracted from previous steps, and ArSenL lexicon contains candidate opinion, then domain ontology can be updated by adding a new concept or a new feature by entering the new record in the ConceptNet database of new concept with proper relation and concept that related to it.
- If candidate features exist in features list, and candidate opinion doesn’t exist in ArSenL lexicon, then add a new opinion words with polarity to our list.

4.4.2 Determine Polarity of Opinion Words

After opinion words extracted from previous steps, the polarity (positive or negative) of these words must be identified and these opinion words can be used to detemine the averall polarity of the review. Therfore, we use the Arabic Sentiment Lexicon ArSenL (Qatar University, 2014) lexicon to determine the polarity of these words. Table (4.3) shows the example of Arabic sentiment root words. It contains words with POS tag for opinion word (adjective, verb or noun), positive score and a negative score. Positive and negative scores having values between 0.0 and 1.0. In general, if positive_score is greater than negative_score then it considers the opinion word as positive polarity otherwise negative polarity.

```
If (positive_score – negative_score > 0) then
  P=positive
Else
  P=negative;
```

Table (4.3): Example of Arabic SentiWordNet.

Opinion word	POS	Positive score	Negative score
جيد	A	0.625	0
أحب	V	0.125	0
سيئ	A	0	0.75
حسن	A	0.625	0

4.4.3 Factors Effect on opinion polarity

In this section we present the important factors in our method that improve the performance in opinion mining class. In the following points, we introduce some of these factors:

Negation: It is important factor affect to the opinion words. Due to some negation such as "لم", "don't" and "ليس", "Not", the polarity of opinion turns to its reverse sign.

Intensifiers: Some particles intensify the strength of the polarity like "جدا" in the example "هذا الكتاب جميل ***جدا***". Words like "جدا" in the previous example intensify the strength for opinion word “جميل”. Because that, we manually form a list of 72 words such as “جدا”, “كثيرا”, “تماما”, “بدون شك”, “بشدة”, “أوي”, etc. These lists will be represented by constants **c** to be added to the polarity of opinion words. In our method we try to choose **c**= 0.2;0.4;0.6;0.8 and we notecable that the opinion mining classification performance improved when using **c**= 0.4.

4.4.4 Determine the Overall Polarity (OP) of the Review

After features are extracted from the review document, and then it is matched in the ontology. The level of ontology where it is located determines the importance of the feature. The features located at higher level near to the root of the ontology are considered to be more important as compared to the lower level features. Further, opinion word corresponding to this feature is detected using opinion extraction process. Further, the polarity value is retrieved from ArSenL lexicon. Finally, the overall polarity of the review is determined by summing up the opinion polarity multiplied by the height of ontology for each feature with respect to c factor that mentioned in previous section. In general, the following formula was proposed to determine the overall polarity of the review:

$$OP = \sum_{k=1}^{fn}((P * h) + \mathrm{c}); \tag{3.1}$$

Where f_n number of feature, P is polarity of opinion, h is height of feature in ontology and c if exist, is intensifier factor that effect to opinion polarity that explained in subsection 4.4.3.

For example, the review:

” فندق جيد من حيث الاقامة ولكن توجد لدي ملاحظة من ناحية التلفاز ان نوعه قديم بالنسبة لمستوى الفندق“

“The hotel is good for staying but I have a remark about its TV. It's kind of an old version and doesn't match with the hotel”

(الاقامة/قام، جيد/جود، +، 4.0)

(التلفاز/تلفز، قديم/قدم، -،3.0)

OP=4*(1) +3*(-1) =1; //Positive with determine important the features.

OP=1+(-1) =0;// neutral, when using ontology baseline.

4.5 Stage 4: Feature-Based Opinion Summary

Finally, after the feature and opinion extraction process is done, we are ready to generate the final feature-based summary. Feature-based summary which is different from the straightforward summary. Our summary depends on the ontology to identify the opinion summary of each feature in the whole corpus by identifying the opinion of its sub-class terms in the ontology. For instance, as seen in figure (4.7), the opinion summary of "غرفة" feature induces the aggregation of sub-class opinion summary such as "المطبخ", "التلفاز", "الباب", etc.. That belongs to the "غرفة" feature in the ontology.

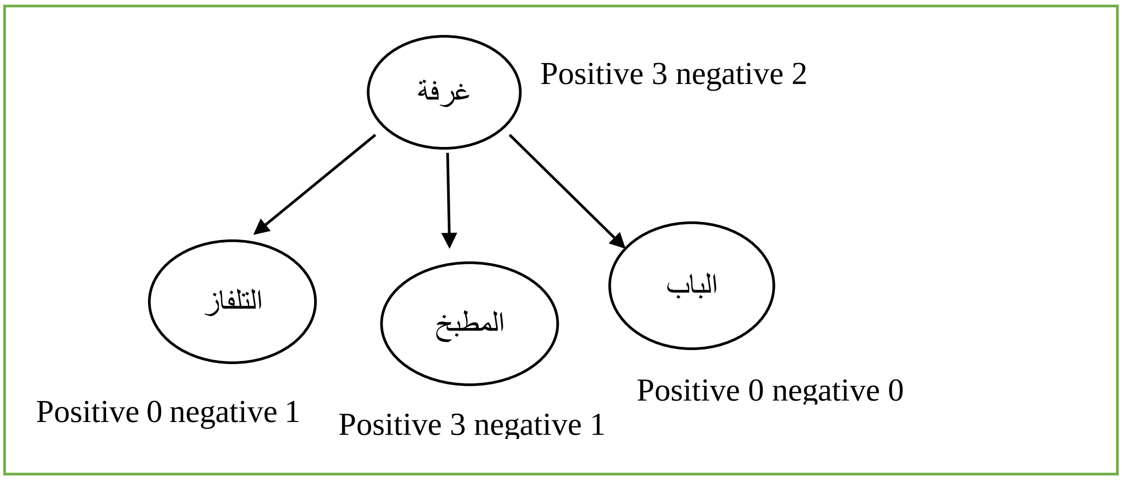

Figure 4.7): Example of feature-based summary for "غرفة" feature.

4.6 Stage 5: Evaluate the Performance

We used two approaches to evlaute our system, objective and subjective as follows:

4.6.1 Objective Evaluation

In this section we discuss the evaluation of our method. The measures evaluating of the performance of classification are a confusion matrix, which is also called a performance vector that contains information about realistic and predicted classifications.

Table (4.4): Confusion matrix table (Holte, 1993).

Predicted			
True		**Positive**	**Negative**
	Positive	(TP) True Positive	(FN) False Negative
	Negative	(FP) False Positive	(TN) True Negative

The entries in the confusion matrix are (Holte, 1993):

- The number of correct predictions that an instance is positive (TP).
- The number of correct predictions that an instance is negative (TN).
- The number of incorrect predictions that an instance is positive (FP).
- The number of incorrect predictions that an instance is negative (FN).

From the entries in the confusion matrix several concepts have been computed. These concepts will be used in later chapters to evaluate the performance of Appling unsupervised and supervised classification methods. These include Recall, Precision, F-Measure, and accuracy.

Accuracy: Is the proportion of the total number of predictions that were correct. It is determined using this equation (Holte, 1993).

$$\text{Accuracy} = \frac{TP + FN + FP + TN}{TP + TN} \tag{4.2}$$

Recall: True positive rate, Recall, or Sensitivity which is the proportion of Real Positive cases that are correctly predicted positive. This measures the Coverage of the Real Positive cases by the (Predicted Positive) rule. Recall is defined, with its various common appellations, by equation (Holte, 1993).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.3}$$

Precision: True False Accuracy, Precision or Confidence (as it is called in Data Mining) denotes the proportion of Predicted Positive cases that are correctly Real Positives. This is what Machine Learning, Data Mining, and Information Retrieval focus on, Precision is defined, with its various common appellations, by equation (Holte, 1993).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.4}$$

F-Measure: F-Measure or F-Factor is the ratio between recall and precision measurements F-Measure is defined, with its various common appellations, by equation (Holte, 1993).

$$\text{F} - \text{Measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.5}$$

4.6.2 Subjective Evaluation

Because our method is subjective and depends on how much feature and opinion process correctly extracted, subjective evaluation must be done to evaluate the effectiveness of our method and classification. We evaluate our results with the help of an experienced human judge. Therefore, we picked around 100 reviews for only two product feature from two different datasets (hotel and book). Then we asked a human judge to manually extract a tuple of (feature, opinion, polarity) for each review. The final evaluation is measured by comparing the results generated by our system with the manually generated results by the judge. We use recall, precision and f-measure in formula 4.6,4.7,4.8 to subjective evaluation of our method.

$$\text{Human Recall} = \frac{\text{the number of correct tuples marked by the system}}{\text{the total number of tuple marked by human}} \tag{4.6}$$

$$\text{Human Precision} = \frac{\text{the number of correct tuple marked by the system}}{\text{the total number of tuple marked by system}} \tag{4.7}$$

$$\text{Human F} - \text{Measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.8}$$

Chapter 5
Experiments and Results

Chapter 5
Experiments and Results

In this chapter, we describe the conducted experiments to evaluate our approach. We made three experiments which are lexicon baseline, ontology baseline (without consider important features) and ontology with consider important feature in the review. In order to compare our result, we used three classifiers which are Decision Tree (DT), Naïve Bayes (NB), and K-Nearest Neighbor (K-NN). We explain the machine environment, and the tools used in our experiment. In addition, we present the evaluation measurements for classification model by using the calculation of accuracy, precision, recall, and f-measure. Finally, we generate feature-based review summary and present the subjective evaluation for our method.

5.1 Datasets

Our method is performed in two domains corresponding to hotels and books. The two corpora are in Arabic, and each consists of 2000 reviews of equal number of positive and negative datasets. We acount the total number of tokens in both datasets as seen in Table (5.1). The hotel dataset has 18970 tokens. The book dataset has 21051 tokens. All the features/words extracted from the review documents are reduced to their root form for better matching of features in the ontology.

Table (5.1): Statistics on the dataset.

Domain	Positive reviews	Negative reviews	No. of tokens
Hotel	1000	1000	18970
Book	1000	1000	21051

5.2 Experiments Setup

In this section, a description about the experimental environment, tools used in experiments, measures of performance evaluation of classification methods.

5.2.1 Experimental Environment and Tools

We applied experiments on a machine with properties that is Intel (R) Core (TM) i3-3110M CPU @ 2.40 GHz, 4.00 GB RAM, 320 GB hard disk drive and Windows 7 operating system installed. To carry out our work (including the experimentation), special tools and programs were used which are:

NetBeans IDE 8.0.2 (Community, 2000): to build and evaluate our method and ontology.

Arabic Stanford Parser (Stanford NLP Group, 2013): tool used for Arabic POS tagging for reviews.

RapidMiner application program (RapidMiner Studio, 2016): used to do supervised classification methods, and extracting the required results that compared with our method.

MySQL 6.3 (Workbench, 2016): to handle with ConceptNet database and sentiment lexicon database.

5.3 Experiments

5.3.1 Preprocessing

Preprocessing is a necessary for both unsupervised and supervised algorithms. In our experiments, two preprocessing processes are used as shown in the following subsections:

5.3.1.1 Preprocessing for Unsupervised Approach

We apply a number of preprocessing algorithms which are sentence splitting, tokenization, POS tagging and Stemming algorithms. Figure (5.1) shows an example the preprocessing result that we applied in our datasets.

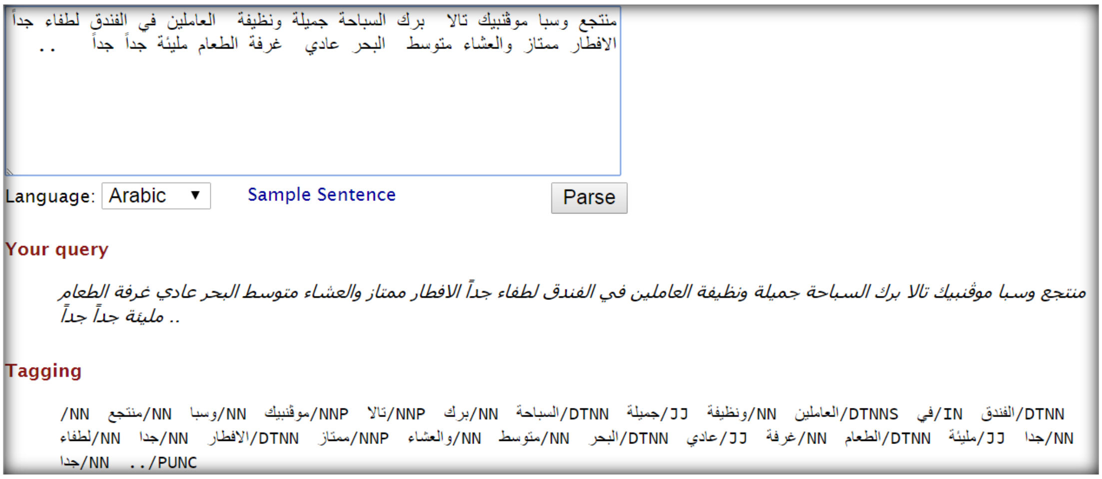

Figure (5.1): Example of POS tag for each token using Standford Parser tool.

5.3.1.2 Preprocessing for Supervised Approach

There are number of preprocessing processes used in supervised learning algorithms such as tokenization, filter stop words, and stemming technique as seen in figure 5.2:

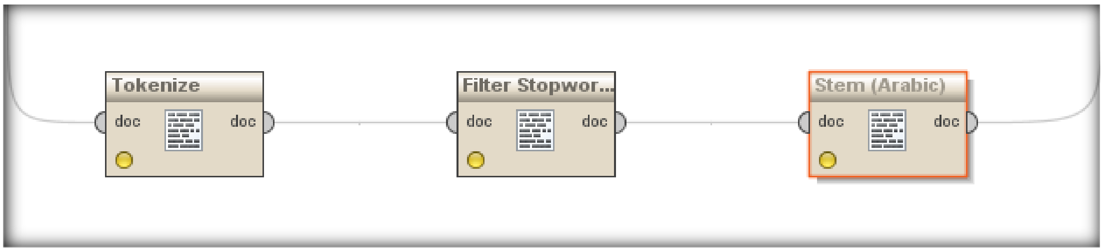

Figure (5.2): Preprocessing process used in supervised machine learning algorithms.

5.3.2 Ontology Construction

The ontology construction is done using ConceptNet. We extract the concepts/features from the ConceptNet up to level 4. We notice from our experiment that some features/concepts don't exist in ConceptNet database, therefore we manually add it in the ConceptNet database. Table (5.2) shows the statistics of the ontology tree in hotel and book domains. It consists of 223 nodes/features in hotel ontology such as room "غرفة", bed "سرير" and food "طعام'"; and 264 nodes/features in book domain such as story "قصة" and language "لغة" and style "أسلوب". We notice that ontology of hotel domain has little nodes comparing with book ontology, this is because WordNet database cover sufficiently the book domain better than hotel domain. Figure (5.3) shows the result of constructing ontology in the hotel domain.

Table (5.2): Ontology Tree Statistics.

Domain	Ontology Nodes Using ConceptNet	Ontology Nodes Manually Added	Total Ontology Nodes
Hotel	215	8	223
Book	251	13	264

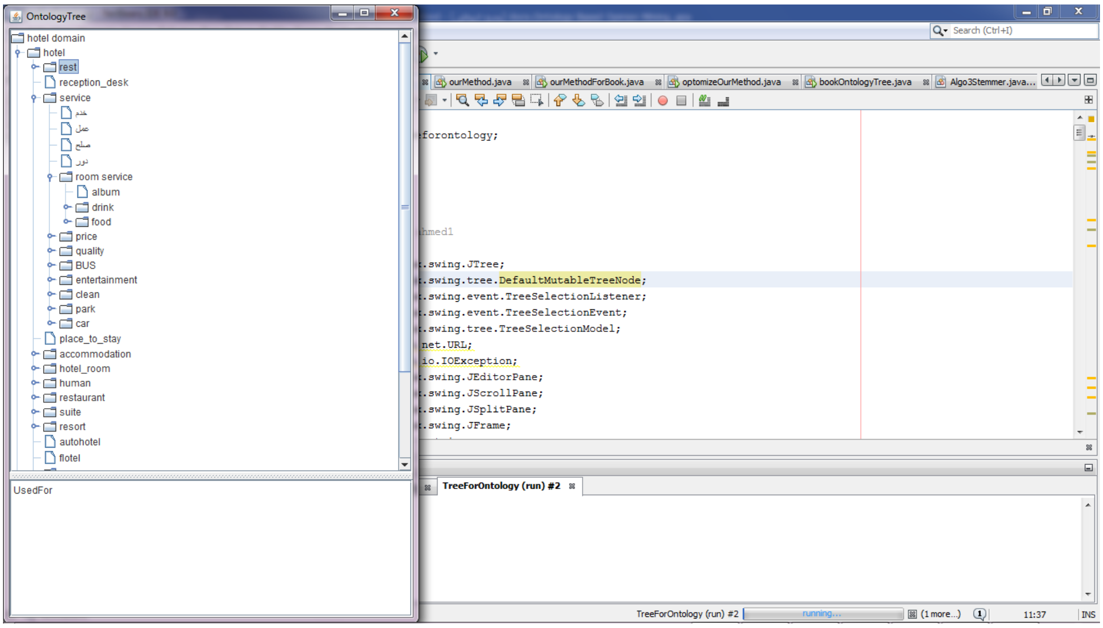

Figure (5.3): Example of ontology treein the hotel domain.

5.3.3 Supervised approach

In this section, we present the evaluation results of major supervised classification algorithms and compare the results with our methods. We use algorithms such as Decision Tree (DT), Naïve Bayes (NB), K –Nearest Neighbor (K-NN) which are provided in Rapid Miner environment. We decided to use the 10-fold cross validation splitting in supervised learning methods. The following subsections present these classification algorithms and evaluation results.

5.3.3.1 Decision Tree(DT)

Used the DT classifier in RapidMiner is presented in figure (5.4). The default parameters were used in DT classifier such as criterion = gain_ratio, maximum depth =20, using apply pruning and confidence = 0.25, as seen in figure (5.5). Table (5.3) shows the confusion matrix for DT classification approach. The correct positive number classified by the system is 465 in hotel domain and 508 in book domain. And, The correct negative number classified by the system is 924 in hotel domain and 947 in book domain. The false positive number in hotel and book domain are 76 and 53 respectively. The false negative number in hotel and book domain are 535 and 492 respectively. The result produced from this classifier in hotel and book review dataset are shown in Table (5.4). The DT classifier gives the f-measure of 60.35% for hotel review dataset and 65.09 % for book review dataset. Form the result, the NB classifier has less f-measure comparing with other supervised methods. Also there is a gap between precision and recall in both datasets. Furthermore, the f-measurement result of our method is better than DT supervised classifier.

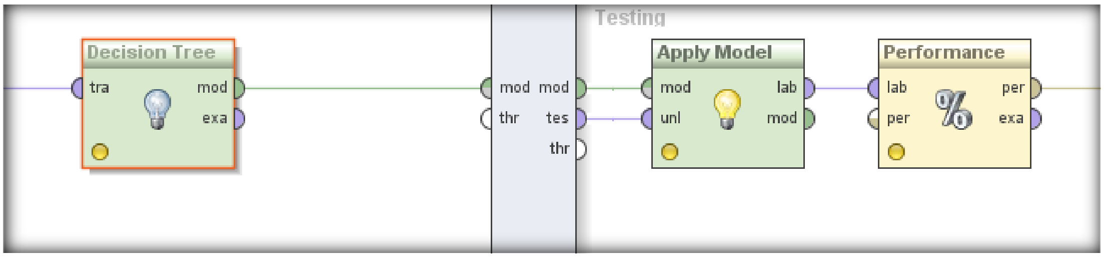

Figure (5.4): Illustrates the DT classifier in RapidMiner.

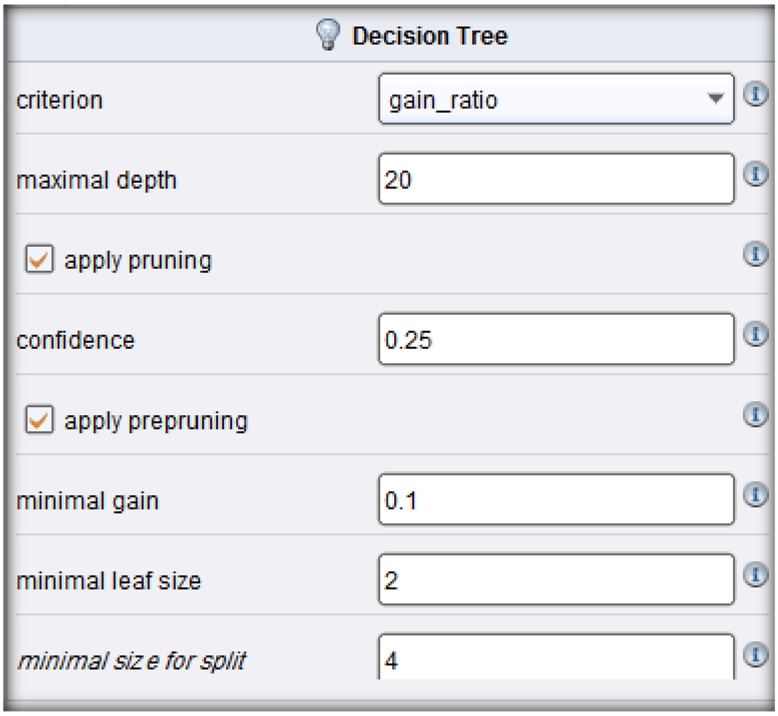

Figure (5.5): Decision tree parameters.

Table (5.3): Confusion matrix table for hotel and book domain using DT classifier.

Domain	N=2000	Positive	Negative
Hotel	**Positive**	465	76
	Negative	535	924
Book	**Positive**	508	53
	Negative	492	947

Table (5.3): Accuracy, Precision, Recall and F-Measure for hotel and book domain using DT classifier.

Domain	Accuracy	Precision	Recall	F-measure
Hotel	69.45%	85.95%	46.50%	60.35%
Book	72.75%	90.55%	50.80%	65.09%

5.3.3.2 Naive Bayes (NB)

Used the NB classifier in RapidMiner is presented in figure (5.6). Table (5.5) shows the confusion matrix for NB classification approach. The correct positive number classified by the system is 783 in hotel domain and 790 in book domain. And, The correct negative number classified by the system is 547 in hotel domain and 617 in book domain. The false positive number in hotel and book domain are 453 and 383 respectively. The false negative number in hotel and book domain are 217 and 210 respectively. The result produced from this classifier in hotel and book review datasets are shown in Table (5.6). The NB classifier gives the f-measure of 70.04% for hotel domain and 72.71% for book domain. From the result, our method overcomes of NB classifier in the performance.

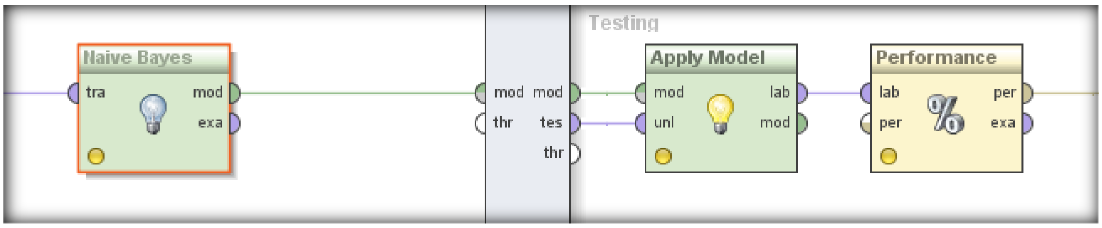

Figure (5.6): Illustrates the NB classifier in RapidMiner.

Table (5.5): Confusion matrix table for hotel and book domain using NB classifier.

Domain	N=2000	Positive	Negative
Hotel	**Positive**	783	453
	Negative	217	547
Book	**Positive**	790	383
	Negative	210	617

Table 5.6): Accuracy, Precision, Recall and F-Measure for hotel and book domain using NB classifier.

Domain	Accuracy	Precision	Recall	F-measure
Hotel	66.50%	63.35%	78.30%	70.04%
Book	70.35%	67.35%	79.00%	72.71%

5.3.3.3 K-Nearest Neighbor (K-NN)

Using K-NN classifier in RapidMiner is presented in figure (5.7). We use a parameter of k=1 as shown in figure (5.8). Table (5.7) shows the confusion matrix for K-NN classification approach. The correct positive number classified by the system is 787 in hotel domain and 829 in book domain. And, The correct negative number classified by the system is 662 in hotel domain and 694 in book domain. The false positive number in hotel and book domain are 338 and 306 respectively. The false negative number in hotel and book domain are 213 and 171 respectively. Experiment shows the f-measure of 74.07% of hotel domain and 77.66% of the book domain as seen in Table (5.8). We note that K-NN classifier with k=1 has better result over the other supervised methods, but is less performance compared with our method.

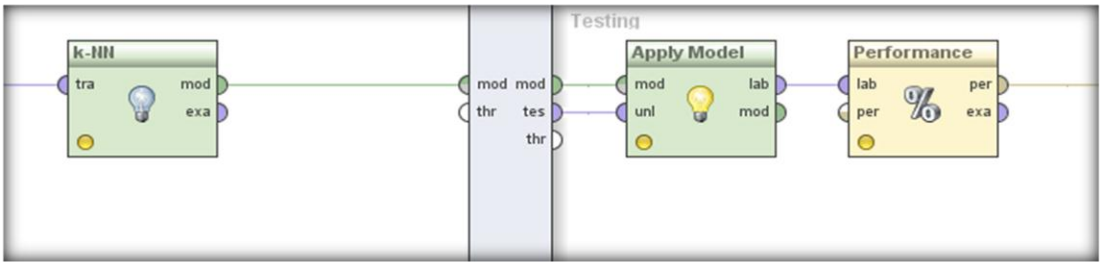

Figure (5.7): Illustrates the K-NN classifier in RapidMiner.

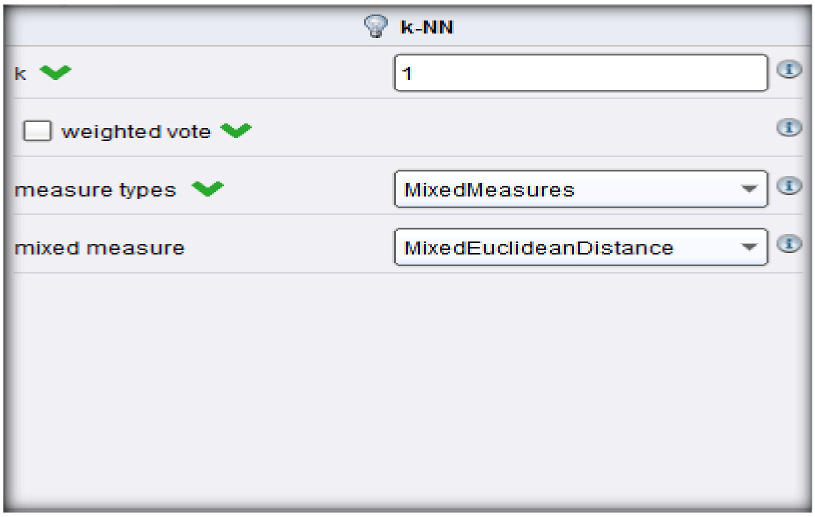

Figure (5.8): K-Nearest Neighbour parameters.

Table (5.7): Confusion matrix table for hotel and book domain using K-NN classifier with k=1.

Domain	N=2000	Positive	Negative
Hotel	Positive	787	338
	Negative	213	662
Book	Positive	829	306
	Negative	171	694

Table (5.8): Accuracy, Precision, Recall and F-Measure for hotel and book domain using K-NN classifier with k=1.

Domain	Accuracy	Precision	Recall	F-measure
Hotel	72.45%	69.96%	78.70%	74.07%
Book	76.15%	73.04%	82.90%	77.66%

5.3.4 Unsupervised approach

5.3.4.1 Lexicon Baseline

A simple lexicon based approach is considered as baseline in our experiments. In this approach, Arabic sentiment dictionary is taken to retrieve the polarity of all the words extracted from the review document. Then, it sums up the total number of positive and negative polarity of all the words of the document; if the total number of positive polarity is greater than total negative polarity value, then the positive polarity was assigned to the document and vice versa. Table (5.9) shows the confusion matrix for lexicon baseline classification approach. The correct positive number classified by the system is 622 in hotel domain and 673 in book domain. And, The correct negative number classified by the system is 703 in hotel domain and 765 in book domain. The false positive number in hotel and book domain are 297 and 235 respectively. The false negative number in hotel and book domain are 378 and 327 respectively. Table (5.10) shows the f-measure of 64.83% of the hotel review dataset and 70.55% of the book review dataset. We notice that f-measure result using lexicon baseline method is not applicable, because the feature extraction process doesn't exist. Also, detection opinion words based on the adjective words exists in the lexicons is not enough to determine the overall polarity of the review.

Table (5.9): Confusion matrix table for hotel and book domain using lexicon baseline.

Domain	N=2000	Positive	Negative
Hotel	Positive	622	297
	Negative	378	703
Book	Positive	673	235
	Negative	327	765

Table (5.10): Accuracy, Precision, Recall and F-Measure for hotel and book domain using lexicon baseline.

Domain	Accuracy	Precision	Recall	F-measure
Hotel	66.25%	67.68%	62.20%	64.83%
Book	71.90%	74.12%	67.30%	70.55%

5.3.4.2 Ontology Baseline

In this experimental setting, we extract the features with noun tag from the review documents; further, extracted features are matched in the ontology to select only the product features. Then, get the opinion words corresponding to the features extracted from previous step. Next, Arabic sentiment lexicon is used to get the polarity of the opinion words. Finally, the overall polarity review is determined. Table (5.11) shows the confusion matrix for ontology baseline classification approach. The correct positive number classified by the system is 810 in hotel domain and 746 in book domain. And, The correct negative number classified by the system is 713 in hotel domain and 839 in book domain. The false positive number in hotel and book domain are 287 and 161 respectively. The false negative number in hotel and book domain are 190 and 254 respectively. Table (5.12) shows the f-measure of 77.25% of hotel and 78.24% of book review dataset respectively. We notice that incorporating ontology information during the feature extraction process from the corpus improves the performance from 64.83% to 77.25% in hotel domain and in book domain from 70.55% to 78.24%.

Table (5.11): Confusion matrix table for hotel and book domain using ontology baseline.

Domain	N=2000	Positive	Negative
Hotel	**Positive**	810	287
	Negative	190	713
Book	**Positive**	746	161
	Negative	254	839

Table (5.12): Accuracy, Precision, Recall and F-Measure for hotel and book domain using ontology baseline.

Domain	Accuracy	Precision	Recall	F-measure
Hotel	78.20%	73.84%	81.00%	77.25%
Book	79.25%	82.25%	74.60%	78.24%

5.3.4.3 Ontology with Important Features

This experiment is to investigate the effect of considering the importance feature in determining the overall polarity of the review. This approach is similar to previous approach; we consider the importance of the feature by looking at the level of match of the feature in the domain specific ontology. Table (5.13) shows the confusion matrix for ontology with consider important features classification approach. The correct positive number classified by the system is 838 in hotel domain and 755 in book domain. And, The correct negative number classified by the system is 712 in hotel domain and 848 in book domain. The false positive number in hotel and book domain are 288 and 158 respectively. The false negative number in hotel and book domain are 162 and 239 respectively.

The results for this method for hotel and book datasets shown in Table (5.14). We found that our method improves the performance of the opinion mining methods by considering the importance of the features. f-measure improves from 77.25% to 78.83% for hotel review dataset, and from 78.24% to 79.18% in book review dataset. Figure (5.4) shows the sample semantic orientation of an review using our method in java tool.

Table (5.13): Confusion matrix table for hotel and book domain using ontology with consider important features.

Domain	N=2000	Positive	Negative
Hotel	Positive	838	288
	Negative	162	712
Book	Positive	755	158
	Negative	239	848

Table (5.14): Accuracy, Precision, Recall and F-Measure for hotel and book domain using ontology with consider the important features.

Domain	Accuracy	Precision	Recall	F-measure
Hotel	79.85%	74.42%	83.80%	78.83%
Book	80.15%	82.69%	75.96%	79.18%

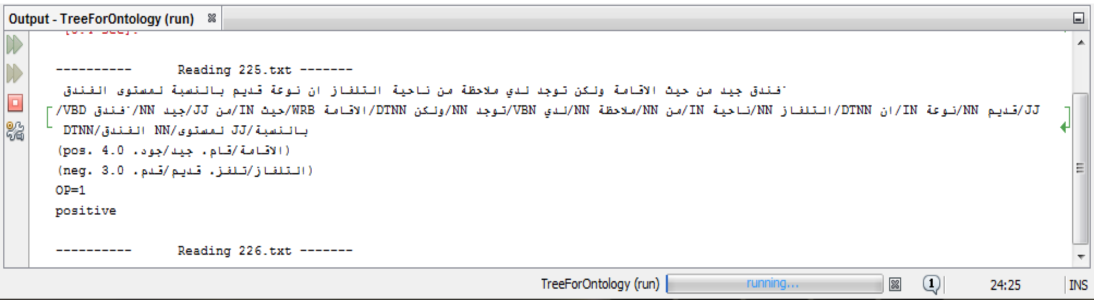

Figure (5.9): The semantic orientation of an review in java tool .

5.4 Discussing the Performance

Table (5.15) presents the f-measure results of all the methods with two hotel and book datasets. In unsupervised learning experiments, the dictionary baseline method gives the f-measure of 64.83% in hotel review dataset and 70.55% for book review dataset. Next, the f-measure is improved using ontology baseline method by incorporating domain specific ontology to get only domain related features. For example, f-measure is increased from 64.83% to 77.25% (+12.42%) for the hotel review dataset, also increased from 70.55% to 78.24% (+7.69) for book review dataset. Further, the performance improves the efficiency of the opinion mining classification by considering the importance of the features. F-measure improves from 77.25% to 78.83% (+1.58%) for hotel review dataset and from %78.24 to 79.18 %(+0.94%) for book review dataset. We conclude, by adding two factor *h* and *c* in our formula, improve the performance of opinion mining classification.

In order to evaluate the performance, we use three supervised learning algorithms such as DT, NB and K-NN classifiers. These classifiers did not give the desired results and f-measures doesn't exceed the 74.07% in hotel and 77.66% in book datasets. Figure (5.10) shows the chart of f-measures for all experiments.

Table (5.4): F-Measure (In %) of various methods on different datasets.

Method	Hotel	Book
Decision Tree	60.35%	65.09%
Naive Bayes	70.04%	72.71%
K-NN with k=1	74.07%	77.66%
Dictionary Baseline	64.83%	70.55%
Ontology Baseline	77.25%	78.24%
Ontology with important feature	78.83%	79.18%

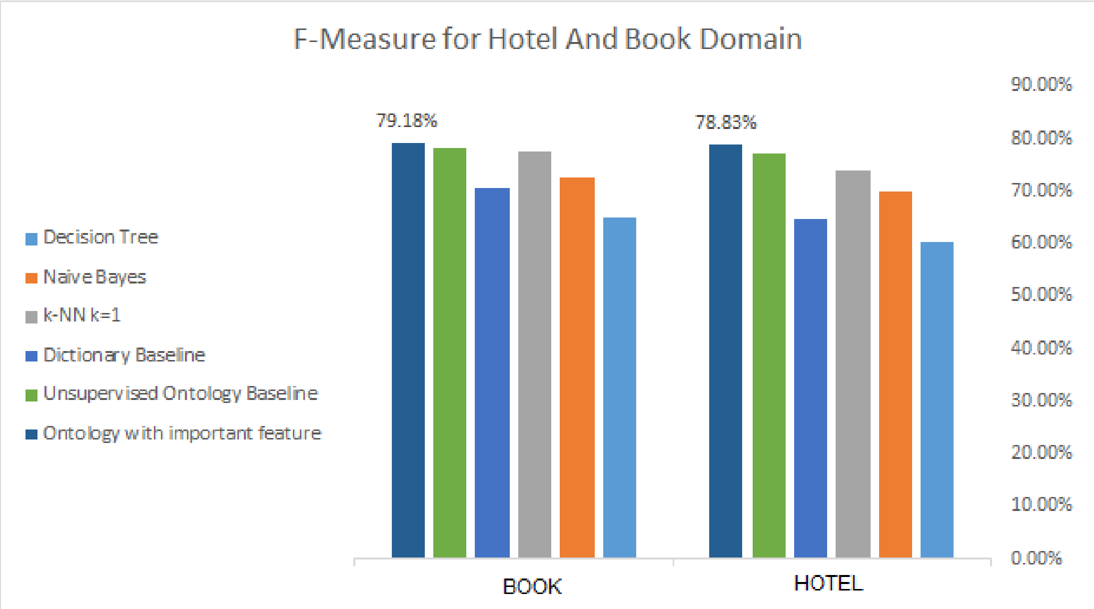

Figure (5.4): Chart to compare different opinion mining methods.

5.5 Summarization

Finally, after completing the summarization process mentioned in section (4.11), we put the summarization results into the ontology to get other advantage form the ontology tree created. The figure (5.11) and (5.12) shown an example of summary of feature in a hotel and book domain respectively.

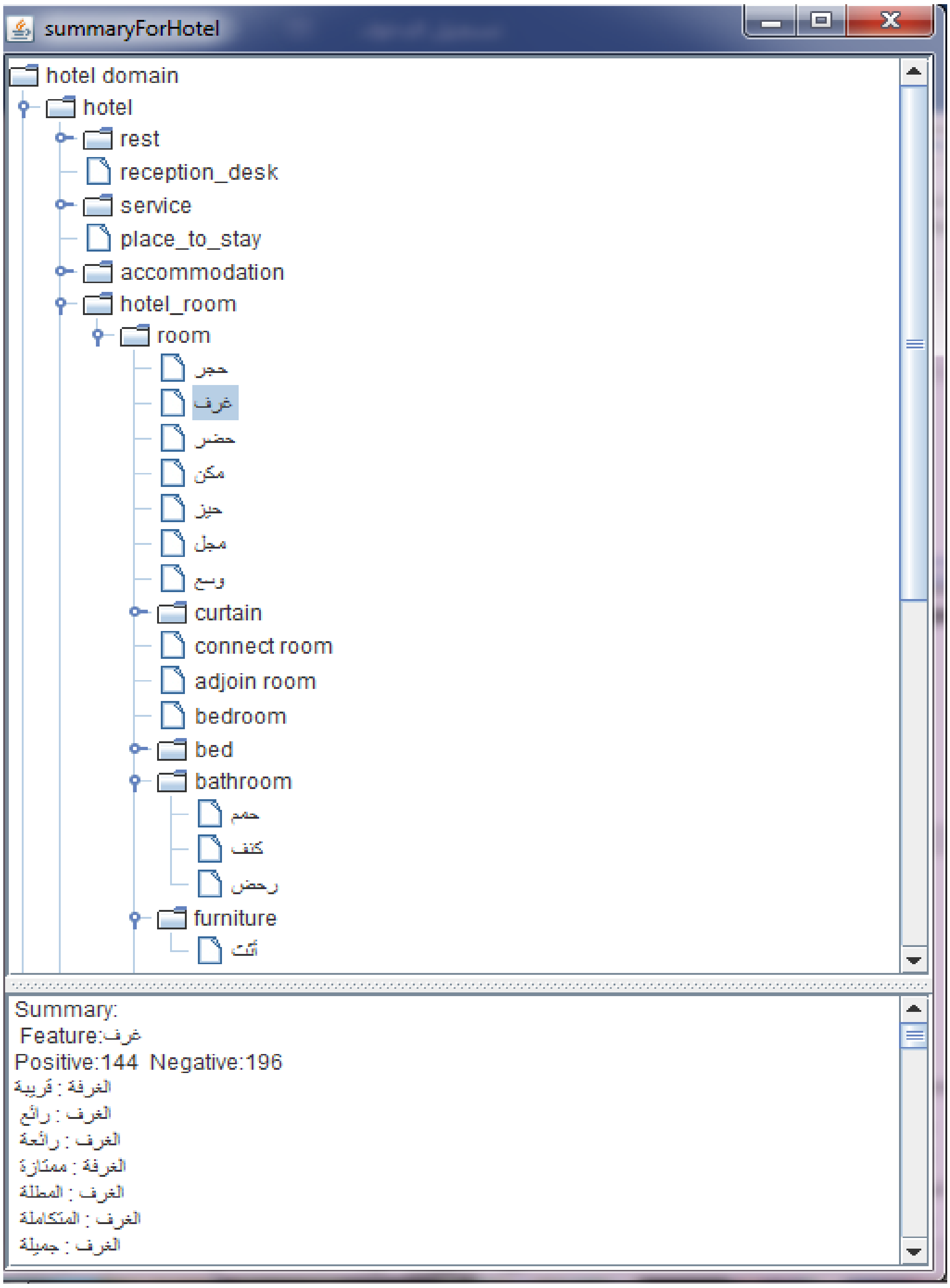

Figure (5.5): Summarization for "غرفة" feature in hotel domain.

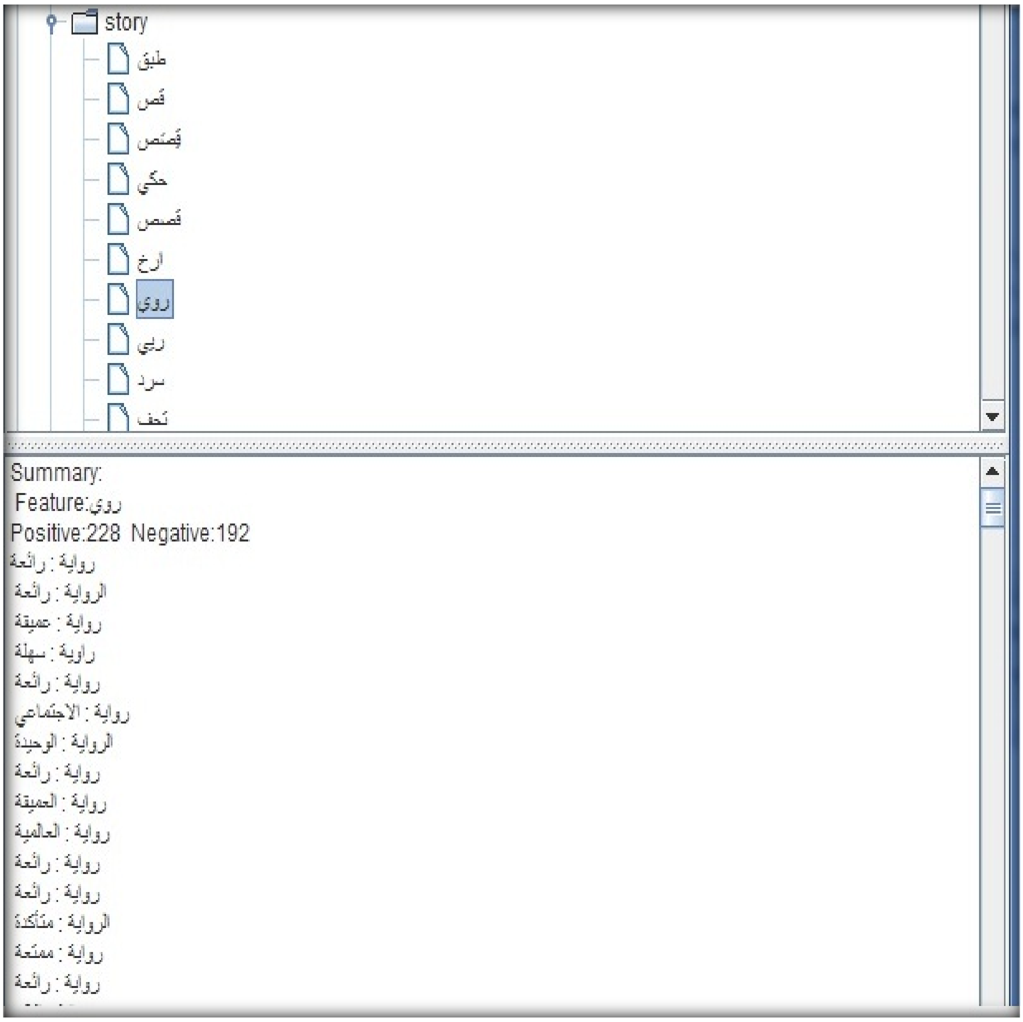

Figure 5.6): Summarization for "رواية" feature in book domain.

5.6 Subjective Evaluation

Because the result of classification is not enough, we want to know whether the system has extracted the feature correctly. Because that, we need someone have expert to evaluate feature and opinion extraction process with it's polarity.

We manually evaluate our method using recall and precision formula that shown in section (4.13). We select randomly 100 reviews for each product features in both domains and manual extract tuple (feature, opinion, polarity) for the reviews with the help of someone who has experience. We choose two product features in hotel domain, such as: "غرفة", "room" and "مطعم", "restaurant" and two product feature in book domain "الرواية", "novel and "اللغة", "language". Table (5.16) shows the review in the first column that need to extract tuple from it. In the second column present the feature extracted from the review. Opinion and polarity about feature presented in the third and fourth column respectively. The five column shows the domain name for the review.

Table (5.5): Manual extract tuple (feature, opinion, polarity) in specific domain.

Review	Feature	Opinion	Polarity	Domain
الخدمة ممتازة. الاستقبال كان جميل العاملين رائعين. الاكل ممتاز. الغرفة نظيفة انصح بالذهاب الي الاوتيل.	الغرفة	نظيفة	+	Hotel
الغرف ممتازة والديكورات تحفة	الغرفة	ممتازة	+	Hotel
الفندق ممتاز جدا من ناحية الراحة و لكن غرفته صغيرة بالكثير لشخصين في كل غرفة فقط.	الغرفة	صغيرة	-	Hotel
الذي يميز الفندق الهدوء وموقعه المتميز وسط النيل وكل المطاعم متميزة وتطل على النيل.	المطاعم	متميزة	+	Hotel
من أسوأ الأماكن في شرم الشيخ: 1. طعام ليس نظيف. 2. الأواني متسخة. 3. الغرف قديمة4 . موظفي الفندق وكأنهم لا يحصلون على مرتبات.5. الشاطئ بعيـــد للغاية. 6. أحواض السباحة متسخة.	طعام	ليس نظيف	-	Hotel
عزازيل رواية رائعة تجمع بين الفلسفة والحب والتصوف والتاريخ	رواية	رائعة	+	Book
رائعة هي هذه الرواية. تقلب الانسان وحيرته وشكوكه بين المقدس والدنيوي، التطرف الديني والغاء الأخر يكتب عنها يوسف زيدان ببراعة واتقان. أحببتها.	رواية	رائعة	+	Book
راوية سهلة رائعة تجسد المعاناة الفلسطينية بعد النكبة ومحاولة الفلسطيني سلوك كل الوسائل لخلق حياة كريمة له لكنة يواجه بظلم العدو والصديق هذه الرواية من أفضل ما كتب كنفاني اضافة لرواية عائد الى حيفا.	رواية	سهلة	+	Book
من كتر ما هي رواية مقززة مش ها كتب عنها اكتر من انه خاض في تفاصيل قذرة لا تليق بعمل ادبى.	رواية	مقززة	-	Book

Table 5.6): Recall, precision and F-measure for feature, opinion and polarity generation for two product features in hotel domain.

Features	No. of tuple extracted by Human	No. of system tuple Extracted	Correct system	Recall	Precision	F-Measure
غرفة	78	70	64	82.05%	91.43%	86.49%
مطعم	76	69	60	78.95%	86.96%	82.76%

Table (5.7): Recall, precision and F-measure for feature, opinion and polarity generation for two product features in book domain.

Features	No. of tuple extracted by Human	No. of system tuple Extracted	Correct system	Recall	Precision	F-Measure
الرواية	80	73	67	83.75%	91.78%	87.58%
اللغة	69	58	54	78.26%	93.10%	85.04%

5.6.1 Conclusion

Tables (5.17) and (5.18) shows the recall and precision result for hotel and book domain in two distinct product features. Column 1, lists of each product features. Column 2, number of tuple generated manually. Column 3, number of tuple generated by our system. Columns 4, number of correct tuple generated by our system.Columns 5 and 6 give the recall and precision of our method generation for each product feature. We notice from the results that our system has good recall and precision in predicting of features with their opinion. The average f-measure of product features in hotel and book domain is 84.62% and 86.31% respectively. Results show us the effectiveness of our method.

Chapter 6
Conclusions and Future Works

Chapter 6
Conclusions and Future Works

6.1 Conclusions

Research in opinion mining has been very limited for the Arabic language at feature-level classification. In this work, we proposed approach work at feature level opinion mining classification to detect polarity of Arabic opinion reviews. Furthermore, we combined our approach with ontology information to give better opinion mining classification performance. Using ontology in our method has several advantages, such as extract explicit product features from the review, also to determine the important features from the review, and to generate feature-based summary. Our approach is very applicable for any product domain that requires a domain name and number of level of the ontology parameter and using ConceptNet and WordNet databases to automatically construct domain specific ontology tree.

All the experiments are performed on two Arabic review datasets, namely, hotels and books. The data were collected from the websites, obtaining a total of 2000 reviews in both hotel and book domain with equal number of positive and negative reviews. We notice from our experiments that our method improves the performance over supervised and unsupervised approaches. Furthermore, we have produced summarization product reviews, to provide a feature-based summary of a large number of customer reviews. Subjective evaluation results indicate that the proposed method are very effective in feature and opinion extraction process.

6.2 Future works

In the future works, discovering methods to enrich the the ontology. Extract ontology from reviews is also a big task to deal with in future. We wan to use Web Ontology Language (OWL) to represent ontology and represent concept and feature. Also, we want to take more benefit from the ontology by expanding our ontology by more than one ontology to improve the performance. Also, we want to incorporate our method with supervised classification approaches. We need to apply our method in different domains such as mobile, computer and cars etc. Also, we want to use more than one public Arabic lexicon to improve the performance. We will try to apply the light stemming technique for our datasets. Evaluate the performance of our method used large data set. Evaluating the effectiveness of feature and opinion extraction process used more than two features.

The Reference List

Abdullah M.& Abeer M.(2016). Semantic Feature Based Arabic Opinion Mining Using Ontology. International Journal of Advanced Computer Science and Applications(ijacsa), 2016,1-7.

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),2008,1-35*.

Agarwal, B., Mittal, P., & Garg, S. (2015). Sentiment Analysis Using Common-Sense and Context Information. *Computational intelligence and neuroscience, 2015,*1-9.

Aggarwal, C. C., & Zhai, C. (2012). *Mining text data*. USA: Springer Science & Business Media.

Al Shboul, B., Al-Ayyouby, M., & Jararwehy, Y. (2015). *Multi-way sentiment classification of arabic reviews*. Paper presented at the Information and Communication Systems (ICICS), USA.

Almas, Y., & Ahmad, K. (2007). *A note on extracting 'sentiments' in financial news in English, Arabic & Urdu.* Paper presented at the The Second Workshop on Computation, al Approaches to Arabic Script-based Languages, USA.

Aly, M. A., & Atiya, A. F. (2013). *LABR: A Large Scale Arabic Book Reviews Dataset.* Paper presented at the ACL, Bulgaria.

Baziz, M., Boughanem, M., Aussenac-Gilles, N., & Chrisment, C. (2005). *Semantic cores for representing documents in IR*. Paper presented at the Proceedings of the 2005 ACM symposium on Applied computing, USA.

Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., & Fellbaum, C. (2006). *Introducing the Arabic wordnet project.* Paper presented at the Proceedings of the third international WordNet conference, Korea.

Cadilhac, A., Benamara, F., & Aussenac-Gilles, N. (2010). *Ontolexical resources for feature based opinion mining: a case-study*. Paper presented at the 23rd International conference on computational linguistics, USA.

Community, N. (2000). *NetBeans Tool*. Retrieved December 11, 2015 from: https://netbeans.org/downloads/8.0.2/

Ding, X., & Liu, B. (2010). *Resolving object and attribute coreference in opinion mining.* Paper presented at the Proceedings of the 23rd International Conference on Computational Linguistics, USA.

Duo, Z., Juan-Zi, L., & Bin, X. (2005). *Web service annotation using ontology mapping.* Paper presented at the Service-Oriented System Engineering, SOSE 2005, IEEE International Workshop, USA.

Duwairi, R. M., Marji, R., Sha'ban, N., & Rushaidat, S. (2014). *Sentiment analysis in arabic tweets.* Paper presented at the Information and communication systems (icics), 2014 5th international conference on, Jordan

Eirinaki, M., Pisal, S., & Singh, J. (2012). Feature-based opinion mining and ranking. *Journal of Computer and System Sciences, 78*(4),2012,1-20.

El-Halees, A. (2011). *Arabic opinion mining using combined classification approach*. Paper presented at the Naif Arab University for Security Sciences, KSA.

El Hadj, Y., Al-Sughayeir, I., & Al-Ansari, A. (2009). *Arabic part-of-speech tagging using the sentence structure.* Paper presented at the Proceedings of the Second International Conference on Arabic Language Resources and Tools, Egypt.

Elhawary, M., & Elfeky, M. (2010). *Mining Arabic business reviews.* Paper presented at the Data Mining Workshops (ICDMW), USA.

Esuli, A., & Sebastiani, F. (2006). *Sentiwordnet: A publicly available lexical resource for opinion mining.* Paper presented at the Proceedings of LREC, Italy.

Fariscs, I.-Q., Turki, (2012). *Khoja Stemming tool.* Retrieved December 14, 2015 from: https://sourceforge.net/projects/arabicstemmer/.

Farra, N., Challita, E., Assi, R. A., & Hajj, H. (2010). *Sentence-level and document-level sentiment mining for Arabic texts.* Paper presented at the Data Mining Workshops (ICDMW), 2010 IEEE International Conference on, USA.

Fensel, D. (2002). Ontology-based knowledge management. *Computer*, *35*(11), 56-59.

Freitas, L. A., & Vieira, R. (2013). *Ontology based feature level opinion mining for portuguese reviews.* Paper presented at the Proceedings of the 22nd international conference on World Wide Web companion, Brazil.

Ghorashi, S. H., Ibrahim, R., Noekhah, S., & Dastjerdi, N. S. (2012). A frequent pattern mining algorithm for feature extraction of customer reviews. *IJCSI International Journal of Computer Science Issues*, 1-7.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, *5*(2),1993, 1-23.

Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation. *International journal of human-computer studies*, *43*(5),1995, 1-15.

Hadni, M., Lachkar, A., & Ouatik, S. A. (2012). *A new and efficient stemming technique for Arabic Text Categorization.* Paper presented at the Multimedia Computing and Systems (ICMCS), 2012 International Conference on, Morocco.

Hammad, A. S. A. (2013). An Approach for Detecting Spam in Arabic Opinion Reviews. *The International Arab Journal of Information Technology (IAJIT), 2013*, 1-8.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques: concepts and techniques***.** Germany: Elsevier.

Havasi, C., Speer, R., & Alonso, J. (2007). *ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge.* Paper presented at the Recent advances in natural language processing, Bulgaria.

Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine learning, 11*(1),1993,1-27.

Hotho, A., Nürnberger, A., & Paaß, G. (2005). *A Brief Survey of Text Mining.* Paper presented at the Ldv Forum,1-37.

Htay, S. S., & Lynn, K. T. (2013). Extracting product features and opinion words using pattern knowledge in customer reviews. *The Scientific World Journal, 2013*, 1-18.

Hu, M., & Liu, B. (2004). *Mining and summarizing customer reviews.* Paper presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, USA.

jonsafar. (2013). *Morphological Analyzer*. Retrieved December 11, 2015 from: http://sourceforge.net/projects/aramorph/

Kaur, A., & Duhan, N. (2015). A Survey on Sentiment Analysis and Opinion Mining. International Journal For Research In Emerging Science And Technology, 2015, 1-6

Kim, H. D., Ganesan, K., Sondhi, P., & Zhai, C. (2011). Comprehensive review of opinion summarization. *International Journal of Computer Engineering and Applications, 2011*, 1-8

Lau, R. Y., Lai, C. C., Ma, J., & Li, Y. (2009). Automatic domain ontology extraction for context-sensitive opinion mining. *ICIS 2009 Proceedings*, 1-18.

Lazhar, F., & Yamina, T. G. (2012). Identification of Opinions in Arabic Texts using Ontologies. *J Inform Tech Soft Engg, 2(2)*, 1-4.

Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing, 2*, 1-39.

Liu, H., & Singh, P. (2004). ConceptNet—a practical commonsense reasoning tool-kit. *BT technology journal, 22(4), 2004*, 1-15.

Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational linguistics, 19*(2), 1993, 1-17.

Morsy, S. A. (2011). Recognizing contextual valence shifters in document-level sentiment classification. *Department Of Computer Science And Engineering, 2011*, 1-92

Mukherjee, S., & Joshi, S. (2013). *Sentiment aggregation using conceptnet ontology.* Paper presented at the Proceedings of the 6th International Joint Conference on Natural Language Processing, IJCNLP, Japan.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval, 2(1-2)*, 1-135.

Peñalver-Martinez, I., Garcia-Sanchez, F., Valencia-Garcia, R., Rodríguez-García, M. Á., Moreno, V., Fraga, A., & Sánchez-Cervantes, J. L. (2014). *Feature-based opinion mining through ontologies. Expert Systems with Applications, 2014,1-13.*

Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics, 3*(2), 2009, 14.

Princeton University. (2010). *WordNet Software*, Retrieved February 1, 2016 from: http://wordnet.princeton.edu/wordnet/license/

Qatar University, A. U. o. B., and Columbia University, . (2014). 2016. *Arabic Sentiment Lexicon database*. Retrieved February 1, 2016 from: http://oma-project.com/ArSenL/download_intro

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning, 1*(1), 1-25.

RapidMiner Studio. (2016). *RapidMiner Tools*. Retrieved February 1 ,2016 from: https://my.rapidminer.com/nexus/account/index.html#downloads

Rob Speer, a. L. c.-f. (2016). *ConceptNet Database*. Retrieved February 1, 2016 from: http://conceptnet5.media.mit.edu/

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, 2011, 1-10.*

Salloum, W., & Habash, N. (2011). *Dialectal to standard Arabic paraphrasing to improve Arabic-English statistical machine translation.* Paper presented at the Proceedings of the First Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties, USA.

SentiWordNet. (2013). *SentiWordNet Database*. Retrieved February 1, 2016 from: http://sentiwordnet.isti.cnr.it/

Stanford NLP Group. (2013). *Part-of-Speech tool*. Retrieved February 1, 2016 from: http://nlp.stanford.edu/software/lex-parser.shtml

tripadvisor (2016). *Hotel reviews*. Retrieved January 1, 2016 from: https://www.tripadvisor.com/

Turney, P. D. (2002). *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews.* Paper presented at the Proceedings of the 40th annual meeting on association for computational linguistics,USA.

Wawre, S. V., & Deshmukh, S. N. Sentiment Classification using Machine Learning Techniques. *International Journal of Science and Research (IJSR), ISSN:* 2319-7064.

Workbench, M. (2016). *Mysql Database Tool*. Retrieved February 5, 2016 from: https://dev.mysql.com/downloads/workbench/

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H.Philip, S. Y. (2008). Top 10 algorithms in data mining. *Knowledge and information systems, 14(1)*,2008, 1-37.

Yaakub, M. R., Li, Y., Algarni, A., & Peng, B. (2012). *Integration of opinion into customer analysis model.* Paper presented at the Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 03, USA.

Yang, T. (2007). Computational verb neural networks. *International Journal of Computational Cognition, 2007*, 1-5.

Zhao, L., & Li, C. (2009). *Ontology based opinion mining for movie reviews*. USA: Springer.