An Aspect-Oriented Sentiment Analysis Of Arabic Tweets Using Language Pattern and Semantic Analysis Techniques

تحليل الميول وفقاً للملامح في التغريدات باللغة العربية باستخدام أنماط اللغة وتقنيات التحليل الدلالي

A thesis submitted in fulfil of the requirements of the degree of Doctor of Philosophy (Computer Science)

By

Mnahel Ahmed Ibrahim

Supervisor
Prof. Dr. Naomie Binti Salim

August. 2018
“To Whom I love my Parents’ Souls, my Husband, Kids, Brothers and Sisters, for their continuous encouragement and support”

إلى كل الذين أحبهم إلى روح والديّ الطاهرة التي احسب أنها ترفرف في عالي الجنان وتحوم حولي تهدوني الأمان وقوة العزيزة على مواصلة ما كنا يتمناه مني (هولة الدكتوره)، إلى زوجي، بناتي، إبني، إخواني وأخواتي الذين لم يخلوا على يوما بالدعاء أهديكم جميعاً جهدي المتواضع.
ACKNOWLEDGEMENTS

Thank you to the Almighty ALLAH, for providing me with the knowledge, guidance and patience needed to achieve this goal. I wish to express my sincere appreciation to my main thesis supervisor, Prof. Dr. Naomie Binti Salim, for her encouragement, guidance and feedback.

My thanks are also extended to the generous staff of SUST, particularly Professor Izzeldin Mohammed Osman, Dr. Mohammed El hafiz Mustafa, and Dr. Talat Wahbee for their support and unlimited help.

I am grateful to all my family members, especially to my parents, for their prayers and moral support till the last moment in their lives. I am also deeply indebted to my Husband, for his continuous support and inspiration throughout the journey. He is the best companion. Special thanks to my daughters Hadeel and Raseel, and to my son Ibrahim, who inspired me to keep on striving to complete the study.

I am grateful to my many friends and colleagues who have listened to my questions, who have often provided answers and who have been patient with my idiosyncrasies, and to all those who have helped me in one way or another.
ABSTRACT

Sentiment analysis of the Arabic language has gained the attention of many researchers because of the increasing number of Arabic internet users, and the exponential growth of Arabic content online. Despite the language’s popularity, there are limited annotated resources for sentiment analysis, including a comprehensive dataset, labelled corpora, and polarity lexica, and also reliable NLP tools. This is the source of motivation for the study - the need to develop an opinion corpus for text written in Arabic. Due to the complexity of the Arabic language, opinion mining and sentiment analysis can be quite difficult to construct. The work that exists in terms of sentiment analysis is limited to news and to blogs written in Modern Standard Arabic, while there have been few studies on social media content and web reviews written in the Arabic dialect. Moreover, most of the work done has been at the document and sentence level. This is compounded by the fact that in Arabic, different forms of the same word can have a variety of suffixes, affixes, and prefixes. Furthermore, different words with different meanings can be drawn from a common three-letter root. Unlike English, which has a rich morphological structure, the Arabic language has complex, varied structures, which can be more effectively handled using natural language processing.

This thesis models an analysis of the sentiments in Arabic customer reviews, especially through Twitter. Particularly, it considers the task of aspect-oriented sentiment analysis, focusing on the two main subtasks which include (i) identifying relevant product aspects, and (ii) determining and classifying expressions of sentiment. The thesis experiments with dictionary-based methods, and several supervised approaches. For aspect detection, it casts the task as a terminology-extraction problem. With regards to sentiment analysis, detailed studies of sentiment lexicon acquisition and sentiment polarity classification are presented.

A set of Arabic language corpora from restaurants has been used to evaluate the proposed sentiment analysis methodologies. In addition, an Arabic Sentiment Classifier (TASC) has been implemented at the document-level which yield higher accuracy (88.00%) for SVM classifier. Feature selection, using Ontology and Information Gain, has also been used. The aspect-based Arabic Sentiment Analyzer (TASA) framework takes a collection of review texts, where the task’s goal is to detect individual aspects that reviewers have commented on, and deciding whether the comments are positive or negative. For purposes of aspect selection, a hybrid approach is proposed, which combines the existing information gain technique (association rules), with the ontology base technique. The existing approach basically extracts frequent aspects of text. However, analysis shows that not all aspects occur frequently in the texts. Therefore, the hybrid technique has been proposed as a means for extracting both frequent and infrequent aspects. The proposed approach employs an
incremental technique for improving the performance of existing aspect selection, by extracting infrequent features through ontology with frequent features based on the lexical dictionary. The results show that the hybrid proposed technique outperforms existing techniques; which yield (75.9%) accuracy.

The approaches used in this thesis have shown significant improvements, in comparison to relevant state-of-the-art methods, such as the lexicon-based or dictionary-based approach. It can be concluded that customer review mining systems can benefit from the methods proposed in this thesis.
المستخلاص

حُظي تحليل الميول للغة العربية باهتمام العديد من الباحثين نظراً للتزايد المطرد
ل المستخدمي الإنترنت باللغة العربية، والنمو المتسارع للمحتوى العربي على الإنترنت. و على
الرغم من أن اللغة العربية لغة شائعة الاستخدام ، إلا أن مواردها لتحليل الميول محدودة، وهذه
الموارد تتمثل في مجموع بيانات شاملة (comprehensive dataset)، والمجموع الموضحه
(polarity lexica) و المعاجم ذات القطبية (labelled corpora) الطبيعية المؤثرة بها (NLP)، وهذا يمثل الحافز الرئيسي لإجراء هذه الدراسة -بالإضافة إلى
الحاجة إلى تطوير مجموع نصي للنصوص المكتوبة باللغة العربية. نتيجة لأن اللغة العربية معقدة،
يمكن أن يكون تنقيح الآراء وتحليل الميول عملية باللغة الصعبة. إذا فقد اقتصر العمل في هذا
المجال حالياً على الأخبار والمدونات المكتوبة باللغة العربية الفصحى الحديثة في حين أن هناك
القليل من الدراسات حول محتوى الوسائط الاجتماعية ومراجعات الشبكة الالكترونية المكتوبة
باللغة العربية. علاوة على ذلك فإن معظم العمل المنجز تم على مستوى المستند و الفلم. ومما
يضاعف الصعوبات في التعامل مع اللغة العربية، أنه يمكن أن يكون لكلمة الواحدة عدة أشكال
مختلفة و أن يضاف لجذر الكلمة ملحقات قليلة وبعيدة كما يمكن أيضاً استخلاص كلمات مختلفة
ذات معان مختلفة من جذر مشترك مكون من ثلاثة أحرف . بخلاف اللغة الإنجليزية، التي لديها
بنية مورفولوجية غنية، فإن اللغة العربية لديها هياكل معقدة ومتروعة، والتي يمكن معالجتها
بطريقة أكثر فاعلية باستخدام معالجة اللغات الطبيعية (NLP).

تهدف هذه الرسالة إلى تحليل نماذج الميول في مراجعات العملاء باللغة العربية والتي
تحمل أفكارهم ومشاعرهم، لا سيما من خلال توتر. وينظر على وجه الخصوص في مهمة تحليل
الميول وفقًا للملامح، مع التركيز على المهمتين الفرعيتين الرئيسيتين اللتين تشملان (1) تحديد
ملامح المنتجات ذات الصلة، و (2) تحديد وتصنيف أشكال التعبير عن الميول وفقًا للطرق التي
تستند على القاموس، والعديد من الأساليب الإشارافية الأخرى. إذ أن المهمة تتطلب في إيجاد حل لمشكلة استخراج المصطلحات التي تتعلق بتحليل المعنوي وكيفية اكتسابها وكذلك تقسيماتها المختلفة. وتم تقديم دراسات تفصيلية حول بناء المعجم ومصنيف قطبية المعنوي.

تم تجميع واستخدام مجموعة من مفردات اللغة العربية المستخدمة في المطاعم لتقييم مهنيات تحليل المعنوي المقترحة. بالإضافة إلى ذلك، تم تطبيق نموذج تصنيف الإراءة باللغة العربية (TASC) على مستوى المستوى والذي أعطى دقة أعلى (88.00٪) لمصنف SVM. تم أيضا استخدام (Ontology and Information Gain) في اختيار الخصائص. يأخذ إطار تحليل المعنوي العربي الهجين وفقًا للملاحج (HASA) مجموعة من النصوص الاستعراضية، و التي تهدف إلى الكشف عن الملاحج الفردية التي علق عليها المراجعون، وتحديد ما إذا كانت التعليقات إيجابية أم سلبية. لأغراض اختيار الملاحج، تم إقتراح نهج هجين، والذي يجمع بين تقنية كسب المعلومات (Information Gain) وقواعد الوبائط بينها، مع تقنية قاعدة الأنطولوجيا (Ontology). النهج الحالي يستخرج ملاحج متعددة من النص. ومع ذلك، يظهر التحليل أن ليس كل هذه الملاحج تحدث بشكل متكرر في النصوص. لذلك، تم إقتراح التقنية الهجينة كوسيلة لاستخلاص الملاحج المتكررة وغير المتكررة. يستخدم النهج المقترح أسلوبًا إضافيًا لتحسين أداء التحليل الحالي، عن طريق استخلاص ميزات غير متكررة من خلال قاعدة الأنطولوجيا (lexical dictionary). أظهرت النتائج أن التقنية الهجينة المقترحة تتفوق على التقنيات الحالية بدقه تصل إلى 75.9٪.

وقد أظهرت الأساليب المستخدمة في هذه الرسالة تحسنات كبيرة، بالمقارنة مع الأساليب الحديثة ذات الصلة، مثل النهج القائم على القاموس المعجمي (Lexicon-Based) أو المعجم (Dictionary). يمكن أن نستنتج أن أنظمة التقنيب عن أراء العملاء يمكن أن تستفيد من الأساليب المقترحة في هذه الرسالة.
# LIST OF PUBLICATIONS

<table>
<thead>
<tr>
<th>#</th>
<th>Paper Title</th>
<th>Journal</th>
<th>Publication Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>A Hybrid Pattern Extraction Algorithm Based on Semantic Relations and Context Dependency</td>
<td>Int. J. of Data Mining and Bioinformatics</td>
<td>Submitted for publication (under review) May 2018</td>
</tr>
<tr>
<td>5</td>
<td>Ontology based Aspect Extraction of Arabic Restaurants reviews</td>
<td>-</td>
<td>A paper under preparation</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS........................................................................................................ III  
ABSTRACT ............................................................................................................................ IV  
المستخلص ............................................................................................................................ VI  
LIST OF PUBLICATIONS ..................................................................................................... VIII  
TABLE OF CONTENTS ........................................................................................................ IX  
LIST OF TABLES .................................................................................................................... XIV  
LIST OF FIGURES ................................................................................................................ XVI  
LIST OF ABBREVIATIONS .................................................................................................... XVII  
LIST OF APPENDICES ......................................................................................................... XVIII  
CHAPTER ONE ..................................................................................................................... 1  
1 INTRODUCTION ................................................................................................................. 1  
1.1 Overview ......................................................................................................................... 2  
1.2 Background of the Problem .............................................................................................. 5  
1.3 Problem Statement .......................................................................................................... 9  
1.4 Research Question .......................................................................................................... 10  
1.5 Research Goal ................................................................................................................ 11  
1.6 Research Objectives ....................................................................................................... 11  
1.7 Research Scope ............................................................................................................. 12  
1.8 Significance of the Research ......................................................................................... 13  
1.9 Major Contributions ....................................................................................................... 14  
1.10 Thesis Outline .............................................................................................................. 15  
1.11 Summary ..................................................................................................................... 17  
CHAPTER TWO ................................................................................................................... 18  
2 LITERATURE REVIEW ..................................................................................................... 18  
2.1 Introduction to Opinion Mining ....................................................................................... 18  
2.1.1 Opinion Mining and Sentiment Analysis .................................................................. 20  
2.1.2 Subjectivity Analysis ................................................................................................. 21  
2.1.3 Sentiment Classification .......................................................................................... 23  
2.1.4 Sentiment Lexicon Construction ............................................................................. 33
5.2 Aspects Identification Techniques ................................................................. 121
  5.2.1 The Supervised Learning Approach .......................................................... 122
  5.2.2 The Unsupervised Learning Approach ....................................................... 123
5.3 Aspect Opinion Mining Challenges ............................................................... 125
5.4 Aspect Extraction and Engineering ............................................................... 126
  5.4.1 Term Presence vs. Frequency .................................................................. 126
  5.4.2 Parts-of-Speech ....................................................................................... 127
  5.4.3 Term Position .......................................................................................... 127
  5.4.4 Negation .................................................................................................. 127
  5.4.5 Syntactic Dependency Tree Patterns ......................................................... 128
  5.4.6 Opinion and Emotion Words and Phrases ................................................. 128
  5.4.7 Valence Shifter ....................................................................................... 129
  5.4.8 Comparative Terms and Phrases ............................................................... 129
5.5 Selection of Candidate Aspects using Linguistic Patterns ............................ 129
  5.5.1 Identification of Aspect Boundaries for Patterns ....................................... 131
  5.5.2 Issues and Limitations of Existing Approaches ......................................... 131
  5.5.3 FP-growth Algorithm .............................................................................. 132
5.6 Hybrid Pattern Framework ............................................................................ 133
5.7 Patterns for Aspect Candidate Selection ....................................................... 135
  5.7.1 Linking Verb Based Patterns ................................................................... 136
  5.7.2 Preposition Based Noun Phrase (iBNP) .................................................... 138
  5.7.3 Subjective Base Noun Phrase (sBNP) ....................................................... 138
5.8 Candidate Aspect Selection ........................................................................... 139
  5.8.1 Pre-processing ......................................................................................... 139
  5.8.2 The Candidate Selection Algorithm ......................................................... 140
  5.8.3 Grouping Candidate Aspects ................................................................... 141
5.9 Experiments and Results ............................................................................... 144
  5.9.1 Results .................................................................................................... 149
5.10 Discussion ..................................................................................................... 151
5.11 Summary ..................................................................................................... 152
CHAPTER SIX ..................................................................................................... 154
6 ONTOLOGY-BASED CANDIDATE FEATURES SELECTION .......................... 154
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Previous Studies Regarding Document-level Sentiment Classification</td>
<td>25</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>The Most Widely-Used Features in Sentiment Classification</td>
<td>27</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Machine Learning Techniques vs. Lexicon-based Techniques</td>
<td>32</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>Summary of Lexicon Creating Methods Discussed in the Literature</td>
<td>39</td>
</tr>
<tr>
<td>Table 2.5</td>
<td>Aspect Extraction Techniques</td>
<td>41</td>
</tr>
<tr>
<td>Table 2.6</td>
<td>Aspect Extraction Approaches</td>
<td>46</td>
</tr>
<tr>
<td>Table 2.7</td>
<td>Summary of the Aspect Extraction Method Strength and Limitations</td>
<td>49</td>
</tr>
<tr>
<td>Table 2.8</td>
<td>The Evaluation of Sentiment Classification Techniques</td>
<td>52</td>
</tr>
<tr>
<td>Table 2.9</td>
<td>Summary of Arabic Lexicons</td>
<td>67</td>
</tr>
<tr>
<td>Table 2.10</td>
<td>Normalization Rules</td>
<td>69</td>
</tr>
<tr>
<td>Table 2.11</td>
<td>Variety of Word Spellings in Arabic Dialects</td>
<td>72</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Samples of Positive and Negative Tweets</td>
<td>82</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Qaym Dataset Statistics</td>
<td>83</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Trip Advisor Dataset Statistics</td>
<td>84</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>Samples of Positive and Negative Reviews</td>
<td>87</td>
</tr>
<tr>
<td>Table 3.5</td>
<td>The Confusion Matrix</td>
<td>89</td>
</tr>
<tr>
<td>Table 3.6</td>
<td>Dialect Words and their Corresponding MSA Words</td>
<td>95</td>
</tr>
<tr>
<td>Table 3.7</td>
<td>Dialect Words and Their Orientation</td>
<td>95</td>
</tr>
<tr>
<td>Table 3.8</td>
<td>Review of the Research Methodology</td>
<td>99</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>ML Results Without Pre-processing</td>
<td>112</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Accuracy Results of Different Classifiers</td>
<td>113</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Precision Results of Different Classifiers</td>
<td>114</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Recall Results of Different Classifiers</td>
<td>115</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>F-Measure Results of the Different Classifiers</td>
<td>115</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Dependency Relations for Nominal Sentences</td>
<td>124</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Features Extracted from Tweet Corpus</td>
<td>130</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>POS Categories and Their Abbreviations</td>
<td>135</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>Examples of cBNP Patterns</td>
<td>141</td>
</tr>
<tr>
<td>Table 5.5</td>
<td>Aspect Extraction Using Association Rules Without N-grams</td>
<td>143</td>
</tr>
<tr>
<td>Table 5.6</td>
<td>Aspect Extraction using Association Rules with N-grams</td>
<td>143</td>
</tr>
<tr>
<td>Table 5.7</td>
<td>Set of Aspects with the Same Meaning</td>
<td>143</td>
</tr>
<tr>
<td>Table 5.8</td>
<td>Evaluation of the Extracted Aspects</td>
<td>143</td>
</tr>
<tr>
<td>Table 5.9</td>
<td>The Comparative Results of Pattern-based Candidate Selection</td>
<td>150</td>
</tr>
<tr>
<td>Table 6.1</td>
<td>The Number of Positive and Negative Classes within the Resource</td>
<td>156</td>
</tr>
<tr>
<td>Table 6.2</td>
<td>Concepts and Relations in the ConceptNet Database, for the Restaurant Domain</td>
<td>159</td>
</tr>
</tbody>
</table>
Table 6.3: Examples of Manual Labelling ................................................................. 161
Table 6.4: Extracted Aspects Resulting from the use of Ontology and Patterns ........ 166
Table 6.5: Results for Sentiment Analysis .............................................................. 167
Table 6.6: Examples of the Extracted Aspects Using Ontology ............................. 167
Table 7.1: SO Results Without Pre-processing ...................................................... 175
Table 7.2: SO Results Using Normalized Tweets .................................................... 176
Table 7.3: SO Results Using a Normal Stemmer ................................................... 177
Table 7.4: SO Results Using a Light Stemmer ...................................................... 177
Table 7.5: Results of Using CRF ............................................................................ 179
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>Different Challenging Problems in Opinion Mining and Sentiment Analysis (Liu, 2012)</td>
<td>21</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Example of Customer Review Text Documents Used in Sentiment Classification (<a href="https://ar.tripadvisor.com">https://ar.tripadvisor.com</a>)</td>
<td>24</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>The Taxonomy of Document-level Sentiment Classification Techniques</td>
<td>28</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Aspect-based Sentiment Analysis</td>
<td>42</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Operational Research Frameworks</td>
<td>79</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Proposed HASA Frameworks</td>
<td>80</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Support Vector Machines Classifier</td>
<td>105</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>NB Classifier</td>
<td>106</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>KNN classifier</td>
<td>108</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Method of TASC</td>
<td>110</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Summary of the Accuracy Results of Different Classifiers</td>
<td>114</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Summary of the Precision Results of Different Classifiers</td>
<td>114</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Summary of the Recall Results of Different Classifiers</td>
<td>115</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Summary of the F-Measure Results of different classifiers</td>
<td>116</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Method of Aspect Extraction</td>
<td>137</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Sample POS Tagged Text</td>
<td>140</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>Algorithm for Candidate Selection(Khan et al., 2014)</td>
<td>148</td>
</tr>
<tr>
<td>Figure 5.4</td>
<td>Pattern-based Candidate Selection, for Performance Measures</td>
<td>151</td>
</tr>
<tr>
<td>Figure 5.5</td>
<td>Summary of Average Pattern for the Two Corpora</td>
<td>151</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>Sample Ontology for the Restaurant Domain</td>
<td>158</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>Restaurant Ontology Class Hierarchical</td>
<td>160</td>
</tr>
<tr>
<td>Figure 6.3</td>
<td>Data Property Hierarchical of the Restaurant Ontology</td>
<td>160</td>
</tr>
<tr>
<td>Figure 6.4</td>
<td>Algorithm for Creation Ontology Tree</td>
<td>163</td>
</tr>
<tr>
<td>Figure 6.5</td>
<td>Rules Used to Detect the Adjective and Verb Opinion Words Related to the Feature</td>
<td>165</td>
</tr>
<tr>
<td>Figure 6.6</td>
<td>Summary of Aspects Results Using Ontology and Patterns</td>
<td>167</td>
</tr>
<tr>
<td>Figure 7.1</td>
<td>Hybrid Arabic Sentiment Analyzer (HASA)</td>
<td>171</td>
</tr>
<tr>
<td>Figure 7.2</td>
<td>Example of Polarity Identification</td>
<td>180</td>
</tr>
<tr>
<td>Figure 7.3</td>
<td>Tweet Target Identification</td>
<td>181</td>
</tr>
<tr>
<td>Figure 7.4</td>
<td>An Example Summary of the Feature “ السعر”</td>
<td>181</td>
</tr>
<tr>
<td>Figure 7.5</td>
<td>Handling the Negation Word</td>
<td>181</td>
</tr>
</tbody>
</table>
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HASA</td>
<td>Hybrid Arabic Sentiment Analyzer</td>
</tr>
<tr>
<td>TASC</td>
<td>Twitter Arabic Sentiment Classifier</td>
</tr>
<tr>
<td>ArSenL</td>
<td>Arabic Sentiment Lexicon</td>
</tr>
<tr>
<td>MSA</td>
<td>Modern Standard Arabic</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>SO</td>
<td>Semantic Orientation</td>
</tr>
<tr>
<td>TDASA</td>
<td>Twitter Dataset for Arabic Sentiment Analysis</td>
</tr>
<tr>
<td>OWL</td>
<td>Web Ontology Language</td>
</tr>
<tr>
<td>NLP</td>
<td>Natural Language Processing</td>
</tr>
<tr>
<td>OM</td>
<td>Opinion Mining</td>
</tr>
<tr>
<td>BOW</td>
<td>Bag-of-Words’</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>KNN</td>
<td>K-Nearest Neighbor</td>
</tr>
<tr>
<td>POS</td>
<td>Parts-Of-Speech Tagging</td>
</tr>
<tr>
<td>TF-IDF</td>
<td>Term Frequency - Inverse Document Frequency</td>
</tr>
<tr>
<td>DA</td>
<td>Dialectal Arabic</td>
</tr>
<tr>
<td>IR</td>
<td>Information Retrieval</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Networks</td>
</tr>
<tr>
<td>FP-Growth</td>
<td>Frequent Pattern Tree Algorithm</td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Samples of tweets and reviews used in classification and manual annotation</td>
<td>199</td>
</tr>
<tr>
<td>B</td>
<td>Samples from the list of stop words used in the pre-processing stage</td>
<td>207</td>
</tr>
<tr>
<td>C</td>
<td>Samples from the lists of sentiment words</td>
<td>207</td>
</tr>
<tr>
<td>D</td>
<td>Samples from the lists of negation and intensifier words</td>
<td>208</td>
</tr>
<tr>
<td>E</td>
<td>Traditional Arabic grammar</td>
<td>209</td>
</tr>
</tbody>
</table>
CHAPTER ONE

1 INTRODUCTION

The tremendous growth of the internet, along with the number of social platforms and blogs, has provided avenues for millions of people to post comments and opinions on a wide range of subjects, including food, electronics, automobiles, politics, movies, and many others. The huge amount of opinions available online have piqued the interest of companies and government organizations, as well as people who are interested in using them for their own personal benefit. The growing popularity of e-commerce has encouraged more people to purchase products from online stores, where choices are made based on reviews and comments of others who have purchased similar products (Rushdi-Saleh et al., 2011). Companies and governments have an interest in obtaining the public’s opinions. Companies require user opinions to improve and market products and services, while governments require public perspectives in order to set up or update rules, policies and regulations. After the information technology revolution, researchers have developed an immense interest in the analysis of people’s opinions, particularly on social networks.

In this chapter, section 1.1 provides an overview of the study. Section 1.2 describes the background of the problem chosen for this study. Section 1.3 then gives the problem statement, followed by section 1.4 which presents the research question. Section 1.5 gives a description of the research goal, and section 1.6 lists the research objectives. Section 1.7 highlights the research scope, while section 1.8 gives a description of the research’s significance. Section 1.9 presents the expected research contributions. Section 1.10 outlines the thesis, and finally, section 1.11 summarizes the chapter.
1.1 Overview

There has been a boom in internet communication in recent years, especially through the use of social media. With the introduction of Web2.0, and the availability of a wide range of social media platforms, such as Facebook or Google+, online review websites, personal blogs, microblogging services (e.g., Twitter), photo sharing apps (e.g., Instagram), and professional networking tools (e.g., LinkedIn), an environment has been created in which public opinions about upcoming events, products, companies, and even trending news, are all highly valued.

Scanning opinion sites and summarising the information gathered has become a challenge, owing to the tremendous growth in the number of websites online. A huge volume of opinionated text is available on each site, but its analysis and summarising into a useful format remains a difficult process.

The researchers’ interest in the processing and analysing of people’s comments and reviews regarding a variety of topics, has led to the development of an automated opinion identification tool for documents, sentences or phrases (Liu, 2010). This field of research is relatively new, and is supported through machine learning (ML), computational linguistics, and natural language processing (NLP). Keywords encountered in this research include sentiment analysis, sentiment orientation, subjectivity analysis, and opinion mining (OM).

Traditional Natural Language Processing (NLP) applications have mostly focused on topical text characterization, which deals with communicated facts and the objective presentation of information. Recent research by the natural language community has shown the benefit of analysing emotions and opinions shared in public texts, otherwise referred to as the process of opinion mining. Opinion mining refers to a technique in which meanings are automatically extracted from opinions put forth in text (Liu, 2012).

Text sentiment classification is a term which refers to the determination of a text’s sentiment polarity, be it either positive or negative (Liu and Zhang, 2012).
Sentiment classification has become a subject of discussion within the natural language processing research community, due to its usefulness in classifying online product reviews and in summarizing opinions (Kang et al., 2012; Ku et al., 2006).

Sentiment classification methods fall into two main groups, which include lexicon-based classifications, and machine-learning based classifications. Lexicon-based classification methods analyse the words and phrases in a text, and help determine its polarity accordingly. If the positive terms outnumber the negative terms, then the document is classified as being positive, and vice versa, if the negative terms outnumber the positive ones (Turney, 2002; Taboada et al., 2011). A sentiment lexicon is needed for the deterring the polarity of each term. Machine learning-based classification, on the other hand, involves the use of machine learning classification algorithms for training a sentiment classifier, using labelled data (Pang et al., 2002; Moraes et al., 2013). The classification’s accuracy is based on the quality and quantity of the labelled data used for the training set. Based on these two method groups, sentiment lexicons and annotated sentiment data can be seen as being the most important resources for sentiment classification.

These classification methods have been applied in English language processing over the past few years, according to Pang et al. (2002) and Turney (2002). However, this processing has not been effectively applied to other languages, such as Arabic.

Recent research has looked into the large-scale sentiment analysis of comments posted on micro-blogging services, such as Twitter. The processing of sentiment analysis varies depending on the scenario, and the relevant tasks and subtasks. In addition, the complexity and methods used to carry out the analysis depend on the nature of the text. For example, formal language might be used in newswires, while informal language may be used on micro-blogging sites.

This research is concerned with aspect-based sentiment analysis of the Arabic language. It aims to classify aspects as being positive, negative or neutral sentiments, irrespective of their source (blog, review or tweet). The scope of this study is limited to tweets written in the Arabic language. The study entails carrying out a comprehensive
overview of mining, after which a detailed search is performed for the purposes of aspect-oriented sentiment analysis. The main goal of this research is the development of an automatic sentiment assessment and classification method, applicable to the individual aspects of a product.

If you consider a query to find the satisfaction level of customers who visit a certain restaurant to be based on the quality of meals, and the complaints made about the service, a sentiment analysis system should be able to produce a detailed summary of customers’ opinions. This is done in a manner which ensures there is no information overload.

The most common approach to carrying out sentiment analysis involves supervised machine learning and rule/dictionary-based approaches, both requiring significant manual effort for establishing an adequate system. Supervised machine learning requires the creation of labelled corpora, while rule/dictionary-based systems depend on the availability of comprehensive lexicons, and the manual fine-tuning of rule sets. Recent sentiment analysis of Arabic language has gained the attention of many researchers, due to the increasing number of Arabic internet users, and the exponential growth of Arabic content online. After reviewing previous research in this area, this study proposes a hybrid model of Arabic tweet sentiment analysis, which presents improved sentiment classification.

Rankings issued by the Internet World State\(^1\) (2017), have shown that Arabic is one of the 10 most popular languages on the Internet, one spoken by hundreds of millions of people worldwide. Despite the language’s popularity, there are limited annotated resources for sentiment analysis, for instance labelled corpora and polarity lexica, which act as the source of motivation behind this study for developing an opinion corpus for text written in Arabic.

The interest of research in the mining of text, and the retrieval of information in

\(^1\) [http://www.internetworldstats.com/stats7.htm](http://www.internetworldstats.com/stats7.htm)
the Arabic language, has resulted in the accumulation of resources, corpora and text classification tools (Kuwaiti et al., 2009), and also name entity recognition (Shaalan and Raza, 2009). However, due to the complexity of the Arabic language, related opinion mining and sentiment analysis are quite difficult to undertake (Al-Shalabi and Obeidat, 2008). This is compounded by the fact that in Arabic, different forms of the same word may have a variety of suffixes, affixes, and prefixes. Furthermore, different words with different meanings can be created from the same three-letter root.

The Arabic language is used by over 388 million people in more than 22 countries all over the world (2018). Unlike English, which has rich morphological structure, the Arabic language has complex, varied structures which can be more suitably handled through natural language processing. Some of the basic tools developed for Natural Language Processing (NLP), in the Arabic language, include the morphological analyser, speech tagger, and syntactical parser (3). Farghaly and Shaalan (2009) have stated that NLP in the Arabic language is still in its infancy.

This study focuses on an analysis of sentiments in the Arabic language, due to the wide-scale use of the Arabic language around the world, its online usage, its history, and its rich cultural and strategic value. In addition to this, there are limitations in NLP regarding the language and the tools employed.

1.2 Background of the Problem

User reviews are valuable in the fields of business, education, and especially in e-commerce, because customers’ online purchasing decisions are based on these reviews. Companies also use these reviews to obtain feedback about customers’ evaluation of their products. The textual information here can either be categorised as

2 http://www.internetworldstats.com/stats7.htm
3 https://nlp.stanford.edu/links/statnlp.html
Facts focus on the objective transmission of data, while opinions focus on the expression of thoughts or sentiments about a subject or item. Sentiment analysis determines the tone of the document’s content, and a person’s attitude towards a certain topic. This attitude may relate to the judgement passed, its emotional state or its related effect. The key task in sentiment analysis is determining the polarity of the document’s content, in its entirety, at the sentence and phrase levels (Michelle, 2010).

Research work on sentiment analysis began in the early 2000s, with initial studies conducted by Pang et al. (2002) who applied supervised machine learning, and Turney (2002) who proposed a lexicon-based method. Supervised learning relies on a large set of labelled data used to train a classifier, which is used afterwards to determine the polarity of unlabelled test data. Most of the existing studies handle sentiment classification as a supervised classification problem (Pang et al., 2002, Riloff et al., 2006; Prabowo and Thelwall, 2009; Ye et al., 2009; Zhang et al., 2009; Kang et al., 2012). In supervised methods, different feature sets are considered, and various feature selection techniques are used to increase the performance of sentiment classification. The ‘Bag-of-Words’ (BOW) approach is the most popular technique used for text representation in sentiment classification (Pang et al., 2002; Wang et al., 2014). The major shortcoming of the supervised methods is the difficulty faced in implementing and annotating large amounts of labelled training data.

At the same time, research on sentiment lexicons has been undertaken, classifying documents based on the sentiments in them. The establishment of the sentiments in a document is done through the calculation of the sentiment orientation of words in a document, with the help of a dictionary or search engine which associates words with a known polarity seed set (Turney, 2002; Harb et al., 2008). These falls into the category of lexicon-based methods of sentiment analysis, since the analysis results are dependent on sentiment lexicons.

Sentiment classification requires two essential resources, particularly labelled corpus and sentiment lexicons. Few recent studies have delved into the sentiment
classification of other languages, such as Arabic. Other researchers have studied aspect-oriented sentiment analysis, which not only analyses sentiments but also evaluates them based on specific product aspects.

Although a lot of research has been conducted in this area of sentiment analysis and opinion mining, there are still limitations which relate to accuracy, scalability, quality, data standard, and difficulties faced in understanding natural language. The main challenges encountered in the use of NLP, which makes OM difficult, are context dependency, semantic relatedness, and the ambiguity of conclusions drawn. Given that accuracy is important in real-life practical applications, some of the work still needs to be carried out manually. Private Blogs provide useful sources of data for OM. However the posts are written using informal language and are diverse, which makes analysis difficult.

Opinions are collected from the World Wide Web, in order to facilitate the OM process. Given that the web contains a very large, diverse collection of information, which is also multi-dimensional and at times redundant, opinions are gathered only from specific websites or from several websites as and when needed (on an ad-hoc basis), so that large-scale opinions can be afterwards processed. A great deal of research on opinion mining is domain-dependent. Machine learning systems, for instance, are domain-dependent and require the manual labeling of data, thereby making it difficult to make generalizations. For this reason, generalized domain-independent algorithms are required for the automatic identification and classification of opinion components.

Another additional challenge in this research field is the acquisition of a standard data set, because there are only a few which are useful for classification, benchmarking and analysis.

Most of the previous approaches of sentiment analysis have focused on classifying data written in Modern Standard Arabic (MSA) at a document level (Abdul-Mageed and Diab M,2014; Abdul-Mageed,2011; Rushdi-Saleh et al.,2011; Abbasi et al.,2008). Few research studies look into dialectical Arabic, which includes the translation of dialectical words into corresponding MSA words (Duwairi,2015; Mourad
and Darwish, 2013; Shoukry, 2013). Most existing works in this approach have used machine translation systems to translate labelled training data from Arabic into English, and to perform sentiment classification using English datasets. Although machine translation is used extensively in the field of sentiment classification, working with translated data implies an increasing number of features, sparseness, and noise in datasets. Some other researchers have tried to deal with Arabic sentiment analysis at an aspect level (Elarnaoty et al., 2012; SemEval-2016; Al-Smadi, Qawasmeh, Talafha and Quwaider, 2015). Some researchers have used Twitter as a data source for their work for both document and aspect level sentiment analysis (Alhazmi, M., and Naomie, S., 2015; Aldayel, H.K. and Azmi, A.M., 2016; Refaee, E. and Rieser, V., 2014; Al-Horaibi, L., & Khan, M. B., 2016; Kuwaiti, R. M., and Qarqaz, I., 2016).

Although recent research works have tried to overcome some problems faced in Arabic sentiment classification, there are still several research gaps in this area of research which have not been considered in the literature. These gaps can be summarized as considering translation errors, information loss during this process, the problem of the scarcity of Arabic Dialects resources, and the need to deal with the problems of low accuracy and precision, and aspect extraction, when implementing aspect-based sentiment analysis.

Taking into account these gaps, this research seeks to deal with the problems of the aspect-based sentiment classification of Arabic tweets, within the umbrella of a semi-supervised learning strategy. Pure classification only gathers information for the purposes of determining the number of satisfied and unsatisfied customers. The trends for customers’ perception of a product are established based on satisfaction count, although the exact reasons for a customer’s satisfaction or dissatisfaction still remain unknown. Similarly, customer likes and dislikes are also difficult to determine. Aspect-oriented mining is the one that serves to analyse customers’ sentiments, based on the aspects of each product.

Classification only considers a single dimension, specifically sentiment polarity, while aspect-oriented mining involves a joint analysis of two dimensions. The first dimension establishes the products’ relevant aspects, while the second dimension
identifies related expressions, using them to determine the sentiment’s polarity. The problem highlighted is better classified under text categorization, rather than information extraction. For example, unstructured information of a tweet text is transformed into a structured, aspect-oriented summary.

The Arabic language is selected as the subject of the study because of several factors. First of all, Arabic sentiment analysis is gaining popularity due to its wide-scale use. Second, it is challenging and interesting to work with the Arabic language due to its history, its rich culture and strategic value. There are six dominant Arabic dialects, namely Egyptian, Moroccan, Shami, Iraqi, Gulf, and Yemeni. These dialects may vary in vocabulary, morphology, syntax and in the dictation of classical Arabic, and most of them lack correct spelling.

Internet, social media and social network websites such as Twitter⁴, Facebook⁵, and many more, are playing an active role in adopting Arabic as an informal language. In order to overcome the aforementioned problems, multiple types of semi-supervised learning models have been proposed in previous works, and will be discussed in this studies’ literature review.

1.3 Problem Statement

The main aim of this research is to overcome the problem of the aspect-based sentiment analysis of Arabic tweets.

Sentiment analysis, which can also be referred to as opinion mining, is an emerging area of research. A massive amount of text is available from online sources such as reviews, blogs and social networking websites, which can be processed through this means. Previous research in this field has only been undertaken for the English

⁴https://twitter.com/
⁵https://www.facebook.com/
language, while other languages such as Arabic and its dialects have been ignored, due to the absence of adequate resources for conducting the analysis.

Based on the review of previous literature as conducted in this study, the main issues in the area of Arabic aspect-based sentiment analysis include:

- Pre-processing methods: Most of the text pre-processing tools and methods (such as stemmers, stop-words lists, etcetera) are designed for the modern standard Arabic (MSA), excluding dialect-specific rules.
- Feature sets: The second issue is the lack of suggested feature sets and classification algorithms, which can be used in the classification of the dialect-specific text.
- Dialect-specific lexicon: The third issue is the absence of a dialect-specific lexicon, with weights for each sentiment word.
- Aspects extraction: The fourth issue is the lack of suggested methods that can be used to extract aspects from reviews written in Arabic dialect and determine its Polarity.

The combination of these issues has contributed to the limited number of studies conducted in the field of sentiment analysis for certain dialects.

1.4 Research Question

This study aims to overcome the aforementioned problems of the aspect base sentiment analysis of Arabic dialect, by answering the following research questions:

a. Which pre-processing techniques can be used to enhance the sentiment analysis of the Arabic dialect?
b. Which techniques and features can be used to identify the opinion aspect of the Arabic dialect?
c. Is the use of dialect lexicon effective in enhancing the sentiment analysis of the Arabic dialect?
d. Which technique and features can be used to enhance the polarity identification of the Arabic dialect?

1.5 Research Goal

This study aims to propose a suitable technique for extracting opinions and opinion aspects from Arabic tweets. The important factor in the extraction of opinions from any tweet, is to classify the tweet as being either subjective or objective tweet. An opinion from a tweet may hold facts which, after being correctly classified, can be used to extract opinions.

The aim of this research is to propose an aspect-based sentiment analysis model for Arabic tweets, using integrated techniques that represent a combination of suitable features for the improvement of classification performance and aspect identification. By addressing the existing problems in previous works, this research strives to design and develop learning models in the above-mentioned framework, for a specific dialect of the Arabic language. The ultimate goal is to improve the performance of machine learning methods and lexicon-based approaches, for the sentiment analysis of Arabic tweets.

1.6 Research Objectives

To meet this research’s goal, the following objectives have been identified:

1. To investigate the impact of different pre-processing techniques on sentiment classification accuracy.
2. To investigate different features and techniques that can be used to identify the opinion aspects of Arabic tweets.
3. To identify the polarity of Arabic sentences, based on their aspects and
patterns.

4. To propose a hybrid aspect extraction model for Arabic tweet sentiment analysis, and to measure the resulting enhancement in the quality of sentiment classification, if any.

In order to achieve the mentioned objectives, the approach taken starts with the development of a pre-processing mechanism for tweets, which will normalize, stem and remove stop words. The effect of this mechanism on the performance of machine learning and SO is measured. This is followed by the identification of the features used for machine learning, and potential patterns including Bag of Words, TF-IDF and N-grams. The next step involves the building of an annotated corpus for the training and validation of the most suitable classifier, while a sentiment lexicon is constructed from the corpus. Experimentation is undertaken through different combination schemes for these approaches, so as to enable selection of a scheme that utilizes the combined benefits of each approach. Based on the experiment, the best pre-processing, machine learning and association rules mining techniques are chosen and are then combined with lexicon-based approaches. This combined scheme is then proposed as being a hybrid model for the analysis of sentiments written in Arabic tweets.

1.7 Research Scope

The scope covered in this research study, in the course of solving problems associated with the aspect-based sentiment analysis of Arabic tweets, is as follows:

I. The focus of the research is on the identification of the opinion aspects of Arabic tweets.
II. This research focuses on classifying Arabic tweets regarding restaurants, based on the overall sentiment orientation of each tweet.
III. This research focuses on classifying Arabic tweets on restaurants Gulf dialect.
IV. This research only focuses on increasing the performance of machine learning methods and lexicon-based approaches within Arabic sentiment classification.
V. The content similarity of documents is used as a simple structural similarity measure, introducing the intrinsic structure of documents in the graph-based method. Other methods of introducing intrinsic structure (for instance, opinions, methods of expressing the sentiment and opinion holder characteristics) are not considered in this study.

1.8 Significance of the Research

Sentiment analysis has become a topic of interest in the past few years, with several companies using it for business applications. IBM SPSS\(^6\) is an example, providing quantitative summaries of sentiments from survey data on products which are useful in providing businesses with information regarding consumer attitudes. OpSec\(^7\) mines user-generated data from social media. Wall Street began using sentiment analysis in their trading algorithms, and the OpFine\(^8\) company provides up-to-date sentiment tracking of financial news. The American media conglomerate, The Washington Post\(^9\), uses sentiment analysis to provide social media statistics regarding political figures. Some Arabic companies use social media networking, including Twitter and Facebook, to compile customer feedback reviews about their service and products. However, to the best of the author’s knowledge, there are no published studies by these companies for learning how they benefit from or use these reviews.

A study by Pang and Lee (2008) has shown that 81% of internet users have searched for a product online at least once, and that out of this amount, 73% to 87% of them have made purchases whose choice was motivated by reviews. This indicates that the sentiment classification of reviews is useful for customer product selection, thereby

---

\(^{6}\)http://www-01.ibm.com/software/analytics/spss/

\(^{7}\)http://opsecsecurity.com/brand-protection/online-brand-rotection/sentimentanalysis

\(^{8}\)http://www.opfine.com/

\(^{9}\)http://www.washingtonpost.com/politics/mention-machine/
drawing researchers’ interest.

One of the most accurate methods introduced for sentiment classification is machine learning. However, most languages have inadequately-annotated sentiment resources for application in supervised classification, while the alternative, manual construction of labelled corpus, is challenging and takes a lot of time to implement.

This study aims to create aspect-based sentiment classification models for Arabic tweets which serve to improve classification performance. This is urgently needed in today’s sentiment analysis applications. Arabic is a native language for 290 million people around the world, with the majority of Arabic speakers generally use dialects in daily interactions, and those dialects also being generally used on the Internet and through social networking sites such as Twitter, in lieu of standard Arabic. This adds a new challenge to the existence of systems dealing with this type of text, in order to analyse sentiment.

1.9 Major Contributions

In this thesis, we introduce aspect-based sentiment analysis of Arabic tweets, and we enhance the technique using association rules in general, patterns and ontology. The contributions to this study have been grouped as follows:

1. Designing a new aspect-based sentiment classification model for Arabic tweets, based on association rules mining.

The Association Mining approach for product features extraction was first used by Hu and Liu (2004), who through their work has extracted frequent features using the association rule mining technique (Agrawal and Srikant, 1994). The algorithm’s earliest implementation was in market basket analysis, whereby the level of dependency of a sold item is measured against another item. Using this analogy, Hu and Liu (2004) assumed that words in a sentence are sold items, while the association between the terms
can be used to predict features and opinions. This method was found to be useful for feature extraction. Wei et al. (2010) later extended this approach for application within semantic-based pruning, to the refinement of frequent features and the identification of infrequent features. This refined approach showed improved results in terms of opinion target identification.

2. Enhancing the new aspect-based sentiment classification model, using noun phrase patterns.
3. The design of a novel domain ontology for restaurants, consisting of concepts and attributes associated with taxonomic and non-taxonomic relationships between them.
4. Using the domain ontology during the aspect selection stage of opinion mining, and the extraction of sentiments associated with the aspects.
5. Enhancing the new aspect-based sentiment classification model using patterns.
6. The use of a combination of these approaches, to enhance the overall model accuracy.

All the mentioned contributions will be discussed later, in detail, in the thesis chapters.

1.10 Thesis Outline

This thesis has been organized into eight chapters, as follows:

Chapter 1 gives an introduction to the research topic. After that, the research background and research problems are explained. The research questions and objectives of the research are then introduced. Finally, the importance of the research is expressed.

Chapter 2 provides background information and reviews the previous studies in this field, which is followed by the identification of research gaps, and the formulating of a research problem. Then, an overview of opinion mining studies is then presented.
The review covers the basic method of opinion lexicon creation, the types of feature-based summarization methods, and the methods that have been used in Arabic opinion mining.

Chapter 3 explains the methods and datasets used in this research, for each of the aspect-based sentiment classification approaches and detail their combination to form a hybrid approach. It covers tweets pre-processing, whereby an explanation is provided of the different types of text pre-processing applied to Arabic tweets. The research flow is then described systematically in this chapter. The evaluation metrics and evaluation framework are also explained in this chapter. The chapter overviews the design and development steps of the opinion lexicon, which is later used in HASA.

Chapter 4 covers the development process of the first proposed model for the Twitter Arabic Sentiment Classifier (TASC), which provides an overall opinion on an entity, topic or event, by using opinion mining classification at the document level. The construction of the TASC is dependent on a manually-annotated corpus. This chapter proposes a supervised machine learning technique, including naïve Bayesian, Support Vector Machines (SVM) and KNN classifiers. This model is evaluated and compared with some other baseline methods within this chapter.

Chapter 5 covers pattern-based candidate features selection. It addresses the design and development procedure for the second proposed model patterns, based on candidate selection from unstructured reviews, thereby explaining the modelling of patterns and structures. An analysis of the proposed patterns, based on the candidate selection algorithm, has been provided. This model is an enhancement of the first proposed model which uses the noun phrase. Corresponding results and evaluations are also given in this chapter.

Chapter 6 covers the ontology-based candidate features selection, describing the implementation process for the proposed model, which employs ontology as a means of developing a structured knowledge representation and a common vocabulary for a domain, for example the restaurant domain. Web Ontology Language (OWL), a W3C standard, is used to represent ontology. OWL represents the concepts and features of the
application domain, which in this case is the restaurant domain. Chapter 6 also shows the results obtained from the proposed model and compares this model’s performance with other similar existing methods.

Chapter 7 introduces the Hybrid Arabic Sentiment Analyzer (HASA), which serves the purpose of mining and summarizing opinions from customer reviews. HASA takes advantage of both frequency-based and relation-based approaches for identifying opinion sentiments, involving mining a set of aspects from frequent noun phrases in the reviewed texts, and using a novel technique for grouping synonymous aspects. In addition, it determines whether an opinion is positive or negative, and generates a related summary. Corresponding results and evaluations are also given in this chapter.

Chapter 8 concludes the research, providing a list of contributions, states the limitations of proposed models, and expresses some recommendations for future study.

The thesis concludes with five appendices at the end of the document. Appendix A shows some sample tweets from the datasets used in conducting the experiments. Appendix B shows samples from the list of stop words used in the pre-processing stage. Appendix C shows samples from the list of positive sentiment words, and samples from the list of negative sentiment words. Appendix D sample from the lists of the negation and intensifier words used. Finally, Appendix E shows the traditional Arabic grammar.

1.11 Summary

The principles of the research, and the essential parts of this study, have been introduced in this chapter. The information provided has included an overview of the research topic, the background of the research problem and the related problem statement, along with research questions, the research goal and objectives and the scope of the current research. The significance of this research is also described, in order to present its benefits. The aim of this chapter is to provide an overall description of this research’s main elements.
CHAPTER TWO

2 LITERATURE REVIEW

Every individual has their own set of opinions and emotions, about all matters pertaining to life. These opinions and emotions govern how people communicate with others, their motivations, and their participation in everyday life. There are a number of potential applications which could be developed through the use of sentiment analysis and opinion mining, which would be of benefit to organizations and businesses. These benefits include applications that have the capacity to deduce opinions regarding certain topics, the construction of an automatic recommendation system, the extraction of customer sentiments about a certain product, and many more (Pang and Lee, 2004).

The chapter gives an overview of previous research work in the field of sentiment analysis and aspect-based opinion mining. The rest of the chapter is organized in the following way. Section 2.1 begins with an introduction to the field of opinion mining. Section 2.2 follows with a description of opinion mining and sentiment analysis. Section 2.3 presents a subjectivity analysis, and then in Section 2.4 there is a discussion of sentiment analysis. Section 2.5 offers detailed information about aspect-based opinion mining. Section 2.6 provides a discussion of Arabic opinion mining, and finally, Section 2.7 summarizes the literature review.

2.1 Introduction to Opinion Mining

Recently, web documents have received great focus within a new medium, which can be used to obtain personal opinions and experiences. This brings more attention to technologies for mining and analysing opinions, through web documents
such as forums, weblogs, and customer review websites (Serrano-Guerrero et al., 2015). The growth of research in this area is as a result of the easily-accessible documents across the web. Another catalyst is the maturity of machine-readable techniques. Furthermore, machine learning methods in Natural Language Processing (NLP), as well as Information Retrieval (IR), have seen a considerable increase in the development of practical methods.

Although natural language processing and linguistics have been studied for a long time, few research attempts had been made within the field of mining people's opinions and attitudes, before the year 2002. From that date onwards, this area attracted interest and became active. The main reasons behind this activation are, firstly, that the area of research has a wide range of applications in different domains. Secondly, there are many challenging and unsolved research problems in this field.

In recent years, many researchers have directed their focus to this field of research. They are trying to use technology to extract opinion information, and then to analyse this information automatically. There is a large amount of information shared by users online, ranging from reviews of products and movies, to posts on forums and blogs regarding a variety of topics. Once the data is collected, the challenge comes in analysing and summarising the opinions. Up until now, researchers have proposed several methods for solving this problem. Given that the most recent studies have focused on the sentiment analysis of English language, not much information is available regarding sentiment resources in other languages, specifically Arabic (Montoya et al., 2012).

As mentioned in the previous chapter, research in the recent past has focused on opinion mining and sentiment analysis. The work by Pang and Lee (2008) has provided a very insightful and comprehensive overview of research in the area. A more recent survey has been provided by Liu and Zhang (2012), as well as introductory text regarding this research area. In this chapter, an overview of research studies within opinion mining and sentiment analysis has been provided.

Aspect-based opinion mining is made up of two processes, specifically aspect
extraction and aspect sentiment classification. Aspect extraction techniques are further split into four categories, including i) extraction of frequent nouns and noun phrases, ii) using opinion and target relations, which is challenging to implement with the Arabic dialect because of a lack of reliable NLP, iii) using supervised learning, and iv) using topic models. On the other hand, aspect sentiment classification is further split into two categories, including supervised learning and lexicon-based approaches.

The supervised learning approach is characterized by its dependency on training data, and therefore a model trained using labelled data from a certain domain shows adverse performance when applied to a different domain. Due to the length of documents and the features contained in them, current methods only consider document-level sentiment classification, rather than looking into the specific sentences and clauses contained in the document. For this reason, supervised learning is not easily scalable to large application domains and text. The lexicon-based approach, on the other hand, is easily scalable to a large range of domains, characterised by unsupervised learning. In this approach, sentiment lexicons contain words, phrases and idioms used. This thesis experiments with this approach, through creating a dialect lexicon.

2.1.1 Opinion Mining and Sentiment Analysis

The purpose of opinion mining is to identify, mine or retrieve subjective information from documents containing textual data. Sentiment analysis, on the other hand, aims to identify the text writer’s attitude towards the subject matter, or the sentiment polarity of a given document. This attitude represents his or her evaluation, disposition or judgment about a given topic, event, matter or product. Some problems are encountered in the process of opinion mining and sentiment analysis. Figure 2.1 presents some of the main problems encountered in this field.

The rest of this chapter provides a discussion of problems, and then shifts focus to aspect sentiment orientation detection, which is a sub-problem of aspect-based opinion mining, and is the main problem addressed in this study.
2.1.2 Subjectivity Analysis

Subjectivity analysis (sometimes called subjectivity detection or subjectivity classification) tries to distinguish texts that express factual information (called objective text), from texts that express subjective information (called subjective text). For example, the sentence “Xperia Z is a Sony product” is an objective sentence, while “I like Xperia Z” is a subjective sentence. Subjectivity detection is always used as an initial step in sentiment classification, used to eliminate objective sentences from text. Pang and Lee (2004) have used subjectivity detection to remove objective sentences from review documents and have shown that subjectivity analysis can contribute towards improving sentiment polarity classification, by compressing and cleaning review documents.
A great deal of the existing methods in subjectivity detection are based on supervised learning. For instance, the Naïve Bayes classifier was used by Wiebe et al. (1999) to perform subjectivity classification. In this research work, a set of binary features has been used to represent each text, such as the presence of a pronoun in the sentence, an adjective, a cardinal number, a modal other than ‘will’, and an adverb other than ‘not’. In the same manner, Pang and Lee (2004) built classifiers to be used for the purpose of subjectivity detection in sentences, using semi-automatically constructed data sets as a starting point.

When annotated corpora are not available, a rule-based system is a natural choice for building a subjectivity classifier. One of the most-frequently used rule-based systems in this area is Opinion Finder (Wiebe and Riloff, 2005). This system is capable of checking text through the evaluation of the presence of words from a large lexicon and using this observation to determine the text’s subjectivity. The Opinion Finder uses a high-precision classifier, which classifies sentences as being either subjective or objective, depending on three select heuristics. If a sentence contains two or more subjective terms, the sentence will be classified as subjective. Conversely, if a sentence contains no strong subjective term, but the prior, current and next sentences contain minimum two-week subjective terms, the sentence will then be classified as being a subjective sentence. In the case where neither of these instances is true, then the sentence is classified as being unknown. The classifier makes use of clues obtained from a subjectivity lexicon, as well as rules described above, to classify sentences as being either subjective or objective. A set of extraction patterns is automatically identified from this data. A greater set of subjective and objective sentences is harvested through the use of extraction patterns.

Research work in this area is not restricted to the binary distinctions of subjective and objective content, but also evaluates the clause-level opinion strength (Wilson et al., 2004). It does this by answering questions such as "How angry are you?". Other related works look into the relationship between word sense disambiguation and subjectivity (Wiebe and Mihalko, 2006).
2.1.3 Sentiment Classification

Sentiment classification is defined as the process of establishing a polarity of sentiments (Liu, 2012). For example, in e-commerce websites, customers can write opinions about products. A product review document expresses a customer's opinion about a specific product. For instance, a customer may have a positive opinion about that product, and therefore he or she writes a review that expresses his or her positive opinion in a natural language. The customer's opinion may be negative towards a product, and therefore, he or she may write a negative review. Two examples of positive and negative customer product reviews, as extracted from a Trip Advisor website, have been presented in Figure 2.2. Some additional examples of customer restaurant reviews, which have been used in this thesis, are given in Appendix A. The process of identifying the sentiment polarity of a given reviewed text is referred to as sentiment classification.

Recent research in the natural language processing research community has been centred on sentiment classification, owing to its wide range of applications. These applications include the classification of product reviews posted online and the summary of these opinions (Kang et al., 2012; Ku et al., 2006). This area of research regarding sentiment classification has been studied at different levels of detail. Document and sentence-level sentiment classification will be described in the following sub-sections, while aspect-level sentiment analysis will be introduced in a separate subsection.

2.1.3.1 Document Sentiment Classification

The task of labelling documents as representing either positive or negative opinions, is referred to as sentiment classification (Pang and Lee, 2008). It is assumed that opinionated texts express attitudes regarding a single target, and that the opinions belong to a single opinion holder, which holds true in the case of customer product reviews. Reviews on movies, books or products written by a reviewer expressing their opinion, could be either positive or negative. Document-level classification is used in making this distinction.
There are two classification techniques used in document-level sentiment classification. These include machine learning techniques and lexicon-based techniques, which are shown in Figure 2.3. Machine learning techniques can be further divided into supervised methods and semi-supervised methods. On the other hand, lexicon-based techniques, which are sometimes called unsupervised methods, are categorized as either corpus-based methods, or dictionary-based methods. However, any existing supervised learning method can be used for supervised sentiment classification on the document level. The most commonly used methods in this regard include support vector machine (Pang et al., 2002; Morae’s et al., 2013), Maximum Entropy (Pang et al., 2002), Naïve Bayes (Zhang et al., 2011; Kang et al., 2012), and Artificial Neural Network (Chen et al., 2011; Moraes et al., 2013).

In semi-supervised methods, unlabelled documents can be incorporated into the
process of learning, in order to enhance the performance of the sentiment classification process when there is insufficient labelled corpus for training a sentiment classifier model. Graph-based methods (Goldberg and Zhu, 2006; Ren et al., 2011) and Transductive SVM (Dasgupta and Ng, 2009), are two semi-supervised techniques used for sentiment classification at the document level.

An overview of previous studies on document-level sentiment classification has been provided in Table 2.1, with details of the classification methods, features and data sets used to conduct the research. Selecting and engineering a set of effective features is essential for tasks related to sentiment classification. Various features have been examined by several researchers in different domains of this field of work. Some of the outlined studies have been discussed in further detail through the following subsections.

<table>
<thead>
<tr>
<th>Type</th>
<th>Methods</th>
<th>Features</th>
<th>Dataset</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>SVM, Naïve Bayes</td>
<td>Unigrams, bigrams, trigrams</td>
<td>Restaurant reviews</td>
<td>(Zhang et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>SVM, Naïve Bayes, Maximum Entropy</td>
<td>Unigrams, bigrams, adjective, position of words</td>
<td>Movie reviews</td>
<td>(Pang et al., 2002)</td>
</tr>
<tr>
<td></td>
<td>Naïve Bayes</td>
<td>Unigrams, bigrams</td>
<td>Restaurant reviews</td>
<td>(Kang et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>Neural network</td>
<td>Adjectives, adverbs, nouns</td>
<td>Blog, MP3 reviews, movie reviews</td>
<td>(Chen et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>SVM, neural network</td>
<td>Unigrams</td>
<td>Movie reviews, GPS, book and camera reviews</td>
<td>(Moraes et al., 2013)</td>
</tr>
<tr>
<td></td>
<td>SVM, Naïve Bayes, Maximum entropy</td>
<td>POS-based features, word relations</td>
<td>Movie reviews, book, DVD, electronics and kitchen reviews</td>
<td>(Xia et al., 2011)</td>
</tr>
<tr>
<td>Semi-supervised</td>
<td>TSVM</td>
<td>Unigrams</td>
<td>Movie reviews, book, DVD, electronics and kitchen reviews</td>
<td>(Dasgupta and Ng, 2009)</td>
</tr>
<tr>
<td></td>
<td>Graph-based</td>
<td>Adjectives, adverbs, nouns</td>
<td>Book, hotel, notebook reviews</td>
<td>(Ren et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>Graph-based</td>
<td>Unigrams</td>
<td>Movie reviews</td>
<td>(Goldberg and Zhu, 2006)</td>
</tr>
<tr>
<td>Lexicon-based</td>
<td>Corpus-based</td>
<td>Adjectives and adverbs</td>
<td>Automobile, bank, movie, travel reviews</td>
<td>(Turney, 2002)</td>
</tr>
<tr>
<td></td>
<td>Corpus-based</td>
<td>Adjectives and adverbs</td>
<td>Movie reviews</td>
<td>(Harb et al., 2008)</td>
</tr>
<tr>
<td></td>
<td>Dictionary-based</td>
<td>Adjectives, nouns, verbs adverbs, intensifier, negation</td>
<td>Movie reviews, camera reviews and opinions</td>
<td>(Taboada et al., 2011)</td>
</tr>
</tbody>
</table>
2.1.3.1.1 Machine Learning Techniques

Machine learning techniques always rely on large volumes of labelled data to be used for the purposes of training a model to be used as a sentiment classifier, and then employing the trained model to predict the polarities of sentiments in unlabelled documents. These methods can be divided into two separate groups, namely supervised methods and semi-supervised methods. These two groups will be discussed in the following subsections.

- **Supervised Methods**

  Sentiment classification can be expressed as a supervised learning problem comprised of two classes, namely positive and negative. In order to train the test data, product reviews are the most-commonly used data.

  There are various supervised learning techniques, such as naïve Bayes and Support Vector Machines (SVM), any of which are applicable for sentiment classification. However, SVM has slightly better performance when compared to Naïve Bayes, and even with the rest of the classifiers. According to Cui et al. (2006) the SVM has a superior performance, because it is a discriminative classifier which makes it well suited for the classification of mixed sentiments. In addition, SVM requires the input of a large data set, so as to be able to generate a high-quality classifier. In the event that only a small amount of training data is available, then the Naïve Bayes classifier is more appropriate.

  Sentiment classification requires the selection of a suitable set of features. Table 2.2 shows some of the most widely-used features.

  Pang et al. (2002) is one of the pioneers in the supervised methods of sentiment classification. In their work, they used three machine learning methods for the classification of sentiments in movie reviews, using the standard bag-of-words (BOW) framework. Several features have been tested in order to arrive at an optimal feature set, which is made up of unigrams, bigrams, adjectives and word positions. In order to cut down the number of features, only unigrams and bigrams with an appearance frequency
of at least four and seven times, respectively, have been used. The results obtained from the study have revealed that performance was better when unigrams were used in conjunction with the SVM classifier. The performance was better while using feature presence to determine the weight of terms, as opposed to using feature frequency. In another study, Xia et al. (2011) used word syntactic relations together with the POS-based feature set, employing three different classification techniques within an ensemble framework, in order to perform sentiment classification.

Table 2.2: The Most Widely-Used Features in Sentiment Classification

<table>
<thead>
<tr>
<th>Feature</th>
<th>Meaning of the feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terms and their frequency</td>
<td>Refers to individual words, along with their frequency of appearance in the sentiments. These features are effective for sentiment classification, and have been widely implemented (Pang et al., 2002; Xia et al., 2011).</td>
</tr>
<tr>
<td>Parts-of-Speech information</td>
<td>POS information, such as adjectives, are an important indicator of a document’s sentiment (Pang et al., 2002).</td>
</tr>
<tr>
<td>Opinion words</td>
<td>Opinion words, sentiments or phrases are used to express sentiments which may be positive or negative. Words such as ‘good’, ‘amazing’, and ‘brilliant’ express positive emotions, while words such as ‘bad’, ‘slow’, and ‘poor’ express negative emotions. Opinion words can adjectives, adverbs, nouns or verbs (Hung and Lin, 2013; Montejo-Ráez et al., 2014).</td>
</tr>
<tr>
<td>Negations</td>
<td>Words of negation can transform a sentiment’s polarity. For instance, if you consider the statement “I don’t like this mobile”, the sentence has negative orientation even though there is the positive opinion word ‘like’ (Na et al., 2005).</td>
</tr>
<tr>
<td>Syntactic dependency</td>
<td>Refers to features that are word dependent and are generated from a dependency tree or parsing (Matsumoto et al., 2005; Zheng-Jun et al., 2014).</td>
</tr>
</tbody>
</table>

Whitelaw et al. (2005) combined the BOW framework with a shallow parsing technique aimed at finding opinion phrases, classified by the orientation and taxonomy of attitude types from the appraisal theory as presented by Martin and White (2005), all of which were specified using a hand-constructed attitude lexicon. Text classification was conducted using a support vector machine, with the vector being word frequency (for BOW), and the percentage of the classified appraisal group over the attitude taxonomy.
Ye et al. (2009) introduced the application of classification techniques in destination reviews, using three supervised learning algorithms including SVM, Naïve Bayes, and the character-based N-grams model, for supervised learning. Additionally, information gain (IG) was used for feature selection. In this study, word frequency was used for document representation, as opposed to the use of word presence. Overall, the study showed that SVM has better performance than the other two.

Recently, Artificial Neural Networks (ANN) has also been employed as a sentiment classifier by some researchers, in regards to document-level sentiment classification. Moraes et al. (2013) provided an empirical comparison of ANN and SVM, within document sentiment classification. The work argued that ANN could perform better than SVM, in the case of unbalanced data. In another study, Chen et al. (2011) proposed an approach based on neural networks, which were used in combination with the semantic orientation index in order to determine the sentiment polarities of bloggers. Although ANN showed a good performance in comparison with SVM, ANN contained more sensitivity to noise features than SVM, especially when the
discrepancy in data was increased.

In almost all research, SVM showed better performance in comparison to other methods, particularly with a large labelled training set. The greatest challenge experienced in supervised learning classification methods, is the sensitivity of the outcome to the quality (bias), and the quantity of training data. Another challenge is sentiment classification at the sub-document level, due to the unavailability of sufficient information.

- **Semi-supervised Methods**

  Semi-supervised learning methods are trained using a combination of labelled and unlabelled data. This differs from the supervised learning methods, which are trained using labelled data. The use of semi-supervised learning (SSL) in a sentiment classification is relatively new, brought on by the unavailability of labelled data in real world scenarios. The idea in SSL is that even though unlabelled data may not contain information about its classification, being either positive or negative, it may hold information on the joint distribution of classification features. Therefore, in the case where there is insufficient labelled data on the target domain for use in supervised learning, unlabelled data can be used with SSL to provide improved performance.

  Previous work by Ren et al. (2011) investigated the effectiveness of the label propagation technique, a special kind of graph-based semi-supervised learning approach, for sentiment classification in resource-scarce languages. They found that label propagation outperforms SVM and transductive SVM in document-level sentiment classification, for small-labelled training data. Goldberg and Zhu (2006) introduced graph-based semi-supervised learning technique for document sentiment classification. A graph was created on both labelled and unlabelled documents based on documents similarity, and then technique-solved an optimization problem, used to find a smooth rating function over the whole graph. They also showed that the semi-supervised learning approach outperforms supervised methods when the labelled training set is not large enough.
Transductive SVM (TSVM) is another semi-supervised learning technique, which was employed by Dasgupta and Ng (2009) for the sentiment classification of movie and product reviews. Research first identified unambiguous review documents using semi-supervised spectral clustering, and then employed active learning to select and manually label some small number of relevant documents, to create a small training set. Ultimately, they used TSVM to classify review documents, by incorporating unlabelled documents along with a small training set.

2.1.3.1.2 Lexicon-based Techniques

A great deal of researchers have used words and phrases for lexicon-based classification, due to the straight-forward nature of classification. The lexicon-based classification involves the calculation of the semantic orientation of words or phrases, followed by the establishment of the sentiment orientation of the entire document based on the average semantic orientation of the words and phrases contained in it (Turney, 2002; Harb et al., 2008). The establishment of the semantic orientation of words and phrases in a document can be done using either corpus-based methods, or dictionary-based methods.

A. Corpus-based Methods

The first approach is the corpus-based method, which relies on patterns that co-occur with a seed set of opinion words, to identify other opinion words and their sentiment orientations in a large corpus. In an earlier research work, Turney (2002) proposed a simple lexicon-based approach for the classification of reviews into two categories, being either ‘recommended’ or ‘not recommended’. The polarity of words was determined by computing the words' Point-wise Mutual Information (PMI) for their co-occurrence with a positive seed word (‘excellent’), and a negative seed word (‘poor’). PMI was estimated by issuing some queries to a search engine, to calculate the number of hits when querying a phrase with a positive seed word and a negative seed word. This value was labelled as the Semantic Orientation (SO). In this method, a review is scanned in order to find phrases that match certain speech patterns, such as adjectives and
adverbs. Afterwards, the semantic orientation of the phrases is added up, in order to determine the review’s polarity. The results of the study showed 74% accuracy in the classification of a corpus of product reviews.

Harb et al. (2008) carried out blog classification in his study by using 2 sets of seed words with positive and negative semantic orientations to start with, as is the case in Turney (2002). The Google search engine was then used to generate associations, which helped find more words that would provide insight into the document’s sentiment classification. Afterwards, the documents were classified on the basis of comparing the number of positive and negative adjectives. The drawback of this method, is its heavy reliance on labelled seed words, such as ‘excellent’ and ‘poor’. This creates a domain-oriented sentiment lexicon, in which is not possible to find the mutual relationship between words and documents.

B. Dictionary-based Methods

The second approach uses a sentiment dictionary word, such as SentiWordNet (Baccianella et al., 2010), instead of searching large corpora to determine the sentiment orientation of words or phrases contained in a document.

Taboada et al. (2011) proposed a dictionary-based sentiment classifier, made up of a dictionary of positive and negative-polarized words. A semantic orientation calculator (SO-CAL) was built on the basis of this dictionary, by combining intensifiers and negation words. Upon testing this approach, it was found to be 59.6% to 76.4% accurate on the 1,900 documents of the dataset which contained movie reviews.

Machine learning and lexicon-based techniques have been extensively used in document-level sentiment classification. Each of these two techniques has its own strengths and weaknesses. Table 2.3 shows some of these strengths and weaknesses, in regards to document-level sentiment classification.

The greatest strength of the sentiment classification of documents, is that it is capable of narrowing down the predominant opinion expressed. Its main weakness is that it lacks details regarding people’s likes or dislikes, and cannot be extended to non-reviews, such as forums or blogs which contain multiple entities which require
comparisons to be made.

Table 2.3: Machine Learning Techniques vs. Lexicon-based Techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>Strength</th>
<th>Weakness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexicon-based techniques</td>
<td>Conceptually intuitive. Easy to implement.</td>
<td>Heavy dependence on the quality of opinion words. Needs sophisticated method to identify and extract opinion words. Needs an ability to capture all exceptions. Absence of explicit opinion words in the expression of some opinions.</td>
</tr>
</tbody>
</table>

2.1.3.2 Sentence Sentiment Classification

The classification of sentences, as being either subjective or objective is referred to as subjectivity classification, as discussed in Section 2.3. Once a sentence is classified as being subjective, it is further evaluated and then classified as having either a positive or negative orientation. This is referred to as sentence-level sentiment classification. The key aspect in this classification is the identification of the sentence’s target, without which the detected polarity is not useful.

A similar lexicon-based approach to the one used in Turney (2002), was used by Yu and Hatzivassiloglou (2003) for sentence sentiment classification. In this study, the authors utilized a large set of seed words, instead of using only one positive and one negative seed word. Moreover, the authors have employed a modified version of the log-likelihood ratio, as a means of calculating each word’s sentiment polarity. The average log-likelihood scores of the words contained in the sentences were used by authors as a means of assigning a sentiment label to each sentence.

McDonald et al. (2007) came up with a sentiment analysis model which is applicable to the evaluation of sentiments at various levels of concurrent granularity. The authors used graphical models, wherein the sentiments were sequentially linked.
Here a document-level sentiment is linked to several paragraph-level sentiments, while each paragraph level sentiment is linked to several sentence-level sentiments. The Viterbi algorithm is used to infer the sentiment of each text unit, ensuring that the labels in a particular paragraph or document, are the same for similar paragraphs or documents.

Neviarouskaya et al. (2010) developed a sentiment computation system, which worked by evaluating words in a sentence. The system’s operation was based on the appraisal theory of Martin and White (2005), and the categorization method of Izard (1971). The system utilized a rather complicated rule set, as a means of determining the attitude conveyed in sections of the sentence, and eventually formulating an inclusive label.

Gamon et al. (2005) made use of a semi-supervised algorithm for sentence sentiment classification. The work used a large set of unlabelled sentences, together with a small set of labelled sentences, as a means of training a sentiment classifier. EM-Naïve Bayes was used as the base classifier within their learning algorithm.

### 2.1.4 Sentiment Lexicon Construction

Sentiment words, which are useful for sentiment classification, may come in the form of ‘polar words’, ‘opinion words’, or ‘opinion-bearing words’. Sentiment words may either be categorized as being either positive or negative. Positive sentiments express some desirable states or qualities, while negative sentiments express some undesirable states or qualities. For example, ‘excellent’ is a positive sentiment, while ‘poor’ is a negative sentiment. Examples of positive sentiments in the Arabic language include ضعيف،سيء،ممتاز،مدهش،يحب،امتنان، and examples of negative sentiments include جميل،ممتاز،مدهش،يحب،امتنان. Sentiments can be found in a sentence as an adjective, verb, as a phrase, or as an idiom. Phrases can also be used alongside individual words, for the purposes of sentiment classification. A sentiment lexicon or opinion lexicon contains sentiment words, as well as sentiment phrases.

Sentiment words are further divisible into base types and comparative types. The
examples provided in the previous paragraph belong to the base type category. Comparative and superlative sentiment words express comparative and superlative opinions, respectively. For instance, أَرْخَص, أَفْضَل, أَحْسَن, are comparative and superlative forms of base adjectives or adverbs, such as جَيِّد. Base type sentiment words express a regular opinion, while the comparative form refers to an opinion that relates to more than a single entity - for instance، "وجبة ماكدونالدز أَرْخَص من وجبة هريفي، (أ)" The previous sentence provides a comparison of meal prices in two restaurants.

The identification of opinion words is challenging, due to the domain and context dependency of words. A great deal of researchers have looked into the problem of finding opinion words and have proposed several approaches to tackling the problem. A sentiment lexicon is constructed using any of three methods, these being manual construction, large corpus-based construction, and dictionary construction. Manual construction is cumbersome and time-consuming. It is used in conjunction with other methods for improving performance accuracy, because it is not applicable on its own. The following sub-section provides a discussion of the other two methods.

2.1.4.1 Construction of Sentiment Lexicon Based on Large Corpus

The large corpus-based construction is applicable in two instances. The first is in the presence of a seed list of general purpose sentiment words, into which more sentiment words are discovered, and their orientation is identified with the help of a domain corpus, and added to the list. The second is in the construction of a domain-specific sentiment lexicon, adapted from a general purpose sentiment lexicon, through the use of a domain corpus. The adaptation factors in the context-dependent meanings of words, such as "سعر الوجبة" which is negative in the "سعر الوجبة مرتقب" context, but "مرتقب" is positive in a different context.

Both of these methods use sentiment words, whose polarity is known, while making use of syntactic patterns which are useful for the identification of sentiment words, along with their polarity (Huang et al., 2014).
Hatzivassiloglou and McKeown (1997) were pioneers in the domain of word orientation determination. In their work, they looked at adjectives and phrases containing adjectives connected by conjunctions, like ‘and’ or ‘but’. They went on to construct a technique functioning on the basis of graphs, which facilitated the learning of lexicons by reading from a corpus, whereby adjective pairs are joined using conjunctions. Examples of these may be "simple and well-received" or "fair but brutal". These conjunctions also join morphologically related-adjecitves as such as ‘thoughtful’ and ‘thoughtless’. Finally, a graph was created in which the vertices represented words, while the edges represented pairs which have the same-polarity, or are identified as being opposite-polarity links. From their study, Hatzivassiloglou and McKeown developed a graph-clustering algorithm with 82% accuracy, which served the purpose of creating two clusters of found adjectives, specifically positive and negative word clusters.

Wu and Wen (2010) suggested the use of a method based on syntactic patterns. They also used web search hit counts to arrive at a solution to the same problem but faced in the Chinese language. The authors also looked into word pairs containing quantifiers, like ‘big’, ‘small’, ‘low’ and ‘high’.

Lu et al., (2011) used a similar context definition. Ding et al., (2008) made the assumption that the sets of aspects were provided. The tracking of word pairs is considered to be an optimization problem, which contains some constraints. The objective function and constraints in the optimization problem have been designed, based on clues contained in a general-purpose sentiment lexicon. The information from the lexicon was used to rate sentiments posted in the reviews, with the incorporation of synonyms, antonyms, conjunctions and negations.

Du et al., (2010) proposed an algorithm to adapt the sentiment lexicon, which was not a general-purpose lexicon, from one domain to another domain. The proposed algorithm took in multiple inputs, as a set of labelled documents containing sentiments from the original domain, a set of sentiment words from those documents, and a set of documents from the new domain. The technique used two approaches. Firstly, a document was labelled as being positive or negative, based on the presence of positive
or negative words. A word was labelled as being positive or negative, based on its frequency of appearance in positive or negative documents. This is otherwise referred to as mutual reinforcement relationships. In the second approach, the two domains were taken under different distributions, with any similarities being taken as having the same orientation. The sentiment lexicon adoption was undertaken through the application of the information bottleneck framework, with the solution to this problem being proposed by Du and Tan (2009).

Feng et al. (2011) proposed a graph-based method, which worked on the basis of mutual reinforcement, to serve as a solution to the problem of constructing a connotation lexicon. A sentiment lexicon handles words which explicitly or implicitly convey sentiments, while a connotation lexicon handles words associated with a specific polarity. For instance, the word ‘cancer’ has a negative connotation, while the word ‘award’ has a positive connotation. In terms of building a general-purpose sentiment lexicon, the dictionary-based approach is far more effective as it contains all words. However, the corpus-based approach is applicable to very large and diverse corpuses. Dictionary-based approaches are less suited for domain-specific opinion words. Up until the present day, domain and context-dependent sentiments remain challenging. There are two key difficulties in lexicon construction. The first is handling context or domain-dependent opinion words, without the user’s prior knowledge. The second is handling multiple language constructs, which are capable of changing the semantic orientation of opinion words, such as a negation word.

Corpus-based methods have the capability of producing accurate lists of positive and negative words. However, a great deal of these methods need to use large annotated training datasets, in order to achieve their full potential. The shortcomings of the corpus-based approaches can be eliminated by using dictionary-based methods, whose operation is dependent on lexicographical resources such as WordNet.

### 2.1.4.2 Construction of Sentiment Lexicon Using Dictionary

Sentiment lexicon construction, using dictionary-based methods, alleviates the
need to use large corpora or search engines. Instead, they make use of the existing lexical resources, such as WordNet, which have the ability to produce accurate, comprehensive, and domain-independent lists of words and senses, annotated for sentiment and subjectivity. These methods involve the manual collection of initial seed, through a set of sentimental words and their orientation. This is followed by a dictionary search for synonyms and antonyms, which can be used to expand the set. These are then applied iteratively, so as to generate new sentiment words.

Hassan (2010) used WordNet synonyms and hyponyms to present a Markov random walk model for building a word-relatedness graph for sentiment estimation. The model defined the mean hitting time as ‘h(i| S)’, the node ‘i’ and the set of node words ‘S’, which refers to the average number of steps that a random walker from state ‘i S’ would take in order to enter state ‘k , S’ for the first time. The sentiment orientation of a word ‘w’ was estimated, using the set of positive seed words ‘S’, and the set of negative seed words ‘S-’. The model was also used to compute the hitting times ‘h (w |S +)’, and ‘h (w| S -)’. The rule of thumb used was that if ‘h (w |S +)’ was greater than ‘h (w|S −)’, otherwise the word was classified as negative. Otherwise, the word is classified as positive.

Hassan et al., (2011) defined a multilingual method of determining the sentiment orientations of words in a foreign language. The authors built a word graph for English words, as well as for words in a foreign language. A connection was made through the use of definitions derived from foreign language dictionaries.

Velikovich et al., (2010) presented a sentiment lexicon construction method using a graph propagation algorithm, run on a phrase similarity graph, with the use of data from four billion web pages. In this method, a set of positive and negative seed phrases were passed as input. The candidate phrases obtained from the web pages served as the nodes in the phase graph. Out of these, only 20 million candidate phrases were selected based on heuristics, frequency and word boundaries. For each phrase, a context vector was created with a size-six window, summed up over the entire four billion phrases. The set of edges was built through the use of the cosine similarity of the context vectors of the candidate phrases, which were selected only if they had 25 or more
adjacent edges, with the highest cumulative weight. The proposed graph propagation method was then used to determine a phrase’s sentiment, by summing up the best paths that lead to the seed words.

A different bootstrapping method was proposed by Dragut et al., (2010), using WordNet. The authors used inference rules to determine sentiment orientations, using seed words whose orientation was already known, using them to construct synsets made up of synonyms whose orientation and polarity were deduced from the available data set.

In summary, the dictionary-based approach is advantageous, because it allows for greater ease in finding a great volume of sentiment words, along with their corresponding orientations. Despite the likelihood of errors in classification, a one-time but albeit time-consuming manual check can be done undertaken, in order to alleviate errors. The drawback of dictionary-based approaches is the general nature of sentiment orientations contained in it, thereby making it laborious to apply in specific context-dependent domains. This is because a single sentiment word may portray differing emotions in different domains. The corpus-based approach overcomes this problem.

Table 2.4 provides a summary of the advantages and disadvantages of the different approaches used in creating lexicon, previously discussed in this subsection.

2.2 Aspect-based Opinion Mining

With the steady growth of the number of e-commerce transactions, there is an ever-increasing volume of data on products, and their related online reviews. A great deal of customers are of the opinion that they can make more informed decisions with the help of information acquired from other customers’ user experiences posted online (Yang et al., 2010). However, as the volume of reviews grows on a daily basis, it is more difficult for users to read through all of them. To resolve this problem, several alternatives have been proposed for summarizing, and presenting these evaluations for
the entire range of product features to users.

Table 2.4: Summary of Lexicon Creating Methods Discussed in the Literature

<table>
<thead>
<tr>
<th>Approach</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus-based approach</td>
<td>The possibility of identifying multi-word, opinion-bearing expressions, in finding domain-dependent orientation lexicon.</td>
<td>Requires the processing of large volumes of data.</td>
</tr>
<tr>
<td>Dictionary-based approach</td>
<td>Ease of identifying large volumes of sentiment words, along with their corresponding orientation. Exploration of well-defined, coded and validated semantic relations between words, and a large lexical database.</td>
<td>Requires the time-consuming task of cleaning errors in the resultant list. The general nature of sentiment orientations is of a general nature, thereby making it laborious to apply in specific, context-dependent domains.</td>
</tr>
<tr>
<td>Translation-based approach</td>
<td>Unavailability of linguistic resources, in languages besides English.</td>
<td>Difficulty in translating words between different languages, while retaining the intended meaning.</td>
</tr>
<tr>
<td>Manual-based approach</td>
<td>Provides the ability to find words which bear different opinions in different domains.</td>
<td>It is labor intensive and time-consuming.</td>
</tr>
</tbody>
</table>

Some of the work done on product reviews falls within the category of coarse-grained sentiment analysis, and tries to answer the question of whether the general review of the product is positive or negative (Pang et al., 2002; Blitzer et al., 2007). However, there is a large body of research being carried out on product reviews, with the intention of addressing more fine-grained questions. Examples of these questions can be “what are the most popular product features among customers?” or “what are the most disliked product features?” (Hu and Liu, 2004; Titov and McDonald, 2008). This interest in the fine-grained sentiment analysis of product reviews has led to the inception of the sub-area of sentiment analysis, referred to as aspect-based opinion mining, and also sometimes as feature-based opinion mining (Liu and Zhang, 2012).

Feature-level (aspect level) opinion mining performs a fine-grained analysis, which aims to derive target or feature entities from sentences, and to identify opinion words and expressions. This method looks directly into the opinion itself, rather than looking into language constructs, such as documents, paragraphs, sentences, clauses or phrases.
In many applications, the opinion targets are described by entities or their aspects. Therefore, the goal of this level of analysis is to determine the sentiments of entities and their aspects. For example, the sentence “فمعلومة جميلة لكن سعره غالي” evaluates two aspects of restaurant service, specifically "price" and "views". The sentiment in "فمعلومة جميلة" has a positive expression, but the sentiment in "سعره غالي" has a negative expression. "سعره غالي" and "فمعلومة جميلة" are the opinion targets.

Therefore, unstructured data is transformed into structured data, in the form of opinions regarding entities and aspects, which can then be applied to various qualitative and quantitative analysis facilities. Classification at the document, sentence and feature levels is quite a challenge, with feature-level classification posing the greatest challenge, thereby spurring development of various methods of feature-based opinion mining.

Aspect-based opinion mining consists of two sub-tasks, specifically aspect extraction and aspect sentiment-orientation detection, both of which are described in the next section. Figure 2.4 illustrate the two sub-tasks and the main approaches of each.

### 2.2.1 Techniques Used to Identify Aspects

Aspect extraction is considered to be one of the most complicated elements involved in aspect-based sentiment analysis, which is also referred to as topic, feature, or target extraction. The task involves applying natural language processing techniques, in order to automatically-extract features or aspects from opinion documents. There are four main approaches to extracting aspects, which include:

1. Supervised learning.
2. Topic modelling.
3. Frequent nouns and noun phrases.
4. Exploiting opinion and target relations.

The proposed work in this study focuses on two particular approaches, specifically extraction based on frequent nouns and noun phrases, and extraction by exploiting opinion and target relations. Both of these techniques are more compatible
with the Arabic corpus review, given that they can handle the unavailability of a labelled data set. Being supervised techniques, they require the use of manually-labelled data for training purposes. Training a data set involves the process of manually annotating aspects and non-aspects in a corpus. Topic modelling, on the other hand, is an unsupervised learning method which works on the assumption that each document contains a variety of topics, which have a probability distribution over words. Therefore, a topic model lays out a probabilistic procedure for document generation, with the output being a word cluster set. Each cluster constitutes a topic, with a probability distribution over words in the document collection and is dependent on the size of the corpus.

Several aspect extraction techniques are listed in Table 2.5, along with the authors who published the technique. Kobayashi et al. (2004) applied the unsupervised approach for target feature and opinion extraction, and proposed a semi-automatic extraction process which is applicable in evaluating expressions, concerning target features and objects. The proposed method applied text mining techniques to extracting candidate evaluative expressions, thereby speeding up the manual annotation process used to compose lists of evaluative pairs, which are then applied in the ML training data sets.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Association Mining and Web PMI</td>
<td>(Hu and Liu, 2004)</td>
</tr>
<tr>
<td></td>
<td>(Popescu and Etzioni, 2005)</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>(Yi et al., 2003)</td>
</tr>
<tr>
<td>CRF approach</td>
<td>(Jakob and Gurevych, 2010)</td>
</tr>
<tr>
<td>SVM</td>
<td>(Kessler and Nicolov, 2009)</td>
</tr>
<tr>
<td>HMMs</td>
<td>(Jin et al., 2009)</td>
</tr>
<tr>
<td>Averaged perception</td>
<td>(Stoyanov and Cardie, 2008)</td>
</tr>
<tr>
<td>Double propagation–syntactic relation</td>
<td>(Qiu et al., 2011)</td>
</tr>
</tbody>
</table>
Popescu and Etzioni (2005) proposed the OPINE system, which is an unsupervised learning approach for feature and opinion extraction. The system made use of syntactic patterns for semantic orientation, in order to identify opinion phrases, along with their polarity.

Carenini, Ng et al. (2005) applied user knowledge to come up with a model for building taxonomies of product features. The authors used an unsupervised approach of feature extraction, using the built taxonomy. The combination of the two approaches showed better results than the use of the existing unsupervised approach. The drawback of the combined approach is the domain dependence, which arises from the application of a pre-knowledge base.

Holzinger et al. (2006) made use of the domain ontologies built using tabular web content data, to develop a feature extraction process. The process involved the creation of a wrapper which uses logical rules to extract data from tables, and applies data integration for pinpointing product features.
Research by Zhuang, Jing, and Zhu (2006) centred on opinion mining within the field of movie reviews. In their work, they developed a multiknowledge-based approach, which was combined with WordNet4, a statistical analysis, and a movie knowledge base. The approach used a combination of grammatical rules and keyword lists, in order to identify product features, and the opinions expressed which pertain to these features.

Bloom et al. (2007) proposed an unsupervised feature and appraisal extraction technique, based on the premise that the expression of a product’s assessment is crucial in sentiment analysis. The assessment is presented as text which describes the outcome of a target’s evaluation. This study proposed expressions that are indicative of an evaluation focused on adjectives, and used them to pinpoint opinion targets.

Ben-David et al. (2007) developed a domain classification method using a structural correspondence learning (SCL) algorithm, which applied perception in predicting new domain features, using training data. Training is done using the source domain and is then implemented on the target domain. Later research by Blitzer et al. (2007) extended the SCL algorithm for its implementation in opinion target identification.

Lu and Zhai (2008) developed a semi-supervised topic model, used to solve the problem of integrating opinions contained in reviews scattered across different blogs and forums. The study explored the integration of opinions regarding product and political reviews. Their study focused on coming up with a model of opinion extraction, generalized across different domains.

Ferreira and Jakob et al. (2008) developed an extended pattern-based feature extraction approach, applying a modified Log-Likelihood Ratio Test (LRT) as previously used by Yi et al. (2003), for purposes of target identification. The study also proposed an extension of the annotated scheme, previously proposed by Hu and Liu (2004), for application in product features.

Kessler, Eckert et al. (2010) proposed an annotated corpus, made up of mentions, co-reference, metonymy, sentiment expressions, and modifiers of sentiment expressions,
such as neutralizers, negators and intensifiers. The corpus was designed for the automotive industry, with the intention of quantifying sentiments and target features.

Chao (2010) proposed a feature-based opinion mining model for reviews in the hotel industry. The model utilized a manually-annotated corpus, constructed from opinions in tourism blogs. The model employed a supervised machine-learning approach for training purposes.

Recent studies in feature-based opinion analysis included that of Zhai et al. (2011) in which a semi-supervised feature grouping technique was proposed for application in developing the summary of a collection of opinions. The technique took into account the use of different words (synonyms), or phrases, which refer to the same feature. Consequently, when coming up with an effective summary, it is important to group together these similar phrases and words. An initial list was developed to bootstrap the process, using lexical characteristics of terms, and was later applied in grouping features. An examination of the model showed satisfactory empirical results.

Goujon (2011) developed a linguistic knowledge-based text mining approach for establishing features or targets of expressed opinions, in relation to a particular topic. The study made use of linguistic patterns in pinpointing the subject or target of the expressed opinions.

Khairullah Khan et al. (2014) developed a pattern-based features extraction approach, using a hybrid pattern combination of the syntactic sequence, as well as semantic relation. The semantic relation used adjectives with polarity, while syntactic patterns used two existing patterns and one new pattern, as based on a linking verb. A comparison of the developed approach with existing ones, showed that the new one had a superior performance. The model proposed in this study is similar to the approach of Khairullah Khan et al. (2014) but it is tailored for application in extracting opinions from reviews written in Arabic. Table2.6 summarize different patterns and approach used for aspect extraction.
2.2.1.1 Finding Frequent Nouns and Noun Phrases

This method involves identifying explicit aspect expressions in the form of nouns and noun phrases, from collection reviews belonging to a particular domain. The main advantage of the method is its simplicity, and it gives good results by being empirical, particularly in regards to product reviews. Its main disadvantage is the absence of the normalization of features, and that different heuristics may be required for different domains.

Hu and Liu (2004a) have carried out some initial studies in regards to aspect extraction from customer reviews, in which association rule mining has been used in conjunction with pruning strategies, in order to identify candidate features. The work was based on the premise that product features appear in the form of nouns or noun phrases. The study began with the parsing and tagging of Parts-of-Speech (POS), after which a transaction was created for each noun word in a sentence, and the resulting transactions served as the input for the rule mining algorithm used to identify the frequent sets of items. An itemset is defined as a set of words or phrases which appear together in sentences. The frequent item sets, based on their frequency of occurrence, were identified as product features. The product features were then split into two groups, specifically frequent features and infrequent features. Infrequent feature words were extracted, using the adjacent noun phrases of known opinion words. A priori algorithm was then used to find frequent words, without considering the position of the words in a sentence.

Popescu and Etzioni (2007) removed noun phrases not associated with any features, which could then be used in determining the polarity of a sentence or review. This method entailed the computation of a Pointwise Mutual Information (PMI) score, between the phrase and some metonymy discriminators associated with the entity class.

Blair-Goldensohn et al. (2008) conducted a study of noun phrases found in sentiment-bearing sentences, or in ones bearing syntactic patterns which serve as an indicator of underlying sentiments. Ku et al. (2006) utilized the TF-IDF scheme, which
takes into account terms at both the document and paragraph levels. Moghaddam and Ester (2010), combined the frequency-based approach with a pattern-based filter used to eliminate non-aspect terms. Their approach was applied in predicting aspect ratings. Scaffidi et al. (2007) proposed a method for identifying true aspects, by carrying out a comparison of the frequency of the extracted frequent nouns and noun phrases in a review corpus, considering their occurrence rates in a generic English corpus. Long et al. (2010) proposed a feature extraction method, based on frequency and distance information.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Pattern</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsupervised</td>
<td>evaluating expressions</td>
<td>Kobayashi et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>syntactic patterns</td>
<td>Popescu and Etzioni (2005)</td>
</tr>
<tr>
<td></td>
<td>user knowledge</td>
<td>Carenini, Ng et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>domain ontologies</td>
<td>Holzinger et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>indicative expressions</td>
<td>Bloom et al. (2007)</td>
</tr>
<tr>
<td></td>
<td>structural correspondence learning (SCL)</td>
<td>Ben-David et al. (2007), Blitzer et al. (2007)</td>
</tr>
<tr>
<td>Semi-Supervised</td>
<td>Topic model</td>
<td>Lu and Zhai (2008)</td>
</tr>
<tr>
<td></td>
<td>modified Log-Likelihood Ratio Test (LRT)</td>
<td>Ferreira, Jakob et al. (2008), Yi et al. 2003</td>
</tr>
<tr>
<td></td>
<td>co-reference, modifiers and metonymy, sentiment expressions</td>
<td>Kessler, Eckert et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>feature grouping technique</td>
<td>Zhai et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>linguistic knowledge-based</td>
<td>Goujon (2011)</td>
</tr>
<tr>
<td>Supervised</td>
<td>hybrid pattern combination of the syntactic and semantic relation</td>
<td>Khairullah Khan et al. (2014)</td>
</tr>
</tbody>
</table>

### 2.2.1.2 Using Opinion and Target Relations

An opinion unit is made up of three interrelated items, specifically a product feature or target, an expression of opinion, and a positive or negative emotional attitude. The relationship between the items is beneficial for extracting opinion targets, given that sentiment words are common knowledge. Hu and Liu (2004a) applied this method to extract infrequent features. They used the same sentiment word to describe and modify different features. In the event that a sentence contains no frequent aspect, but has some sentiment word, then the noun or noun phrase closest to the sentiments is extracted.
Since no parser was used by Hu and Liu, the ‘nearest’ function gives an approximation of the dependency between the sentiment word and the noun or noun phrase it modified. This technique showed satisfactory performance. For instance, take the sentence “The software is amazing”. Here, if ‘amazing’ is a sentiment word, then ‘software’ is extracted as an aspect. This concept is useful for finding all the features of a product mentioned in reviews.

Blair-Goldensohn et al. (2008) proposed the sentiment patterns method, which applied a similar idea. Furthermore, they used the method to discover important key aspects or topics within opinion documents. The method was found to be useful, in that an aspect was considered to be important if it had no associated opinion or sentiment.

Zhuang et al. (2006) and Somasundaran et al. (2009) used a dependency parser to identify dependency relations for aspect extraction, whereas Wu et al. (2009) identified candidate aspects as being either noun or verb phrases. Rather than using a normal dependency parses, a phrase dependency parser was used to extract noun and verb phrases. This was followed by the filtering out of unlikely aspects using a language model.

All previous work conducted on a normal dependency parser was concerned with the identified dependency of individual words. A phrase dependency parser, on the other hand, determined the phrase dependency and thereby made it suitable for aspect extraction. The concept of dependency was generalized into a double-propagation method, retrieving sentiment words and aspects as presented by Qiu et al. (2009) and Qiu et al. (2011).

Kessler and Nicolov (2009) focused exclusively on the identification of opinionated expressions, especially their association with each aspect of a product review. Data retrieved from car and camera review sites served as a dataset for annotated opinion expressions and target aspects. This was followed by the training of a machine learning classifier (SVM) to pinpoint related opinion expressions and target aspects. Feature vectors were then constructed on the basis of the syntactic and semantic relationship between opinion expressions and candidate aspects.
The study by Jakob and Gurevych, (2010) used a Conditional Random Fields (CRF) based approach for opinion extraction, using input features like POS tags, short dependency paths, word distance, and opinion sentences. The study utilized data sets from three sources, as a means of proving the effectiveness of the approach in both single and cross-domain settings.

Stoyanov and Cardie (2008) considered aspect extraction to be a topic coreference resolution problem. They proposed a method applicable to cluster opinions, that is, opinions that share a target. They trained a classifier to determine whether two separate opinions share a target, and the result indicated that their approach was supervised. Table 2.7 summarize some aspect extraction method strength and limitations.

### 2.2.2 Sentiment Orientation Detection Techniques

The determination of the sentiment orientation (polarity) expression associated with each aspect in a sentence is the second task involved in aspect-based sentiment analysis. This task also involves determining whether the sentiment orientation is positive, negative or neutral. It can be broken down into the following subtasks:

1. Extracting opinion words or phrases.
2. Identifying the polarity of each opinion word or phrase.
3. Handling the opinion shifters in the form of negation words like ‘no’, ‘not’, ‘don’t’, ‘لا’, ‘لا السما’ and opinion intensifiers such as ‘very’, ‘extremely’ and ‘ جدا، شديد’.
5. Aggregating opinions, in the cases where multiple opinion words or phrases appear in a single sentence.

Hu and Liu (2004a) developed a distance-based approach for extracting opinion words and phrases, once the aspects had been extracted. Adjectives adjacent to the aspect, within a 3-word distance, were taken to be the opinion words. The study used the WordNet lexicon for determining the polarity of opinion words, taking into
consideration the effect of negation words. The effect of intensifiers was not taken into account in the study.

Table 2.7: Summary of the Aspect Extraction Method Strength and Limitations

<table>
<thead>
<tr>
<th>Method</th>
<th>Technique</th>
<th>Strength</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency-based methods (Hu and Liu, 2004; Popescu and Etzioni, 2007)</td>
<td>Aspect identification through applying a set of constraints on the most frequently-occurring noun phrases.</td>
<td>Simple and effective.</td>
<td>Error in aspect identification, resulting in the output of non-aspects. The manual tuning of parameters is required for low-frequency aspects, thereby negatively affecting the portability to other databases.</td>
</tr>
<tr>
<td>Relation-based methods (Liu et. al, 2005; Baccianella et. al, 2009)</td>
<td>Sentiment expresses opinions on aspects, thereby making it easy to identify relationships which can be used, in turn, to identify other new aspects as well as sentiments.</td>
<td>Ability to identify low-frequency aspects.</td>
<td>Production of non-aspects, which show a match with relation patterns.</td>
</tr>
<tr>
<td>Supervised Learning Techniques (Jin et. al, 2009; Li et. al, 2010)</td>
<td>HMM and CRF, which are state-of-the-art sequential learning methods.</td>
<td>The use of a classifier trained using labelled data, belonging to a domain, which thereby eliminates the limitations of the frequency and relation-based methods, by learning the model parameters from the data.</td>
<td>Need for manually labelled data, for training the classifier.</td>
</tr>
<tr>
<td>Topic Modelling Techniques (Mei et. al., 2007; Titov and McDonald, 2008)</td>
<td>The unsupervised learning approach, operating on the assumption that each document contains a variety of topics with a different probability distribution over words. The output of the topic modelling is a set of word clusters.</td>
<td>Does not require the use of manually-labelled data. Carries out aspect extraction and grouping simultaneously, in an unsupervised manner.</td>
<td>Requires the input of a large volume of (unlabeled) data for accurate training.</td>
</tr>
</tbody>
</table>

In the research by Qiu et al. (2011), the authors conducted a study regarding the application of a propagation-based method of simultaneously extracting opinions and aspects. They tried to discover the natural relations between aspects and opinion bearing
words, since the opinion words describe aspects. The proposed approach begins with initial opinion word seeds, which are subjected to syntactic relations intended to find a link between the aspects and opinion words. The link is subsequently used to find more aspects and opinion words. The process comes to an end when there are no more identifiable aspects or opinion words.

Feature extraction and sentiment determination processes are coupled together. The determination of the orientation of sentiments in each aspect of a sentence was previously studied through two approaches, specifically the supervised learning approach and the lexicon-based approach. The latter was studied first in this research, in order to determine the orientation of sentiments.

2.2.2.1 The Supervised Learning Approach

Wei and Gulla (2010) proposed the use of a hierarchical classification model. The limitation encountered in their study was the particular difficulty faced in determining the scope of each sentiment expression. Jiang et al. (2011) proposed a dependency parser, in which a set of aspect-dependent features was used to perform the classification task. Boiy and Moens (2009) used a related approach, which weighed each feature depending on its location, relative to its target feature in the parse tree.

A model built using a classifier, trained using labelled data belonging to a certain domain, often exhibits poor performance in other domains. The recently-published approaches are applicable to document-level sentiment classification, due to their length and their volume of features, as opposed to individual sentences or clauses. Therefore, supervised learning is limited in its ability to scale up to additional domains.

2.2.2.2 The Lexicon-based Approach

Ding et al. (2008) and Hu and Liu (2004) applied the lexicon-based approach, which is more applicable to multiple domains. This approach contains a list of sentiment words, phrases and idioms, which make use of composite expressions, rules of opinions and a parse tree, for establishing the sentiment orientation of features mentioned in
reviews. The approach also put into account sentiment shifters, such as but-clauses and negations.

Ding et al. (2008) introduced a lexicon-based approach, lexicon-in, which operated using an algorithm which follows a sequence of four stages:

1. Marking sentiment words and phrases
2. Applying sentiment shifters
3. Handling but-clauses
4. Aggregating opinions

The algorithm exhibited good performance results in a majority of instances. Hu and Liu (2004a) determined the orientation by computing the cumulative sentiment score for the sentiment words in each sentence, or in segments of the sentence, whereas Kim and Hovy (2004) and Zhu et al. (2009) used the product of the sentiment scores of words.

Blair-Golden Sohn et al. (2008) proposed a method that integrated the lexicon-based method with supervised learning, in order to enhance the above methods. To make this method even more effective, parsing was used to identify the relationship between the words and the scope of the sentiment words. Once this was completed, it was possible to automatically determine the sentiment orientation of context-dependent words, for instance ‘long’ used in different contexts. Table 2.8 gives an evaluation of different sentiment classification techniques.

### 2.2.2.3 Ontology-based Approach

Approaches in this area aim to organize features in a more elaborate representation model. The model contained an ontology, which differs from a taxonomy. The ontology not only illustrates the hierarchical relationship between concepts, but also

---

51
shows paradigmatic relations (for instance synonymy) or complex relationships (composition or space relationships).

<table>
<thead>
<tr>
<th>Methods</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexicon-based</td>
<td>- Domain independent.</td>
<td>- Low accuracy.</td>
</tr>
<tr>
<td></td>
<td>- Easy to implement.</td>
<td>- Need a sentiment dictionary.</td>
</tr>
<tr>
<td></td>
<td>- Need to have NLP tools such as a stemmer or POS tagger.</td>
<td></td>
</tr>
<tr>
<td>Supervised</td>
<td>- More accurate in comparison to lexicon-based methods.</td>
<td>- Suffers from the inequality of term distribution, between training and test data.</td>
</tr>
<tr>
<td></td>
<td>- No need for any NLP tools.</td>
<td>- Highly dependent on the quality of resource projections.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Cannot put into consideration the intrinsic structure of documents.</td>
</tr>
<tr>
<td>Semi-supervised</td>
<td>- More accurate in comparison to other methods.</td>
<td>- Noise sensitive.</td>
</tr>
<tr>
<td></td>
<td>- Used the information obtained from unlabelled data.</td>
<td>- Has a time-consuming learning phase.</td>
</tr>
<tr>
<td></td>
<td>- Can consider the intrinsic manifold structure of unlabelled data.</td>
<td>- Some techniques may have some incorrectly labelled examples added to the training set.</td>
</tr>
<tr>
<td>Domain adaptation techniques</td>
<td>- Task specificity.</td>
<td>- Highly dependent on the selection of pivot features.</td>
</tr>
<tr>
<td></td>
<td>- The need to pivot translation.</td>
<td>- Very sensitive to the translation of pivot features.</td>
</tr>
</tbody>
</table>

Overall, the extracted features corresponded exclusively to terms in the ontology. The feature extraction phase is guided by a domain ontology, which can be constructed either manually (Zhao and Li, 2009), or semi-automatically (Feiguina, 2006; Cheng and Xu, 2008). The newly-constructed ontology is enriched through extraction and clustering terms, corresponding to new feature identification. Terms were extracted using pattern and terminology extractors, trained to use related product features which are manually identified in reviews by Feiguina (2006). This is followed by the grouping of similar features.

Cheng and Xu (2008) developed an ontology enrichment approach which used a domain glossary containing specialized terms, such as jargon, abbreviations and acronyms. Zhao and Li (2009) developed an ontology using a corpus-based method,
which iteratively extracted concepts from sentences containing a combination of known aspects and conjunctions. The process was repeated until there were no more concepts to be found.

Ontologies served the purpose of supporting polarity mining. Chaovalit and Zhou (2008) constructed an ontology specifically designed for movie reviews, combining it with polarity classification, and which thereby resulted in improved performance. The study by Anaïs Cadilhac et al. (2010) regarding the value of ontology in feature extraction and summary development, as based on a case study of French cuisine restaurant reviews, showed a 0.7692 level of precision and a 0.7733 level of recall.

The fuzzy domain ontology extraction algorithm involves a series of steps which include concept extraction, concept pruning, dimensionality reduction, and fuzzy relation extraction. Fuzzy relation extraction is a process which involves the generation of taxonomic relations, with the help of the Structural Similarity Metric (SSIM) applied in image analysis. Both formal concept analyses as applied by Cimiano et al. (2005), and the fuzzy formal concept analysis, have been used in the automatic construction of a domain ontology. Formal concept analysis by Tho et al. (2006) refers to the process of systematically deriving implicit relationships between concepts, as based on their attributes.

In this study, a simplified version of the fuzzy domain ontology model has been used to represent knowledge on sentiments. The research particularly focuses on developing effective computational methods for obtaining knowledge regarding non-taxonomic relations between concepts, in support of opinion mining.

An econometric opinion mining method was developed for analysing product aspect evaluations, as expressed in online consumer reviews by Hu and Liu (2004). In these reviews the product aspects were in the form of nouns which appeared most frequently. The candidate nouns were then manually filtered out, so as to narrow down on the product aspects. The adjectives adjacent to the aspect were selected as the sentiment words. Opinion phrases, being word pairs that comprise the product aspect
and sentiment, were identified and represented as a vector in the tensor product space.

The relative weight of product features, and the strength of the sentiments relating to a product feature, can be estimated through the use of Hedonic regressions. OPINE uses the ‘relaxation labelling’ classification method, which was developed by the computer visioning research community as a way to detect sentiment polarity (Miao et al., 2008). Similarly, the Feature-Based Summarization (FBS) system was developed for the purpose of extracting product aspects and sentiments from sentences (Popescu and Etzioni, 2005).

The Apriori association rule mining algorithm has been used to extract product aspects, for instance, noun phrases, based on their frequency of appearance in product reviews. The ReviewSeer system uses an N-grams approach for extracting aspects from reviews, coupled with a machine learning approach for classifying sentiment polarity (Wanget. al., 2013; Dave et. al., 2003). In both these systems, specifically Apriori and ReviewSeer, the determination of the polarity of sentiments is not domain-specific.

The Entropy Weighted Genetic Algorithm (EWGA) is used to determine the best syntactic and stylistic features for sentiment classification in multiple languages, across various online forums (Abbasi et. al., 2008). The EWGA algorithm operates by first identifying the most informative features and then, through using this information, the features are passed onto a SVM classifier for polarity classification at the document level. The classification is carried out using the bootstrapping technique, with an accuracy estimated 91%, in comparison to a benchmark movie dataset (Pang et al., 2002).

Sentiment orientations can also be determined through using Probabilistic Latent Semantic Analysis (PLSA), which is defined as being a unigram language modelling technique applicable in analysing movie reviews (Liu et al., 2007). The PLSA model has been used in conjunction with a time series analysis model, referred to as an autoregressive model, as a means of predicting the gross revenues of movies. PLSA is also applicable for combining the opinions shared in an expert review with those retrieved from Web 2.0 sources, such as blogs, to develop a more comprehensive
summary of the opinion regarding a product. Probabilistic models are used to facilitate the identification and ranking of sentiments at the document level.

The domain transfer problem refers to the problem of automatically determining the orientation of sentiments used across different domains (Tan et al., 2008). The Relative Similarity Ranking (RSR) method has been used in the selection of the most informative unlabelled opinion documents within a training set, for the purpose of using them to re-train a classifier (for example, a Support Vector Machine). In this research study, the authors opted to use readily-available sentiment lexicons such as SentiWordNet.

Linguistic rules are used while determining the context-sensitive orientations of sentiments or opinions obtained from online customer reviews (Ding and Liu, 2007). For example, in the sentence “The camera takes great pictures and has a long battery life”, the sentiment ‘long’ is classified as having a positive orientation, because of its association with the positive seeding sentiment ‘great’. The Semantic Orientation (SO) analysis method, which is an opinion mining technique whose operation is based on an inference engine, was developed in order to deal with classification instances based on word context. A word’s SO is determined based on the strength of its association with fourteen other seeding sentiment words. The strength of the association is measured through using the Pointwise Mutual Information (PMI) of a pair of words (Hatzivassiloglou and McKeown, 1997).

Context-sensitive sentiment analysis has been a recently-pertinent topic of interest within the Natural Language Processing (NLP) community (Wilson et. al., 2006). The analysis entails parsing a sentence, and then representing it in the form of a dependency tree. Then, using a set of linguistic features, the AdaBoost classifier is trained so as to be used to predict the sentiment orientation of a word. An appraisal group refers to a set of attribute values, which pertain to a task-independent semantic taxonomy (Whitelaw et. al., 2005). An appraisal group can then be used to analyse the extracted sentiments.

A variant of Pointwise Mutual Information (PMI) can also be used in estimating
the strength of associations between product features and sentiment words (Alfonso et. al., 2016). Besides using fuzzy domain ontology, sentiment polarity can also be established using basic sentiment lexicons. However, instead of using NLP techniques which are complex and consume a significant amount of computational resources, the researchers decided instead to apply a light-weight NLP approach, which facilitates the scalability of the opinion mining system, as the number of customers providing feedback on online platforms grows.

Previous works regarding feature identification in reviews have been based on the observation that features are often expressed in the form of nouns or noun phrases. Popescu and Etzioni (2005) considered product features to be concepts possessing specific relationships with the product in question. Using this basis, they attempted to single out features connected the product name, using corresponding metonymy discriminators. It is worth noting that the approach did not include opinion mining but was instead entirely centred on feature identification. Hu and Liu (2004) proposed the inclusion of opinion mining and summarization to the approach, through the application of a lexicon-based method of determining the positive or negative polarity of an expressed opinion, as pertaining to a certain product feature.

Later research works, specifically that of Hu and Liu (2004), are based on the premise that frequent nouns or noun phrases are often featured and then go on to single out frequent features, through the application of association mining. This is followed by the application of heuristic guided pruning, which intends to be used in firstly removing multi-word candidates where the words do not appear together in a certain order, and secondly, removing single-word candidates for which adsorbed super-strings have been collected. This application involves concentrating on specific concepts, for example, replacing ‘music’ with ‘background music’. This method was improved by Ding and Liu (2007), through the application of a more complex method based on the holistic lexicon and the inclusion of linguistic rules, in facilitating the accurate identification of the orientations of context-dependent opinion words. This approach makes use of special words, phrases and language constructs, which have a subsequent effect on derived opinions, depending on their linguistic patterns. Yi et. al. (2003) explored the application
of three strict heuristics for selecting noun phrases, as based on POS tag patterns. Conjunction rules were applied by Hatzivassiloglou and McKeown (1997), and by Kanayama and Nasukawa (2006), in order to find words from a large domain corpora. The conjunction stipulates that in the case where two opinion words are linked by a conjunction in a sentence, their opinion orientations are considered to be the same.

2.3 Arabic Sentiment Analysis

Sentiment analysis for written Arabic text is not popular among researchers, due to certain limitations (Korayem et. al., 2012). These limitations are discussed through the following sub-section, which discusses the following points:

- Structure and morphology.
- Standard Arabic forms.
- Lack of labelling.
- The absence of an opinion Lexicon for the Arabic language.

2.3.1 Structure and Morphology

The Arabic language has a complex structure and morphology, particularly because of the many different parts of speech that exist. In addition, it is a highly-inflectional and derivative language, contains many word forms as well as special labels called diacritics, which are the equivalent of vowels (Korayem et. al., 2012). For example, the word لحمد ثم ‘It can be tagged as being either a noun phrase when the word ‘علم’ is taken as a ‘flag’, or as a verb phrase if it is taken to be ‘knew’. The same three-letter root can produce different words, which each have differing meanings. When used in conjunction with stemming, the same word may possess different forms with varied diacritics. The Arabic language has different types of sentence structures. For instance, a sentence could begin with a verb, a nominal form, or a noun phrase.
2.3.2 Standard Arabic Forms

A majority of opinions in the Arabic language are not written in classical or modern standard Arabic forms. Such text is difficult to find in domains like movie and product reviews. Indeed the languages used in forums and blogs are mostly dialects, which further complicates the use of semantic approaches in mining opinions. It is important to emphasise that an opinion is limited to a specific locality. For instance, “بلشت اليوم زيادة أسعار مطعم ابل بيز” can be considered either negative or positive, depending on the viewer. A Sudanese individual would consider the statement to be a positive opinion (انتهت، ended), whereas a Lebanese individual would view it as a negative opinion (بدأت, started).

2.3.3 Lack of Labelling

Most of the approaches used in previous works focusing on Arabic Opinion Mining employed supervised learning algorithms, while few existing studies used unsupervised learning algorithms. Therefore the construction of an annotated text corpora is a challenging task. There is unlabelled-classical-Arabic-text available, which needs to be input into a Supervised Learning Algorithm.

2.3.4 Opinion Lexicon

The absence of an opinion lexicon for the Arabic language thwarts the polarity measurement of an extracted subjective text. This heightens the difficulty of sentiment mining, which requires that a semantic analysis of words is conducted, followed by a grammatical analysis of the text. Some of the existing works on the subject have been discussed in prior sections of this chapter, specifically those dealing with the Arabic language.
Ahmad et al. (2006) applied their work to reviewing financial news. First of all, they established domain-specific keywords, by identifying and extracting the most frequently-occurring words in the corpus for financial news. They observed that these words appeared less frequently in a general corpus. Secondly, they established the context of these keywords, and using this context they formulated a local grammar which would be used for extracting sentiment-bearing phrases. They went on to extend the application of this approach to Arabic, English and Chinese languages. Upon conducting a manual evaluation of the system, it was found to have accuracy levels which ranged from 60 to 75%, in regards to the extraction of sentiment-bearing phrases. The researchers made the assertion that their proposed approach was applicable to extracting sentiment phrases in the financial domain, as written in any language.

Abbasi et al. (2008) coined the Entropy Weighted Genetic Algorithm (EWGA), a feature selection algorithm which made use of syntactic and stylistic features at the document level. The algorithm actually combined genetic algorithms with information gain (IG), in order to carry out feature selection for both Arabic and English language texts. Specifically, IG was used for the initial collection of the feature set in the genetic algorithm and was also used in the cross-over and mutation stages. EWGA is used in feature selection for sentiment analysis, from a multilingual corpus. Semantic features are excluded due to their language dependency, the need for lexicon features, and the difficulty of including linking features. The study evaluated the proposed system on a benchmark test, consisting of 1,000 positive and 1,000 negative movie reviews. The proposed algorithm was compared with a benchmark test, made up of the same numbers of positive and negative movie reviews. The comparison revealed that the proposed algorithm achieved a 91% accuracy rate, in comparison to the accuracy rates of other systems which remained between 87 and 90%. The EWGA algorithm also exhibited a 92% level of accuracy, when run on Middle Eastern forums, and a 90% accuracy when run on US forums.

Elhawary and Elfeky (2010) used Arabic financial reviews to build a system for sentiment analysis, with the intention of setting up a web search engine which would automatically annotate the returned pages with sentiment scores. Their proposed system
would be made up of a handful of components, with the first component being responsible for classifying an internet page as either containing or not containing a review. The task of the classifier was to tag the Arabic web pages accordingly, from a set consisting of a review, forum, blog, news or shopping store. They collected 2,000 URLs, more than 40% of them reviewed through manual labelling, so as to come up with an Arabic review classifier data set. The dataset was constructed by searching the web using keywords found in the reviews, such as "the camera is very bad". Elhawary and Elfeky then translated the keyword lists, adding them to the Arabic keywords list. Eventually the list was made up of 1,500 features and was used in the construction of an AdaBoost classifier. This classifier used 80% of the data for training, and the remaining 20% for testing. Once a document was classified as an Arabic review, a second component extracted the sentiments. Using a similarity graph, an Arabic lexicon was constructed for implementation with the sentiment component. Eventually, a search engine was developed in which sentiment scores were assigned to a document in the course of the search.

Farra et al. (2010) proposed two methods of sentence-level sentiment analysis, the first making use of the grammatical features of the Arabic language, combined with the nominal sentence structures in the general form, using the notion of actor and action. The method considered that subjects in verbal and nominal sentences were actors, while verbs were actions. In addition, manual POS word tags were used as features for vectors. A feature vector was made up of the following dimensions, including sentence type (for instance, either verbal or nominal), actor, action, object, adjective, pronoun and noun types, transition, word polarity (for instance, positive, negative or neutral), and also sentence class. Using the SVM, the classifier achieved an accuracy level of 80%.

The second method proposed by Farra et al. (2010) was combined with syntactic and semantic features, such as frequency of negation, opinionated words (positive, negative, and neutral words), and special emphasis words (such as ‘really’ and ‘especially’). In the study, a semantic interactive learning dictionary was constructed in which the semantic polarity of word roots, extracted with the help of a stemmer, could be stored. The grammatical method was evaluated using 29 sentences, which were
manually annotated with Parts-of-Speech tags. A comparison revealed that an accuracy level of 89.3% was achieved using the SVM classifier, with 10-fold cross validation. Furthermore, the accuracy of classification ranged from 60% to 80%. A manual evaluation was undertaken using an interactive dictionary, which resulted in errors arising from the fact that multiple words with different polarity shared a stem, thereby resulting in incorrect tags. The semantic and syntactic methods were evaluated using a J48 decision tree classifier, after which the results showed an 80% level of accuracy using manual classification, and 62% accuracy when a dictionary was used. Documents were classified using the entire range of sentence features, by splitting them into sections. An evaluation of the approach showed an accuracy level of 87% when a SVM classifier was utilized in documents, which had been split into 4 chunks and excluded from the neutral class.

The work of Rushdi-Saleh and Martín-Valdivia (2011) used the supervised learning approach to build classifiers, using OCA and EVOCA corpora obtained from movie reviews. They also used Support Vector Machines (SVMs) and Naïve Bayes (NB) classifiers, whose performance showed a 90% F-measure on OCA and an 86.9% recall on EVOCA, while using SVMs. The study proved that SVMs outperformed the NB classifier, in terms of text classification tasks. The proposed work showed that the use of Term Frequency (TF), as opposed to Term Frequency-Inverse Document Frequency (TF-IDF), as weighing schemes was inconsequential. Experiments also showed there is no need for stemming words, prior to extracting features and classification because it has a negative consequence on the results.

El-Halees (2011) proposed a combined classification approach for document-level sentiment classification, through the consecutive use of a series of classifiers. Firstly, a lexicon-based classifier was used to determine a document’s sentiment as based on the aggregate of all the opinion words and phrases. However, due to the shortage of opinion words, some documents required the use of a lexicon-based classifier. Secondly, a maximum entropy classifier was used. The output from the first classifier were the classified documents, which were then used as the training set for the second classifier, which was used to compute the probability of a document falling in a
certain sentiment class. In the case where this probability was greater than 0.75, the document was assigned a class, and otherwise the document was passed on to the next stage. Finally, a k-nearest neighbour (KNN) classifier was applied, so as to find the nearest neighbours in the unannotated document. This process involved using the training set applied in the previous step. The corpus used was made up of 1134 documents from different domains, including education, politics and sports. Out of these, there were 635 positive documents containing 4375 positive sentences, and 508 negative documents containing 4118 negative sentences. The pre-processing involved removing HTML tags and non-textual content, the correction of misspelt words, and the normalization of alphabets. Tokenization was done by removing stop words, stemming the words using an Arabic light stemmer. The TF-IDF weighing was applied in the study. An evaluation of the performance showed an average F-measure of 81.70% across all domains for positive documents, and an average F-measure 78.09% for negative documents. The best F-measure was in the education domain, at 85.57% for the positive class, and at 82.86% for the negative class.

In the field of knowledge-based techniques, a study conducted by Al-Subaihin et al. (2011) using knowledge-based techniques, resulted in the proposal of a new tool for the implementation of Arabic sentiment analysis. The tool accepts informal Arabic language as its input and uses a combination of two techniques within the proposed system, which include natural language processing and human computation. The proposed work contained two components, specifically game-based lexicon and sentiment analyser. The construction of a lexicon based on human computation involved the use of an online computer game, the game presenting many phrases and words extracted from Qaym.com. Then, each player had to determine whether the extracted words were positive, negative or neutral. In the study, the lexicon was constructed automatically, so as to avoid the problems that are commonplace within manual construction. Sentence patterns are another output of the game, which is made up of positive, negative, natural and negation tags, whose polarities are stored in a database. The second part of the tool was the sentiment analyser, responsible for the segmentation of sentences. Afterwards, the words in each sentence were tagged as POS, NEG, ENT or NO, representing positive, negative, neutral and negation words, in accordance with the
game-based lexicon. After tagging, the sentence polarities were detected through matching the result with the stored values of the database. The review polarity was based on the maximum polarities, and the Opinion Corpus for Arabic (OCA), which is a corpus of text from movie review sites, as developed by Rushdi’saleh et al. (2011). OCA included a parallel English version, referred to as EVOCA, which contained 500 reviews, half negative and half positive. The raw reviews were filtered, in order to remove any false or unrelated comments, such as cultural and politically-motivated posts. OCA and EVOCA carried out a standard pre-processing of the corpus, correcting misspelt words, and deleting special characters. The resulting outputs from this work were unigrams, bigrams, and trigrams for the dataset MPQA subjective lexicon, and the Arabic opinion holder corpus.

Abdul-Mageed and Korayem (2010, 2011) extended the previous work by classifying MSA news data at the sentence level, for both subjectivity and sentiment, using a two-stage SVM classifier. First of all, a subjectivity classifier was applied in order to separate the subjective sentences from the objective ones. Secondly, the already-classified subjective sentences were split into positive and negative categories. The study made use of both language-independent and Arabic-specific features, which resulted in the achievement of 95.52% accuracy. The results confirmed the importance of the adjective feature, since it caused a 20% improvement in accuracy. The unique and domain features were equally helpful.

Abdul-Mageed, Korayem, and Diab (2010, 2011, 2012) built subjectivity and sentiment analysis systems, which can work with sentence-level annotated Arabic corpora. In their proposed work, language-independent features, Arabic-specific morphological features, and genre-specific features were all used.

In another study Abdul-Majeed et al.(2012) presented SAMAR, which is a SVM-based system for Subjectivity and Sentiment Analysis (SSA) of the Arabic social media genres and can tackle the problem of sentiment analysis using social media as a source, considering a predominantly linguistic perspective. The technique included lexical information presentation, investigated the importance of standard features, considered the handling of Arabic dialects, and examined the significance of genre-specific features.
Their proposed work was based on support vector machine (SVM) classifiers and carried out SO determination in a series of two steps, specifically the differentiation between subjective (opinionated) and objective cases, and secondly determining polarity of the subjective input using another classifier. The system specifically excluded neutral and mixed cases. The features used by the classifiers included morphological features, Parts-of-Speech (POS) tags, and adjectives. The system’s dialectical performance was evaluated using a tagged dataset made up of 3015 Arabic tweets, 1466 of which were written in MSA, and the rest in different dialects. 80% of the dataset was used for training, 10% for developments, and the remaining 10% for testing. The highest accuracy was observed in dialect-specific sentiment experiments at 71.15%, with F-score of 29.4% for positive cases, and 81.8% F-score for negative cases. The dialect specific dataset was primarily made up of negative tweets, in order to balance the low positive F-score, which contributed towards the overall accuracy.

Mountassir et al. (2012) investigated sentiment classification for Arabic text through a study containing two Arabic corpora of varied sizes, specifically ACOM and OCA. ACOM was developed using news from the Aljazeera website, which was manually annotated into two categories, including positive and negative. It had two data sets, including DS1 which is a collection of 368 comments on a series of reviews, and DS2 which is a collection of 1000 comments in the sports domain. OCA, on the other hand, is a dataset of movie reviews as collected by Rushdi’ Saleh et al. (2011). The study was concerned with conducting an investigation of settings, such as stemming type, term frequency threshold, term weighing, and the N-grams words model. This process was facilitated by the use of three common classifiers, specifically Naïve Bayes, Support Vector Machines, and k-Nearest Neighbour. The classification results of the three classifiers, were applied in the perspective of the Arabic context. The results of the study revealed that the most suitable setting for the classifiers was light stemming, with the elimination of stopwords as the threshold, the combination of unigrams and bigrams, and a presence-based weighing approach.

Misbah and Imam (2012) presented an optimized approach for mining opinions in
Arabic Religious decrees, using an improved semantic orientation and a pointwise mutual information algorithm. Their approach was executed in a series of steps, so as to facilitate the classification of a religious decree as being either *halal* (allowed), or *haram* (prohibited) products. The research included data collection, simple text pre-processing, manual data labelling, advanced text pre-processing, weight calculation, and experimentation using supervised and unsupervised learning algorithms. Results from the original approach showed an accuracy rate of 73.08%. Their proposed approach utilized an improved SO-PMI algorithm, which incorporated advanced steps in the calculation of weights. This improved algorithm increased the accuracy of the unsupervised learning algorithm up to 2,000 but exhibited poor results with the supervised learning algorithm. The recommendation came out that subjectivity classification should be executed before advanced text pre-processing. For this, a classifier is required for classifying sentences as being either objective or subjective. Subjective sentences are required to be checked for their relevance to the asked question, and only sentences which are correlated with the question should be used. Next, advanced text pre-processing and weight calculation are proposed for the extracted sentences.

With all these improvements, it would be expected that the extracted tokens would be in reality, opinion-oriented tokens. The tokens will also be closely correlated with the topic of the decree. The accuracy rate is expected to increase for the sentiment classification of the decrees.

Itani et al. (2012) presented two approaches for the classification of Facebook posts in Arabic language, which are discussed in the following two paragraphs. The first approach used common patterns in Arabic dialects, to express opinions based on the syntactic features. The use of these patterns resulted in high levels of accuracy in terms of the polarity determination of a sentiment, even after conducting tests using the new corpus. This approach was found to be effective for informal Arabic text, which had not been addressed in the past. Different setups were tried, and the highest measures of coverage and accuracy achieved were 49.5% and 83.4%, respectively. The second approach used was the Naïve Bayes classifier, which applied an ordinary probabilistic
model. The approach was based on the assumption that features are independent, while determining the class with the highest coverage. The results obtained showed a coverage of 60.5% for the first approach, and 91.2% coverage when the Naïve search was used as a binary classifier for classify the posts as being either objective or subjective. Table 2.9 summarize the different Arabic lexicons.

2.4 Aspect-based Sentiment Analysis in Arabic

The International Workshop on Semantic Evaluation 2016 (SemEval-2016), established a data set for the task of aspect sentiment analysis, with a sub-task for aspect extraction, but the task for Arabic aspect-based extraction has not received any submission (Pontiki et al., 2016). Al-Smadi et al. (2015) have recently established aspect-based annotation for a book review dataset (LABR). The data set has a baseline for aspect sentiment classification tasks, but to the best of the author’s knowledge, there is no published work regarding the aspect extraction methods on similar Arabic reviews.

Elarnaoty et al. (2012) proposed the use of an opinion holder and a subjectivity lexicon. In the study, an Arabic news corpus was created by crawling 150 MB of Arabic news, and by manually annotating 1 MB of the corpus for the opinion holder. A majority vote was used to remove any conflicts. Pre-processing of the corpus was undertaken through the morphological analysis of Arabic sentences, which were then assigned Parts-of-Speech (POS) tags, using a Research and Development International (RDI) tool. A semantic analysis of the words was then carried out. Arabic Named Entity Recognition (ANER) was used for extracting names from documents (Liu, 2010).

Alkadri and ElKorany (2016) proposed an ontology for classifying Arabic views. They manually collected the test reviews from a variety of related websites, and manually tagged them for evaluation in their proposed method. The level of accuracy achieved remained at 72%.
<table>
<thead>
<tr>
<th>Approach</th>
<th>Performance</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local grammar- sentiment-bearing phrases.</td>
<td>60% - 75% accuracy</td>
<td>Ahmad et al. (2006, 2007)</td>
</tr>
<tr>
<td>Entropy Weighted Genetic Algorithm (EWGA)- syntactic and stylistic features.</td>
<td>91% accuracy</td>
<td>Abbasi et al. (2008)</td>
</tr>
<tr>
<td>Notion of actor/action.</td>
<td>80% accuracy</td>
<td>Farra et al. (2010)</td>
</tr>
<tr>
<td>Syntactic and semantic features.</td>
<td>87% accuracy</td>
<td>Farra et al. (2010)</td>
</tr>
<tr>
<td>Supervised learning approach - Support Vector Machines (SVMs), and NaiveBayes (NB) classifiers.</td>
<td>90% F-measure on OCA and 86.9% recall on EVOCA, while using SVMs.</td>
<td>Rushdi-Saleh and Martín-Valdivia(2011)</td>
</tr>
<tr>
<td>Consecutive use of a series of classifiers(lexicon-based classifier, maximum entropy, k-nearest).</td>
<td>F-measure of 81.70% positive and 78.09% for negative(best F-measure was in the education domain at 85.57% for the positive class, and 82.86% for the negative class).</td>
<td>El-Halees(2011)</td>
</tr>
<tr>
<td>Knowledge-based techniques(natural language processing and human computation).</td>
<td></td>
<td>Al-Subaihin et al. (2011)</td>
</tr>
<tr>
<td>Language-independent features, Arabic-specific morphological features, and genre-specific features were used.</td>
<td>Accuracy</td>
<td>Ageed, Korayem, and Diab.(2010, 2011, 2012)</td>
</tr>
<tr>
<td>Lexical information presentation, and genre-specific features. Support Vector Machine (SVM).</td>
<td>71.15% accuracy, with a F-score of 29.4% positive and 81.8% negative</td>
<td>Abdul-Majeed et al. (2012)</td>
</tr>
<tr>
<td>Common classifiers: Naïve Bayes, Support Vector Machines, and k-Nearest Neighbour.</td>
<td></td>
<td>Mountassir et al.(2012)</td>
</tr>
<tr>
<td>Semantic orientation, and a pointwise mutual information algorithm.(improved SO-PMI algorithm).</td>
<td>73.08% accuracy</td>
<td>Misbah and Imam(2012)</td>
</tr>
<tr>
<td>Common patterns in Arabic dialects, regarding syntactic features.</td>
<td>49.5% and 83.4% accuracy achieved, respectively</td>
<td>Itani et al.(2012)</td>
</tr>
<tr>
<td>The Naïve Bayes classifier, using an ordinary probabilistic model.</td>
<td>Accuracy ranging from 60.5% to 91.2%</td>
<td>Itani et al.(2012)</td>
</tr>
</tbody>
</table>

2.5 Preprocessing of Arabic tweets

Most of the text generated by social media contains noise and is not
appropriately structured, due to the non-standardization of language, spelling errors, incorrect punctuation, short forms and redundancy, and therefore requires text cleansing (Al-Shammari, 2009). This type of text pre-processing is divided into three steps, which includes normalization, stemming, and the removal of stop words. As the first step, normalization serves the purpose of transforming the text into a common form, so as to maintain its consistency while reading. As the second step, stemming converts back the derived words to their original or base words. As the third step, stop word removal removes unnecessary words which are of no consequence to the overall sentence’s meaning, such as ‘في’ (in), ‘على’ (on), ‘من’ (of), and many others in the Arabic dialect language.

2.5.1 Normalization

The normalization of Arabic text ensures text consistency. The process converts back all complicated variants of the same words to the most common form. For example, the word ‘الق’ (الق) can have multiple forms, such as ‘ألف’, ‘ألف’, ‘الق’, and others. In this sense, language detectors consider them to be three separate words. However, for the requirement of language processing, all these forms need to be converted into one common form, which is achieved by applying a normaliser run in Java. This specific normaliser removes diacritics from the letters, for instance by eliminating ‘ا’ (Hamza) for (ا, و, and ی), which are replaced with (bare Alif, ی), and in the same manner, (تاء, هاء, ‘و’, ‘ي’) becomes (هاء, ی). Both (ياء’, ‘ي’، ‘ي’) are converted into (ياء‘، ی‘). The normaliser is considered to be worthwhile, as it is efficient and capable of handling multiple rules simultaneously. Table 2.10 presents the language normalization rules.
### Table 2.10: Normalization Rules

<table>
<thead>
<tr>
<th>Rule</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamza</td>
<td>إ, ى, و → ء</td>
</tr>
<tr>
<td>Alif</td>
<td>أ, ا, آ → ح</td>
</tr>
<tr>
<td>Yaa</td>
<td>ت, ١ → ١</td>
</tr>
<tr>
<td>Haa</td>
<td>ٙ, س → ٙ</td>
</tr>
<tr>
<td>Tashkeel</td>
<td>جميل جدا → جميل جدا</td>
</tr>
</tbody>
</table>

#### 2.5.2 Stemming

Stemming is the process of reducing words to their uninflected base forms. In some situations, the stemmed word can be entirely different from the root, but similar in its contextual meaning. Stem detection maps different words, contextually, within a future estimation that it may be useful, as it is related to the same root. Stemming is quite significant in the Arabic language processing systems. Larkey et al. (2007) explained the complicate process of stemming in Arabic text, complication which is mainly due to high inflexion and the derivation of words. There are two main types of stemming. The first type is aggressive stemming, which reduces all the derivation of words right down to the basic root. The second type is light stemming, which simply adds suffixes and prefixes, which can be further processed intelligently. Aggressive stemmers can result in the loss of valuable meaning in the course of the reduction process, as there is no relation between the subject and the root. The process is somewhat irreversible.

Therefore, aggressive stemmers are inappropriate candidates for systems that require high accuracy, like those for general legal information or psychological analysis. However, on the contrary, they are fast, simple, and the best choice for real-time systems requiring a high response rate with lower accuracy, such as trend analysis or market surveys.

Based on the complexity of the system, a stemmer needs to respond differently in different scenarios. This leads to research on different stemmers, with specific
complexity levels, as required by the addressed information and underlying text mining system. The studies in this area mainly target the Modern and Standard Arabic (MSA) language, with little emphasis placed on the handling of dialect versions of Arabic.

As an example, the application of a stemmer on the word عَلَّامٌ (because), results in its reduction to عَمٌّ (hut). This is because in MSA, when a word ends in ان‘, it reflects duality. This indicates that the word does not need to be stemmed in the first place. This fact led to the application of a customized light stemmer, which served to reduce input words to their shortest form, while maintaining their original and contextual meanings. MSA, being different contextually, is suitable for an aggressive stemmer that maps words to their root. In the case of the aggressive stemming reduction of subject words to their roots, this leads to the unnecessary mapping of multiple terms to a single word, or root, which results in the loss of its unique meaning. The light stemming technique is not only efficient in implementation, but is also effective in information retrieval. One less preferable application of light stemmers is in affixes and broken plurals, which are commonly found in MSA (Larkey et al., 2007).

The customized stemmer uses a combination of rules detailed in research by El-Beltagy and Rafea (2011) and Shoukry and Amira Magdy (2013), with custom rules that handle broken plurals in the Arabic dialect. The customized stemmer constructs two lists, including a list for irregular words, beginning with a prefix or ending with a suffix which requires no stemming, and a list for irregular plurals and singular forms. Both lists are then subjected to normalization and stemming. Each input word is, first of all, evaluated to see if a match exists in either of the two lists. Once a match is found the word is not stemmed. Otherwise, it is flagged for stemming.

This stemmer works through a series of three sequential steps, including prefix removal, suffix removal, and infix removal, for broken plurals. After every step, a decision is made regarding the bases of transformation which are checked from a dictionary. A great deal of the broken plurals’ rules represents adaptations from El-Beltagy and Rafea (2011).
A common problem with DA, is that users tend to write the same word in different ways, as shown in Table 2.11. These spelling variations cannot be easily fixed, due to the lack of a gold standard Arabic DA dictionary. To alleviate this problem, the researchers resorted to simple preprocessing steps, which include the following:

1. Removing punctuation, diacritics and any non-characters.
2. Normalizing letters based on the normalization Rules in Table 2.10.
3. Removing extensions from words - for example، ħāʾ is reduced to ħāʾ.

These steps have been previously used in literature, for similar tasks, and it have been shown to increase accuracy in terms of sentiment analysis (Ahmed et al., 2014). Although these steps will not overcome spelling variations, they will help alleviate the problem by reducing the form for words that have some similar variations. This can be seen in Table 2.11 which have spelling variations in one letter, or sequence of different letters, or in diacritics.

2.5.3 Corpus Development

Al-Sabbagh and Girju (2012) used three venues for building YADAC, including a multi-genre Dialectal Arabic (DA) corpus compilation, Twitter-API-based search engines, online knowledge market services, and blog-based search engines. For each, two different search engines were used to overcome the upper-bound limit of the returned search results, that each search engine sets for each query. Generic queries, each of which consisted of a minimum of three function words, were automatically created by permuting the entries of a 1,527 EA-exclusive function words list. Function words are used to create topic-independent search queries, and there by broadening the search scope and harvesting more data. Out of the created permutations, 15 million search queries were randomly selected and were used to crawl the Web over a period of 7 months, from May 2011 to November 2011.
Table 2.11: Variety of Word Spellings in Arabic Dialects

<table>
<thead>
<tr>
<th>Word</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eat</td>
<td>اكل</td>
<td>حًَ</td>
<td></td>
</tr>
<tr>
<td>Meal</td>
<td>وَجَيَه</td>
<td>جوعانه</td>
<td></td>
</tr>
<tr>
<td>Very</td>
<td>جدا</td>
<td>جدا</td>
<td>جدا</td>
</tr>
<tr>
<td>Hungry</td>
<td>جعَنَه</td>
<td>جوعانه</td>
<td>عطِيشانه</td>
</tr>
<tr>
<td>Thirsty</td>
<td>عطِيشانه</td>
<td>سِوء</td>
<td>ضوء</td>
</tr>
<tr>
<td>Bad</td>
<td>ضوء</td>
<td>ضوء</td>
<td>ضوء</td>
</tr>
<tr>
<td>Light</td>
<td>ضوء</td>
<td>ضوء</td>
<td>ضوء</td>
</tr>
</tbody>
</table>

Eshrag Refaee and Verena Rieser (2014) used the Twitter Search Application Programming Interface (API) for corpus collection, allowing for harvesting a stream of real-time tweets by querying their content. They create a set of search queries, in order to increase the chance of obtaining tweets that convey opinions, attitudes or emotions towards specified entities. An example of this is in Al-Sabbagh and Girju (2012). Note that for training a classifier, these query terms are replaced by placeholders. The extracted data is cleaned in a preprocessing step, for example by normalizing user-names and digits and eliminating Latin characters, such as those in URLs and emails. In particular, the researchers harvested two datasets at two different time steps. These datasets included Development Data, with 7,503 multi dialectal Arabic tweets randomly retrieved over the period from the 20\textsuperscript{th} of January to the 21\textsuperscript{st} of February, 2014, and the Test Data which contained 1,365 instances retrieved during the period from the 6\textsuperscript{th} to 15\textsuperscript{th} of November, 2013.

In related work, the researchers used a similar data set for training and evaluation purposes (Refaee and Rieser, 2014b). It should be noted that the data set described here is different, since the researchers were not allowed to release the original tweets. For further details regarding the release format, see Section 3. The researchers manually annotated a random subset of 8,868 examples of the collected data for subjectivity, for instance by testing apart subjective/polar and objective tweets, and sentiments. Following Wilson et al. (2009) the researchers defined a sentiment as being a positive or
negative emotion, opinion, or attitude. Each data instance (tweet) was marked with only a single tag, denoting that the interpretations was ultimately conveyed by a complete piece of text, taking into account only the writer’s perspectives. The latter were considered as being neutral, mixed, positive and negative, where the latter two are both subsumed under the label polar, for instance the subjective, as seen in Table 2. The mixed label covers cases where tweets are simultaneously composed of both positive and negative emotions (Liu, 2012).

### 2.6 Discussion

Sentiment analysis faces several problems. The first, identifying the set of words that can identify the polarity in the text, is generally a difficult process. Many adjectives are domain dependent, an example being ‘The battery life is long’ vs ‘takes a long time to boot’. Similarly, sentiment and subjectivity are context-sensitive. The sentence “reading the book was very enjoyable” is negative in the movie review context, but positive in the book review context. Finally, some opinions are expressed in idioms and not as individual words. An example of this, is when something can be described as ‘costing an arm and a leg’.

The last problem is that semantics depends a lot on word sequence and sentence structure. Saying “Mac is more expensive than Windows” is not the same as “Windows is more expensive than Mac”. Unfortunately, this problem has not been explored for the Arabic language. Ahmed et al. (2014) relates the lack of reliable Arabic NLP resources, such as a reliable syntactic parser, as the main problem behind not exploring this problem.

The Arabic Language is divided into three types, which include Classical Arabic, Modern Standard Arabic (MSA), and Dialect Arabic (DA) (Soliman et al., 2014). The Arabic language has many different dialects that are used in informal daily communications but are not standardized or taught formally in schools. While there are a variety of dialects, MSA is the only one standard form that has been widely recognized.
and formally taught in schools. MSA is based on Classical Arabic, which is the language of the Qur’an, which is the Islamic holy book (Habash, 2010). MSA is not a native language of any country, and it is largely different from dialect forms. MSA has been studied extensively, and many NLP tools are available for it. Unfortunately, most of the Arabic web content is written in a dialectal form, which has not been extensively studied. To the best of the researchers’ knowledge, there exists no reliable NLP tools for Dialect Arabic.

Arabic is a morphologically-rich language (MRL), where most information regarding syntax and relation is expressed at the word level. English, on the other hand, has much less information expressed at the word level. The Arabic base form of a word can lead to thousands of surface forms, while in English a verb would have three different forms. Therefore, using those forms in a lexicon corpus will lead to data sparseness in Arabic, while in English there is a high chance that the three terms will be present in text (Abdul-Mageed et al., 2012; Ahmed et al., 2014). This suggests that using a compact form of the word, along with POS tagging, will help overcome the problem of data sparseness. Albraheem and Al-Khalifa (2012) also recommended the use of stemming to reduce the size of the lexicon corpus. On the other hand, Rushdi-Saleh et al.(2011) does not recommend the use of stemming for the purpose of opinion mining.

The second challenge related to Arabic is the lack of widely-available Arabic corpora (Abdul-Mageed & Diab, 2012a). There is also a lack of an Arabic lexicon that can be used for sentiment, and a lack of publicly available and reliable NLP tools such as a Parts-of-Speech tagger, and a dependency parser.

The use of the informal form in web content leads to many problems. Arabic users encode Arabic words in the Roman alphabet. For example ‘ الحرب’, which means ‘war’, is written as ‘Al7arb’or ‘Al 7arb’, and there are no defined standards about how this is done. This means that each word can have different variations, depending on the individual writing it (Ahmed et al., 2014). Also, Albraheem and Al-Khalifa (2012), in their study of problems related to DA, indicated that different words with different meanings have the same root. This can impact SA if the wrong root has a different sentiment. Appendix A covers, in much more detail, the nature of the Arabic language.
2.7 Summary

The existing sentiment classification methods were extensively reviewed in this chapter. Sentiment classification has been given a thorough evaluation at different levels of granularity, including at the document, sentence, word and aspect levels. In document-level sentiment classification, the two approaches used included the supervised machine learning approach, and the unsupervised lexicon-based approach. Supervised approaches require a large set of labelled training corpus, in order to learn a sentiment classification model. It was found that in many languages, there is insufficient training data used for supervised sentiment classification methods, and that the labelling of sentiment data is a difficult and time-consuming task. Aspect-based sentiment analysis requires finer-grained analysis, in which all entities and their aspects should be extracted, and the sentiment should be determined accordingly. This requires natural language processing techniques for the extraction and determination of the sentiment orientation of each aspect. Therefore, sentiment classification is a promising potential research area to proceed with.
CHAPTER THREE

3 RESEARCH METHODOLOGY

3.1 Introduction

This research aims to improve the effectiveness of the aspect-based sentiment analysis of Arabic text, and develops a hybrid suitable for the classification and aspect extraction of Arabic text, particularly for restaurant Arabic tweets. It explains the methodology used for developing the proposed Twitter Arabic Sentiment Classifier (TASC) model and the Hybrid Arabic Sentiment Analyzer (HASA) model. A description of the framework and research model used to attain this goal is described through the next section, providing details on its design and implementation. The research study is conducted in six phases. The schematic diagram of the operational framework for this study is shown in Figure 3.1, in which the research phases are illustrated diagrammatically. This framework also depicts the relationships between proposed models and problems to be solved through current research.

This approach incorporates semantic features with language features, which are tested and proven to be vital in sentiment analysis. This is used to come up with a semi-supervised model, which is a considerably-accurate model.

In phase A, the previous research studies related to Arabic sentiment analysis and aspect-based sentiment classification were basically investigated and analysed through the use of a systematic literature review, leading to the identification of problems existing in Arabic sentiment classification methods (Ibrahim, M.A. and Salim, N., 2013). Furthermore, the primary planning of activities, such as problem formulation,
data collection, data pre-processing and identifying performance metrics, were performed in this phase.

In phase B, the opinion corpora construction, the main objective is to prepare three opinion corpora for the research purpose from the Twitter (www.twitter.com), Qaym (http://www.qaym.com) and Trip Advisor (https://ar.tripadvisor.com) websites. The three corpora contain reviews of restaurants written in the Arabic language.

In phase C, a classification model of Arabic tweets (Twitter Arabic Sentiment Classifier, TASC) was designed and implemented, as based on the document level. Here a large scale opinion lexicon has been used to determine document orientation. Frequency and relation-based approaches have been used to build the classification model. Manual annotation corpus has also been employed as a benchmarking dataset for evaluation. Figure 3.1 explains the proposed phase.

In phase D, patterns modelling and restaurant ontology construction were constructed for aspect extraction and were used with TASC to enhance model accuracy. This model tries to cover more vocabularies from the target domain (restaurants), by using different feature sets extracted from customer reviews.

In phase E, the analyser model of the Hybrid Arabic Sentiment Analyzer (HASA), which is an aspect-based opinion mining method, has been proposed for building the model. This model consists of two sub-phases. The first phase is used for identifying aspects and their orientation, while the second one is used for generating a summary.

Analysis of the results, research findings and limitations of the models, conclusions, and recommendations, were stated in the last phase of this research.

This chapter is organized in the following way. Section 3.2 describes the Twitter sentiment classification framework. Section 3.3 describes Phase A, which is primary studies and initial planning. Section 3.4 describes Phase B, which is the construction of opinion corpuses. Section 3.5 describes Phase C, which is the classification model Twitter Arabic Sentiment Classifier (TASC). Section 3.6 presents Phase D, which is
patterns of modelling and restaurant ontology construction. Section 3.7 offers details about the Phase E analyser model, which is the Hybrid Arabic Sentiment Analyzer (HASA). Section 3.8 illustrates the result analysis, findings and the conclusion. Lastly, Section 3.9 lists the evaluation measures for the proposed approaches.

3.2 Twitter Sentiment Classification Framework

As mentioned in the first chapter, the main goal of this research is to propose an aspect-based sentiment classification framework in which the research problems are considered, as a means of improving the classification performance. Figure 3.2 shows this framework, which includes the proposed models designed to overcome the aforementioned problems.

The following sections of this chapter describe the procedural phases involved in the development of the proposed approaches for this research.

3.3 Phase A: Primary Studies and Initial Planning

3.3.1 Existing Literature Analysis and Problem Discovery

In this research, the main problems within the field of Arabic sentiment analysis were determined through several steps. Firstly, a systematic literature review was conducted to generally investigate current methods in sentiment classification, considering the Arabic language as being a special case. Previous research in the Arabic sentiment analysis has been reviewed, in order to investigate the methods used for Arabic language sentiment analysis. Secondly, a traditional literature review has been conducted.
Figure 3.1: Operational Research Frameworks
Input Reviews → Preprocessing
  - Tokenization & Noise Removal
  - Stemming & Stop Word Removal
  - POS Tagging & N-grams Generation
  Pre-processed Review →
  Candidate Aspect Selection
  - Select Patterns
  - Identify Subjectivity
  - Extract Candidate Aspects
  Candidate Aspects List →
  Aspect extraction
  - Apply FP-growth to Find Frequent Aspects
  Frequent Aspects List →
  Apply Semantic Relation Through Ontology
  - Infrequent Aspects List
  Infrequent Aspects List →
  Grouping Candidate Aspects
  Ontology →
  Aspects →
  Polarity Identification →
  Sentiment Analysis and Summary Generation

Figure 3.2: Proposed HASA Frameworks
Based on the outcome of the literature review, as detailed in Chapter 2, two categories for sentiment classification methods have been identified. These include machine learning-based or supervised, and lexicon-based methods. Supervised approaches train a sentiment classifier, based on labelled data, using machine learning classification algorithms. This classifier is then used to predict the sentiment polarity of the unlabelled test data. The lexicon-based methods, on the other hand, establishes a document’s sentiment polarity by looking at the semantic orientation of words and phrases in it. Different methods in each of these two categories have been studied in the review.

An examination of machine learning and lexicon-based methods has indicated that the machine learning method has a better performance than the lexicon-based methods. However, machine learning methods depend greatly on the quality and quantity of the labelled data, when training a sentiment classifier. Given that a great deal of the research in the area of sentiment classification has been conducted in the English language, there is a limited amount of labelled data in other languages such as Arabic. Therefore, one of the contributions of this study is the construction of a manual and annotation of Arabic corpus, which is employed as a benchmarking dataset for evaluation. Figure 3.1 illustrates this proposed phase.

3.3.2 Datasets

The exploration of the performance of various classification models requires the use of real data sets. Unfortunately, there is no standard Arabic dataset in this area. Different researchers have used their own sentiment datasets in their studies. The same approach is adopted in this research, using one scientific dataset that is available for sentiment analysis, referred to as the Twitter Dataset for Arabic Sentiment Analysis.\textsuperscript{10}

---

\textsuperscript{10}https://archive.ics.uci.edu
The characteristics of the aforementioned datasets are described in the following subsections.

As a result of this limitation, the researchers opted to use Twitter’s API\(^{11}\) for retrieving Arabic tweets by setting (lang=ar). It was also crucial that a large collection of Arabic sentences was obtained for training the classifier, and to later classify newly-acquired sentences. Twitter served as a suitable data source, with over 2,000 tweets obtained on a variety of topics.

### 3.3.2.1 Twitter Review Datasets

This collection was made up of two review datasets in the Arabic language, on topics regarding restaurants and hotels. All reviews in this collection were labelled manually by two Arabic experts. Table 3.1 offers an illustration of the samples of annotated tweets, whose sentiment was established with the use of two raters. These raters agreed on the classification of most tweets, but in cases where there was a disagreement a third rater was used to break the tie, thereby helping to settle on a single final sentiment.

<table>
<thead>
<tr>
<th>Positive</th>
<th>مطعم سيزلر هاوس رائع جداً</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The Sizzler House Restaurant is very cool</td>
</tr>
<tr>
<td>Negative</td>
<td>ابل بيي و فراديزي كل مطعم حالف بصير آسوأ من الثاني</td>
</tr>
<tr>
<td></td>
<td>Applebee’s and Fridays both restaurant’s becomes worse than the second one</td>
</tr>
</tbody>
</table>

Out of the 2,000 tweets obtained, 909 of them related to the topic of restaurants, for the purpose of training the corpus. 459 of the tweets were positive, while the remaining 450 were negative.

---

\(^{11}\)https://developer.twitter.com/en/docs
The tweets selected for this study were ones which held a single opinion, excluding those labelled with sarcasm, or those that were objective in nature. Once the tweets were annotated, they were then subjected to the process of extracting and eliminating all usernames, pictures, URLs, and non-Arabic words. This was done in order to ease further analysis.

### 3.3.2.2 The QAYM Dataset

This dataset contains Qaym (http://www.qaym.com/) website restaurants reviews for three different restaurants. The reviews were crawled from the Qaym website in February 2016. The review documents were organized by product category, evaluated, and subsequently labelled as being either positive or negative, depending on their sentiment polarity. A summary of the statistics for the Qaym dataset has been presented in Table 3.2.

<table>
<thead>
<tr>
<th>Restaurant</th>
<th># of positive Samples</th>
<th># of negative Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>MacDonald</td>
<td>759</td>
<td>394</td>
</tr>
<tr>
<td>Herfy</td>
<td>715</td>
<td>292</td>
</tr>
<tr>
<td>Pizza Hut</td>
<td>341</td>
<td>249</td>
</tr>
<tr>
<td>Total</td>
<td>1815</td>
<td>935</td>
</tr>
</tbody>
</table>

### 3.3.2.3 Trip Advisor Dataset

This dataset contains the Trip Advisor (https://ar.tripadvisor.com) website’s reviews for three different restaurants, crawled in February 2016. Each reviewed document is labelled as being either positive or negative, based on its sentiment polarity. The documents in the dataset are organized by product category. Summary statistics of this collection are shown in Table 3.3.
### Table 3.3: Trip Advisor Dataset Statistics

<table>
<thead>
<tr>
<th>Restaurant</th>
<th># of positive samples</th>
<th># of negative samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lusin</td>
<td>215</td>
<td>21</td>
</tr>
<tr>
<td>Nozomi</td>
<td>89</td>
<td>13</td>
</tr>
<tr>
<td>The Globe</td>
<td>243</td>
<td>46</td>
</tr>
<tr>
<td>Total</td>
<td>547</td>
<td>80</td>
</tr>
</tbody>
</table>

#### 3.3.2.4 Previous Datasets

Previous research studies on sentiment analysis have been conducted in regards to customer reviews, often from the domain of consumer electronics (Hu and Liu, 2004; Popescu and Etzioni, 2005; Ding et al., 2008). The most prevalent dataset for Arabic tweets is the Twitter Dataset for Arabic Sentiment Analysis (TDASA), prepared by Abdulla and Mahyoub (2014) using a tweet crawler. 2,000 labelled tweets were collected, half positive and half negative, on a variety of topics ranging from politics to the arts. Some of the tweets were written in Modern Standard Arabic (MSA), while others were in the Jordanian dialect. The labelled tweets expressed positive or negative opinions, which were annotated manually by two human experts. These experts agreed on the classification of most tweets, but in the cases where there were differing opinions, a third expert was called upon to break the tie.

#### 3.4 Phase B: Opinion Corpuses Construction

The language resource construction module has a series of steps, that start with the collection of data (e.g., customer reviews) from the Twitter website. The next step in the module is the simple text pre-processing of the collected data, for the preparation of the content which is then saved in a text file.

The main objective of this stage is the preparation of two opinion corpuses for the selected restaurants in KSA, from the Twitter and Qaym (http://www.qaym.com/) and Trip Advisor (https://ar.tripadvisor.com) review sites. The Twitter corpus contains short sentences due to the fact that tweet have a limited size (only 140 characters).
while the second review corpus sentences are long.

3.4.1 Twitter Corpus

The following are activities conducted to prepare the opinion corpus:

- Crawling tweets from Twitter, using the Twitter API.
- About 2,000 multi-dialectal Arabic tweets are randomly retrieved.
- About 909 of them are annotated.
- Each row in MS-Excel is used to represent a single tweet.
- The tweets are separated into a single review, and saved in one file.
- Some of the repeated letters in Arabic scripts are cancelled.
- The wrongly-spelt words are corrected.

In this corpus the sentences were kept short, at a maximum of 140 characters, for each, because of the length restrictions on tweets.

3.4.2 Review Corpus

The second corpus gathers data on restaurant reviews using information available publicly on the Qaym (http://www.qaym.com/) and Trip Advisor (https://ar.tripadvisor.com) websites. Qaym is a website for Arabic user reviews, regarding restaurants, cafes and bakeries all over the world. Originally, the website started off as a site for reviews of restaurants in Arabic countries. The researchers selected Qaym and Trip Advisor, because they have a greater number of customer reviews in the Arabic language, when compared to other websites that translate content to Arabic. The data collected for this study covered 6 restaurants in KSA, with 3,377 customer reviews. The entries for every restaurant listed on Qaym and Trip Advisor contained general information about the restaurants, including meals, locations, pictures and reviews from previous customers. The review for the corpus had a detailed description of the restaurants’ services, which sometimes included some Arabic Dialects. After collecting the data:
A single review was extracted after stripping out the HTML tags, as well as non-textual content.

The content was saved into a single file (text file).

For Arabic scripts, the repeated letters/words were eliminated, particularly in instances where repetition was used to place emphasis.

Misspelt words were corrected.

A label was assigned according to the following rating system, specifically positive for a rating $\geq$ six, and negative for a rating $<$ six. This resulted in about 3,377 customer reviews, 2,362 of which were positive, while the remaining 1,015 were negative.

The sentences were long in this corpus, as some of them contained some Arabic dialects. Table 3.4 shows some samples of positive and negative reviews.

3.4.3 Data Pre-processing

The tweets underwent pre-processing, prior to the extraction of feature vectors. The pre-processing entails tokenization, or identifying the individual words and reducing the typographical variation. This was followed by the application of the proposed pre-processing mechanism, specifically normalization, stemming, and removal of stop words, on the cleaned tweets. Tokenization can easily be undertaken by using one of the functions of the Rapidminer$^{12}$ tool. Once each process in the pre-processing mechanism is applied accumulatively onto the tweets, the result is stemmed tweets from which stop words have been eliminated.

---

$^{12}$ https://rapidminer.com/
Table 3.4: Samples of Positive and Negative Reviews

<table>
<thead>
<tr>
<th>Orientation</th>
<th>Arabic Tweet</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>أفضل وجهة أكلها في حياتي من ماكدونالدز ، اللحمية مشوية بطريقة ممتازة ، بس اللي مو حلو فيه إنه لفترة محدودة ، وبعد التركي راح يجي الصيني ، أنسحبح أكل تجربه ورح يعجبكم.</td>
<td>The best meal I have ever had in McDonald’s, the meat grilled in an excellent way, but the bad thing it is for a limited time only, and after the Turkish, Chinese will come, I recommend you try it and you will like it.</td>
</tr>
<tr>
<td>Negative</td>
<td>بصراحه مطعم سيء وعالي ان شخصي . نفعت ١٠٠ ريال حق العشاء وطلعت تعريمتا بما كانتة لأن أكلهم سيء جدأ .. وطبعا جو المطعم حلو والمكان خخم بس السحح الناس تطلب من عدهم خلا فقط والطعم مسيئ نظام للدخول بالزي الرسمي أو بدله رسمي ومنعوا شخصي امامي من الدخول لأنبي شورت ونا داخل المطعم وجدت أكثر من واحد لابسين شورت .</td>
<td>Honestly a bad and expensive restaurant. Personally, I paid 600 riyals for dinner and leave to eat in a second place because their food was very bad. Of course, the atmosphere of the restaurant is sweet and the place is luxurious but I advise people to take sweets only, and the restaurant is allowed only a formal dress entry or formal suit. They deny one person to enter because he wore shorts and I was inside the restaurant and found more than one wore shorts.</td>
</tr>
</tbody>
</table>

3.4.3.1 Stop Words Removal

Given that there is no specified stop words list for Arabic, the researchers adopted a custom stop word list, as in Appendix B, which are based on the type of application being implemented. Some researchers have constructed lists made up of common and short function words, such as ‘كَٖ’ (in), ‘ٖٓ’ (of), ‘٠٠٦’ (on), and others. Other authors construct lists made up of the most common words, including lexical words, such as ‘ٞ٦’ (like), ‘٦٦٠’ (want), ‘يقول’ (say), and others. In the event that no stop word list exists for restaurant MSA, one needs to be constructed from scratch. The stop word list construction process requires the identification of words from the whole corpus of 2,000 tweets, which are spread over different frequency ranges.

According to Zipf’s law (Li, 1992), there is an inverse relationship between frequency and word numbers. The process of constructing a custom stop word list began with the selection of a set of 10 words with the highest frequencies. This set was

13 http://www.ranks.nl/resources/stopwords.html
generated after removing all sentiment words, such as ‘ráamú’ (fabulous), ‘sáy’ (bad), among others. Named entities like ‘ unavoidable’ (Leucine), ‘الرياض’ (Riyadh), and others, were consequently removed. The same action was repeated for verbs which included ‘اكل’ (eat), ‘يشاهد’ (view), and others. Finally, the generated custom stop word list was tested in order to evaluate the accuracy of the classifier, by using it in the review’s preprocessing. In the earlier stage, the proposed list’s performance was slightly low, which could have been the result of the removal of important words or the inclusion of removable stop words. A manual scan was done to detect this kind of anomalies, and to subsequently improve the list by removing important words from it. Furthermore, more lists were accumulated for the entire range of frequencies, with list sizes growing up to 120 words. The obtained list improved the classifier’s accuracy by 1.5 percent. Some of the stop words have been listed in Appendix B.

3.4.4 Evaluation Metrics in Sentiment Classification

The evaluation of the performance of sentiment classification has been undertaken through the application of four metrics, including accuracy, precision, recall, and F1-score (Chen et al., 2011; Wan, 2011; Zhang et al., 2011; Moraes et al., 2013). These indices are computed on the basis of the confusion matrix, as illustrated in Table 3.5. The indices can be computed, with the help of the following equations:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)
\]

\[
\text{Precision (Pos)} = \frac{TP}{TP + FP} \quad (3.2)
\]

\[
\text{Precision (Neg)} = \frac{TN}{TN + FN} \quad (3.3)
\]

\[
\text{Recall (Pos)} = \frac{TP}{TP + FN} \quad (3.4)
\]

\[
\text{Recall (Neg)} = \frac{TN}{TN + FP} \quad (3.5)
\]

\[
F_1 (\text{Pos}) = \frac{2 \times \text{Precision (Pos)} \times \text{Recall (Pos)}}{\text{Precision (Pos)} + \text{Recall (Pos)}} \quad (3.6)
\]

\[
F_1 (\text{Neg}) = \frac{2 \times \text{Precision (Neg)} \times \text{Recall (Neg)}}{\text{Precision (Neg)} + \text{Recall (Neg)}} \quad (3.7)
\]
Table 3.5: The Confusion Matrix

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positives</td>
<td># of true positive instances (TP)</td>
<td># of false negative instances (FN)</td>
</tr>
<tr>
<td></td>
<td>Negatives</td>
<td># of false positive instances (FP)</td>
<td># of true negative instances (TN)</td>
</tr>
</tbody>
</table>

Accuracy is the portion of all truly-predicted instances, within all predicted instances. An accuracy of 100% means that the predicted instances are equal to the actual instances. Precision is the portion of truly predicted instances, within all predicted instances for each class. Recall is the portion of truly predicted instances, within all actual instances for each class. F1 is a harmonic average of precision and recall. Moreover, 10-fold cross-validation has also been used to evaluate the performance of the models proposed in this study.

3.4.4.1 Cross-validation

To generate reliable results during the evaluation process, the cross-validation technique was applied in the study. For this task, the review documents were randomly divided into N groups of equal size. N is the number of folds in cross-validation, and in this study, N=10. During each step of cross-validation, N-1 groups of documents were treated as to be in the learning process. The performance evaluation was based on the remaining group, which acted as an independent test set. The final performance results were based on the average of N iterations.

3.4.4.2 Statistical Test

To establish if there were any significant differences in the accuracy of the proposed models, when compared to the standard baseline models (in this study’s case, the manually annotated corpuses), a statistical test was conducted based on the performance results of the N-fold cross-validation. These statistical measures provide an indicator of whether the proposed model has a more superior performance, when
compared to the baseline methods. The paired t-test is used to evaluate whether differences between the two methods are statistically significant.

3.5 Phase C: Classification Model (Twitter Arabic Sentiment Classifier, TASC)

This phase involved the construction of the Twitter Arabic Sentiment Classifier (TASC), which served the purpose of classifying reviews expressed in Arabic as possessing either positive or negative sentiments. This classification task is referred to as document-level sentiment classification, as it considers the entire document (tweet) to be a basic information unit. The TASC is built using a 3-step methodology, detailed as follows. The first step is where the sentiment words are extracted using sentiment annotated tweets, where occurrences of each sentiment word are counted in both positive and negative tweets, and where a weight of two is assigned to each word depending on its number of occurrences in positive and negative tweets. The second step is where the tweets were pre-processed, by normalizing, stemming and eliminating stop words. The third step is where tweets are classified as being either positive, negative or neutral, determined through the sentiment words found in the tweet. The last step involved building a confusion matrix for the tweets classified as positive, another matrix for the tweets classified as negative, and a final matrix for tweets classified as neutral, in order to measure the classification’s accuracy.

3.6 Phase D: Pattern and Restaurant Ontology Construction

3.6.1 Pattern Construction

In this sub-phase, the feature terms are identified on the basis of noun phrases
with certain patterns, otherwise referred to as the Base Noun Phrase (BNP). This pattern has been used for aspect sentiment classification tasks in English (Khan, 2014). However, to the best knowledge of the researcher, there has been no published work regarding the aspect extraction methods for similar Arabic reviews. In this work, we aim at apply these patterns that have been used for English reviews, for the aspect extraction task, and to evaluate their performance in Arabic reviews. The tweet aspects are restricted to the following patterns: NN, NN NN, JJ NN, NN NN NN, JJ NN NN, and JJ JJ NN. Here, NN and JJ are nouns and adjectives. Another pattern is the definite Base Noun Phrase (dBNP), which restricts the tweet target aspect terms to the definite BNPs, preceded by the definite article ﺣ. A Beginning Definite Base Noun Phrase is a dBNP at the beginning of a sentence, followed by a verb phrase.

3.6.2 Restaurant Ontology Construction

This sub-phase proposes an ingenious methodology for use in opinion mining, applying Semantic Web-guided solutions to further improve the results of traditional natural language processing techniques and sentiment analysis processes, relying on the computationally expensive analysis of linguistic rules. The Semantic Web solution technologies are used to provide descriptions supplementing or replacing the content of web documents, in this case tweets or reviews. Therefore the content may manifest itself as being descriptive data stored in web-accessible databases, or as mark-up within the documents. These machine-readable descriptions enable content managers to add meaning to content, for example to describe the structure of the knowledge the researchers have about that content. The outcome of this phase intends to improve aspect-based opinion mining, by using ontology at the aspect extraction stage. The constructed ontology was subjected to thorough testing within the twitter dataset, using the proposed sentiment analyser model to enhance model accuracy. This model tried to cover more vocabularies from the target domain, specifically restaurants, by using different feature sets extracted from customer reviews. To the best of the researchers’ knowledge, there are no Arabic ontologies published for restaurants.
3.7 Phase E: Analyzer Model (Hybrid Arabic Sentiment Analyzer, HASA)

In this phase, an analyser model referred to as the Hybrid Arabic Sentiment Analyzer (HASA) was developed. The model provided a generic framework, aimed at defining automatic tools dedicated for aspect-based classification. The HASA framework consists of two main modules, the first being used to identify aspects and their orientation, and the second one used for generating a summary. The steps adopted in this phase have been described in the following section.

3.7.1 Preprocessing Step

In this first step, the vector representations of the terms in the textual format have been obtained through using term occurrences. This has been carried out in text which has already undergone the different pre-processing techniques, previously mentioned in Section 3.4.3. Afterwards, association rule mining is used to discover correlations among the sets of items in the database.

3.7.2 Extract Frequent Word

The second step intends to find frequent words that are the most popular in a text (nouns and adjectives), and then extracting them. This is done using association rule mining, which finds all frequent item sets (Agrawal and Srikant 1994). The generated frequent item sets, which are also called candidate frequent words, are stored as two sets, specifically frequent adjectives and frequent nouns or noun phrases, for purposes of further processing. The frequent adjectives are then used to identify subjectivity and frequent noun or noun phrases are used for identifying aspects.
3.7.3 Creation of an Arabic Sentiment Lexicon

The third step involves the identification of sentiment words, which are words conveying subjective opinions which may be either in the form of positive or negative opinions or sentiments. These are useful in sentiment analysis. Subjectivity is positively correlated with the use of adjectives.

3.7.3.1 Building a Large Scale Arabic Sentiment Lexicon (ArSenL)

It is preferable to use a large-scale Arabic sentiment lexicon, given that the majority of opinion mining approaches use opinion lexicons to determine the sentiment of words, for instance English SentiWordnet (ESWN) (Baccianella et. al 2010) and MPQA Lexicon (Mihalcea et. al 2007). Despite the recent emergence of Arabic opinion lexicons, there is still a shortage of large-scale Arabic opinion lexicons (Abdul-Mageed et. al 2011). To address this problem, this research proposes the development and refinement of a large-scale Arabic Sentiment Lexicon (ArSenL), comprised of four opinion lexicons holding two lists, of positive polarity words and negative polarity words. The four lexicons utilized in the study were:

- The MPQA lexicon, containing 8,000 manually-annotated English words, identified as being positive, negative or neutral (Mihalcea et. al 2007). The words were derived from translations by Mourad and Darwish (2013), who used the Bing online MT system to translate the MPQA lexicon to Arabic. This lexicon was then refined in order to remove the large number of translation errors present, after which only the positive and negative words were extracted for use.
- The ArabSenti lexicon (Abdul-Mageed et. al 2011), which contains 3,982 adjectives obtained from news data, and then labelled as being positive, negative or neutral. For the purposes of this study, only the positive and negative words were extracted for use.
• Shoukry and Amira Magdy (2013) lists sentiment words, containing 652 Arabic words tagged either as being either positive or negative.

• The Arabic SSL (Mahyoub et. al, 2014) contains 885 positive, 616 negative and 6075 neutral unigrams. The created lexicon is described as being context-independent. Only positive and negative words were extracted for the purpose of this study.

Once the four lexicons were refined, with duplicated words eliminated, a final set of the lexicon was produced. This set contained 5,470 Arabic words, 2,125 of which were positive, with the remaining 3,345 being negative.

3.7.3.2 Dialect Lexicon

A data-driven approach was used to construct the dialect lexicon, beginning with the tokenizing of 2,000 tweets into words, from which the most frequent adjectives in the restaurant corpus were extracted. The extracted adjectives were then merged and manually classified by two annotators, as being either negative or positive words. Each word type was saved into a separate file. Table 3.6 displays some examples of the dialect words, and their corresponding MSA words. Table 3.7 displays some examples of the dialect words, and their orientations.

3.7.4 Design of Sentiment Mining Modules

The fourth and last step involved designing sentiment mining modules, used to analyse Arabic opinions written in colloquial Arabic or Modern Standard Arabic (MSA), or both. In this study, the researcher set out to develop modules which handled standard and colloquial Arabic. Previous research by Rushdi-Saleh et al. (2011) and El-Halees (2011) exclusively dealt with MSA, while research by Almas and Ahmad (2006) dealt with MSA Arabic financial terms. Therefore, the modules designed in this study are expected to be more comprehensive.
Table 3.6: Dialect Words and their Corresponding MSA Words

<table>
<thead>
<tr>
<th>Dialect Words</th>
<th>Corresponding MSA</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>طَلْشان</td>
<td>من أجل</td>
<td>In order to</td>
</tr>
<tr>
<td>بِدُها</td>
<td>تُريد</td>
<td>Want</td>
</tr>
<tr>
<td>شُوَى</td>
<td>قليل</td>
<td>Little</td>
</tr>
<tr>
<td>وَاع</td>
<td>مَفْرَف</td>
<td>Nasty</td>
</tr>
<tr>
<td>رَجَحت</td>
<td>ذِهْبَت</td>
<td>Went</td>
</tr>
<tr>
<td>لِيِش</td>
<td>لمَانَا</td>
<td>Why</td>
</tr>
<tr>
<td>مَانَا</td>
<td>شَوَى وَشَوَى</td>
<td>What</td>
</tr>
<tr>
<td>بِس</td>
<td>لُكْنَ/فِظ</td>
<td>But/only</td>
</tr>
<tr>
<td>كَمَك</td>
<td>طَرْف رَمْوَك</td>
<td>Sleeve</td>
</tr>
<tr>
<td>لَسَه</td>
<td>إِلَى الأنَّ</td>
<td>Till know</td>
</tr>
<tr>
<td>هَبَع</td>
<td>نَوَى</td>
<td>Now</td>
</tr>
<tr>
<td>صَيْح</td>
<td>صَدِق</td>
<td>Truth</td>
</tr>
<tr>
<td>ذِقَه</td>
<td>نَثْقَة</td>
<td>Taste it</td>
</tr>
<tr>
<td>مَرْه</td>
<td>جَدَا</td>
<td>Very</td>
</tr>
<tr>
<td>أِنْطَع</td>
<td>إِفْصَل/أَحْسَن</td>
<td>Best</td>
</tr>
<tr>
<td>أَخْتَيَر</td>
<td>أَحْصَيْنَة</td>
<td>Select</td>
</tr>
<tr>
<td>حَسَبَه</td>
<td>لَلَّاسَف/حَسْرَة</td>
<td>Unfortunately</td>
</tr>
<tr>
<td>وَأْد</td>
<td>جَدَا</td>
<td>Very</td>
</tr>
</tbody>
</table>

Table 3.7: Dialect Words and Their Orientation

<table>
<thead>
<tr>
<th>Dialect words</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>وَاع</td>
<td>Negative</td>
</tr>
<tr>
<td>شُوَى</td>
<td>Negative</td>
</tr>
<tr>
<td>فَازْعَه</td>
<td>Positive</td>
</tr>
<tr>
<td>صَيْح</td>
<td>Positive</td>
</tr>
<tr>
<td>أِنْطَع</td>
<td>Positive</td>
</tr>
<tr>
<td>أَخْسَى</td>
<td>Negative</td>
</tr>
<tr>
<td>خَلَو</td>
<td>Positive</td>
</tr>
</tbody>
</table>

Opinions written in MSA or/and colloquial Arabic need to be classified, and their polarity determined. Classification presents a challenge due to the absence of lexical resources in colloquial Arabic. This implies that in order to conduct this study, a colloquial domain-specific Arabic lexicon was built for the restaurant domain. It was constructed using steps listed in the following section.
3.7.4.1 Extract Feature of the Product

Firstly, the frequent noun was extracted from the Arabic opinions reviews in different datasets, after which the extracted frequent noun was used as the product aspects of the objects, representing the product entity. The product aspect is then stored in an aspect set.

3.7.4.2 Opinion Summarizing

A novel information summarizing and visualization approach was developed, using an NLP technique. The process used to formulate the novel approach is as follows:

- The sentiment sentence was defined. It should be noted that these opinion sentences must contain one or more of the product aspects.
- Using the sentiment lexicon, composed of MSA and colloquial lexicons, the sentence orientations were determined as being either positive or negative, in relation to each aspect in the sentences.
- An aggregate of each aspect was obtained.
- Finally, the results were summarized.

The visualization module aimed to provide users with an effective way of browsing through the set of the aspects, according to the polarity expressed by each review.

3.8 Phase F: Results Analysis, Findings and Conclusion

In order to analyse the results, an evaluation framework was introduced into this section. In this study, standard precision, recall, and F1-score were used as measures for evaluating the proposed models’ performance, while accuracy was used as a measure of
the model’s overall performance. An evaluation framework should be introduced, to compare the effectiveness of the proposed models. Within this framework, the effectiveness of the proposed models is evaluated based on evaluation metrics, through comparing them with some baseline methods and other well-known and best-performing methods within the field of sentiment analysis. These comparisons were conducted in order to reveal the advantages and disadvantages of each proposed model, in comparison to other methods. Some baseline methods were also used in the evaluation process, in order to reveal the effectiveness of different parts of the proposed models. Due to the lack of a standard dataset for use in the comparison process, a manually-annotated dataset was constructed and used for this task. Therefore, an evaluation of the proposed models on the extracted datasets served as a demonstration of the actual performance of these models. All baseline methods, and other comparable methods, were implemented and applied to these datasets in order to show the comparative performance of the proposed models under the same conditions.

If standard measures are available, aspect-based opinion mining performance can be gauged in terms of its accuracy, precision and recall. This is not the case, however, in the real data. Some of the previous work uses human judges to examine reviews, to pick out opinion words, and to rate the reviews with a ‘gold standard’. The precision and recall of aspect extraction are then computed against this gold standard. To achieve a similar standard, this study uses two experts in the Arabic language, in order to determine the polarity of sentiments reviews.

3.9 Evaluation Framework

The models proposed in this study were evaluated using existing reputable methods, including baseline methods (TASC), so as to establish their performance in terms of effective classification. In the study, cross-validation was used as an evaluation method while accuracy, precision, recall and the F1-score were used as evaluation metrics.
3.9.1 The Implementation of Proposed Models

In this study, Java programming technology and Rapidminer software were used to implement the proposed models. For the pre-processing steps of each model, a pre-processor program was designed and developed in Java. The program included tokenization, the removal of white spaces, and other necessary text pre-processing tasks. In addition, the tasks of feature extraction, feature selection and feature weighing, were each performed within this program. The output of this program was a classification file, which was used as the input for classification models, based on the classification algorithm.

Rapidminer software was also utilized in the study, for the purpose of implementing each of the proposed models. All semi-supervised approaches in this study were implemented and tested within this software. Several functions and procedures were implemented in Rapidminer, for the learning and evaluation phases of each proposed model. SVM\textsuperscript{light} (http://svmlight.joachims.org/), a well-known implementation of the SVM classifier, was used as the base classifier for all proposed methods. Naïve Bayes and KNN classifiers were also used.

Web Ontology Language (OWL) was used to describe vocabularies in the domain ontology (Restaurant).OWL is a stable specification, developed by the Web Ontology Working Group. It is considered to be a web standard for the industry and academy.

3.10 Summary

In summary, this chapter provides a description of the research methodology used in this study, including all methods and datasets used to achieve the study objectives. The process followed to achieve this study’s objectives was described diagrammatically, using the operational framework and research process. Different research phases were designed for conducting the research and were consequently
explained in detail. The relationships between the research problems and the proposed solutions were described. The process used to construct the data sets utilized in the evaluation stages was also outlined, and the original datasets used in the construction process were introduced. The different types of text pre-processing, as applied to the Arabic tweets, were also explained. This section also addressed the design and development steps involved in constructing the opinion lexicon, later used in HASA. Two lexicons are created, specifically ArSenL and the dialect lexicon. These lexicons are made up of dialect words and corresponding MSA words. Finally, an evaluation framework is designed to demonstrate how the proposed models would be evaluated in this study. A general overview of the research methodology has been summarized in Table 3.8.

Table 3.8: Review of the Research Methodology

<table>
<thead>
<tr>
<th>Phase</th>
<th>Activity</th>
<th>Deliverables</th>
<th>Objective</th>
<th>Novelty</th>
</tr>
</thead>
</table>
| Phase A: Previous research studies | 1. Study previous research  
2. Identify problems | 1. Investigating and proposing different techniques for sentiment analysis, as applied to customer reviews written in the Arabic Language.  
2. Systematic literature review.  
3. The most frequently used aspect-based techniques.  
4. The most frequently used construct opinion lexicon techniques.  
5. The pros and cons of these techniques. Most research on Arabic opinion mining is listed. | 1. Problem formulation.  
2. Justification of the need for automatic tools used for sentiment analysis.  
3. To build a background of aspect-based opinion mining.  
4. Implement methods for creating opinion lexicons.  
5. Identify existing research in Arabic opinion mining. | -               |
| Phase B: Opinion corpuses construction | 1. Study the twitter APIs.  
2. Collect the relevant tweet data using | 1. Prepare the two corpus (corpus of Arabic tweets, and corpus of Arabic reviews about | Preparing the dataset for experiments. | The release of the Arabic tweets corpus. The release |
| Phase C: Classification model construction | twitter APIs.  
2. Assign labels to each review. | of the Arabic restaurant reviews corpus. |
|------------------------------------------|-------------------------------------------------|---------------------------------|---------------------------------|
| **Phase C:** Classification model construction | 1. Pre-process tweets (normalizing, stemming and eliminating stop words).  
2. Extract sentiment annotated tweets, based on the occurrences of each sentiment word.  
3. Assign weight to each word, depending on its number of occurrences. | 1. Classify reviews as being either positive or negative.  
2. Build a confusion matrix. | Choose the best classifier. |
| **Pre-process corpus** | 1. Tweet pre-processing.  
2. Feature extraction.  
- Manual cleaning, to remove spelling mistakes.  
- Tokenizing.  
- Removing stop words.  
- Parts-of-Speech tagging(POS)  
- Stemming.  
- Creating word vector. | Method for identifying subjective tweets. | - |
| **Extract frequent words** | - Applying association rule mining and NLP techniques to low corpora.  
- Use adjectives, and nouns or noun phrases | - Adjectives, etc. | - |
| **Create large scale opinion lexicon** | - Use adjectives set as opinion words.  
- Use review corpus and existent | - Large scale sentiment lexicon and colloquial lexicon. | Construct opinion lexicons. |
<table>
<thead>
<tr>
<th>Phase</th>
<th>Activity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase D: Ontology construction</td>
<td>Extract product aspects</td>
<td>Use the noun or noun phrase set.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Two aspect sets.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extract object aspects, which represent the product’s entity.</td>
</tr>
<tr>
<td>Phase E: Opinion summarizing</td>
<td>Define sentiment sentence.</td>
<td>Construction of the ontology.</td>
</tr>
<tr>
<td></td>
<td>Identify the orientation.</td>
<td>Restaurant ontology.</td>
</tr>
<tr>
<td></td>
<td>Aggregate each feature.</td>
<td>Use of ontology to enhance accuracy.</td>
</tr>
<tr>
<td></td>
<td>Summarize the results.</td>
<td>Release of restaurant ontology.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Visualization module.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aimed at providing users with effective ways of browsing through the sets of features, according to polarity.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The use of patterns and ontology to identify aspects.</td>
</tr>
</tbody>
</table>
CHAPTER FOUR

4 TWITTER ARABIC SENTIMENT CLASSIFIER

(TASC)

4.1 Introduction

This chapter describes a document-level sentiment classifier, which classifies Arabic reviews as containing positive or negative sentiments. The classification is determined based on the entire document. Previous studies have mostly focused on the classification of online reviews.

Given the opinion document ‘D’, which contains an evaluation of a certain entity, identifying the overall opinion holder’s sentiments ‘S’, concerning that entity, for instance the aspect A in the quintuple(e, A, s, h, t ), where the entity ‘e’, opinion holder ‘h’, and time of opinion ‘t’ are known. A classification problem is defined, depending on whether the formulations are based on their categorical values - for instance, by being positive or negative.

Therefore, in order to ascertain that the classification is meaningful, the implicit assumptions made is that a single document ‘d’ contains an opinion which relates to a single entity ‘e’, given by the one opinion holder ‘h’ (Liu,2010). If a document has opinions relating to multiple entities, or has opinions posted by more than one opinion holder, then it is difficult to assign one sentiment orientation to the whole document. This assumption applies for reviews of both products and services. However, the assumption does not hold for forums and blogs, given that these sources have multiple entities.

This chapter aims to construct a Twitter Arabic Sentiment Classifier (TASC) at a
document level, using review datasets created in Chapter 3. The rest of this chapter contains a brief description of the classification algorithms, in Section 2, followed by a description of the experimental setup in Section 3. This is followed by a description of the validation methods and feature selection techniques in Section 4 and 5, an evaluation of sentiment classification and the experimental results are discussed in Section 6 and 7, and a discussion and conclusion in Section 8 and 9.

4.2 Classification Method

The section contains an evaluation of three popular supervised classifiers, specifically the Support Vector Machines, Naïve Bayes, and KNN. The construction of a classification model requires a set of labelled data, along with a target class. If a model contains two target classes, it is referred to as being a binary classification problem. Otherwise, it is referred to as a multi-class classification problem. The construction of the model involves selecting only the relevant aspects pertaining to the target class, which are extracted from sentences as a means of building an aspect vector representation, along with corresponding values.

Each classifier contains the function \( f(x) : \mathbb{R}^d \rightarrow \mathbb{R} \), which is responsible for assigning sentences to their corresponding classes. For example, a binary classifier classifies a sentence as being positive, if the function yields a value which is greater than or equal to zero and is negative if it yields a value that is less than zero.

This chapter gives a description of the theory behind the SVM, Naïve Bayes and KNN classifiers, which were previously mentioned in the literature review, and were used to conduct the experiments.

4.2.1 Support Vector Machines (SVM)

The Support Vector Machines (SVM) serve the purpose of defining decision
boundaries, depending on the concept put forth by the decision maker. Decision planes are put in place to separate object sets that belong to different classes, leading to the formulation of a linear classifier, which is given by the function:

\[ f(x; w, b) = \langle w, x \rangle + b \]  

4.1

Here \( w \) and \( b \) are the function parameters, and the \('\langle', '>\)' signs are the inner product of the two vectors. The linear classifier identifies the most suitable hyperplane, separating the positive and negative class data points with the largest margin from the hyperplane (Fradkin and Muchnik, 2006). The points along the margin are referred to as support vectors. The training data used in the classifier needs to be linearly separable, so that:

\[ y_i f(x_i) > 0, \forall i = 1, \ldots, l \]  

4.2

This implies that the hyperplanes may be selected in a way that no data points exist between the two hyperplanes (Fradkin & Muchnik, 2006). A larger margin ensures better performance in terms of generalizing test data, particularly because the solution is arrived at based on points that lie on the decision boundary.

The benefits of SVM include good performance and reduced dependence on the dataset’s dimensions. However, its drawback is the need for pre-processing in instances where the data is incomplete, posing a challenge regarding the interpretation of the resulting model.

4.2.2 Naive Bayes (NB)

Naïve Bayes theorem is more suited for high dimension inputs, which operate on the assumption that attributes are conditionally independent of the class they belong to. This independence makes it possible to effectively use a small training data set to compute probabilities, using the Bayes formula, by applying conditional probability. Finally, the posterior probability, which is the product of two probabilities, is used to arrive at the final classification.
The prior probability, set before incorporating the observations from the experimental data, is unconditional and established based on prior experience. The likelihood probability is determined after the prior probability. The prior probability for each class is computed as follows:

\[ P(c = i) = \frac{\text{#data points in class } i}{\text{Total Number of Points}} \]  \hspace{1cm} (4.3)

Given the clustering of data points in a class, in the vicinity of X, it can be reasonably concluded that the new cases belong to that particular class.

The probability can be measured by drawing a circle around X, whose arc covers points chosen as priorities, independent of their class. The quantity of points within the circle, which fall into a specific class, are computed as follows:

\[ P(X|C = i) = \frac{\text{of } i \text{ in vicinity of } X}{\text{Total # of } i \text{ cases}} \]  \hspace{1cm} (4.4)

---

14http://www.statsoft.com/textbook/support-vector-machines
Figure 4.2 shows the training data for the classifier, as well as the objects to be classified. Once the prior probability has been established, the resulting classification is based on the posterior probability, which is computed using Baye’s rule. Baye’s rule is presented in Equation 4.5.

\[
\int_{i}^{NB}(x) = \prod_{j=1}^{n} P(X_j = x_j | C=i)P(C=i)
\]

The posterior probability is computed for each class, and the resulting class is selected as the one with the highest probability.

Research by Xia & Zong (2010) asserts that unigrams exhibit better performance in SVM, while high-order N-grams, along with dependency relations, show much better performance in NB. The difference is attributed to the nature of the individual algorithms. Here the SVM is a discriminative model, while NB is a generative model. The SVM is best suited for identifying complex features and their level of independence, which is especially common in unigrams, while NB is suitable for identifying assumptions pertaining to the independent features which are characteristic of bigrams.
as well as dependency relations (Xia & Zong, 2010).

Naïve Bayes is a probabilistic model, which operates on the assumption of the independence of attributes. Therefore, in the case of ‘N’ documents, each of which is represented by the sequence \( d_j = \{t_1, t_2, \ldots, t_T\} \) comprising ‘T’ terms, the probability that a document ‘dj’ falls in the class ‘ck’ is given by:

\[
p(c_k|d_j) = p(c_k) \prod_{i=1}^{T} p(t_i|c_k)
\]

In the above, \( p(t_i|c_k) \) is the conditional probability that \( t_i \) is contained in a document belonging to the class \( c_k \), while \( p(c_k) \) is the prior probability of a document falling in class ‘ck’. \( p(t_i|c_k) \) and \( p(c_k) \) can be estimated from the training data (Russell, 2009).

The Naïve Bayes classification method has two advantages, the first being that it is easy to interpret, and the second is that it represents a more efficient computation. Its disadvantage is the underlying assumption of the independence of attributes, which does not always apply.

### 4.2.3 K-Nearest Neighbour (KNN)

The K-Nearest Neighbour (KNN) is an unsupervised machine learning algorithm, whereby the classification of an object relies on the class in which the majority of its neighbours belong to, for instance the ‘k’ nearest neighbours. The classification depends on the similarity of the objects classified, with instances in the training data. The majority measure is determined by voting, or by distance-weighted voting.

For example, the classification of an object (shown by the red circle) among some known examples (shown by positive and negative symbols), as based on the number of its nearest neighbours, is presented in Figure 4.3. The KNN classifier can be
KNN requires the use of a large size of training data. If you consider the vector A and the set of M-labelled instances, which is \( \{a_i, b_i\}_{1M} \), a KNN classifier assigns the class label of ‘A’ by first establishing the ‘K’ nearest neighbours of ‘A’, and then applies the majority vote of Russell (2009). The classifier uses the Euclidean distance as the distance metric, which is given by:

\[
\text{Dist} = \sqrt{\sum_{i=1}^{D} (X_i - Y_i)}
\]

Figure 4.3: KNN classifier

4.3 Experimental Setup

The approach proposed in this study was evaluated through experimentation, whose design and review corpus has been described in this section. The experiments comprised of the pre-processing stage and the feature selection used several methods and evaluation metrics, which are shown in Figure 4.4.

\[^{16}\text{http://www.statsoft.com/textbook/k-nearest-neighbors}\]
4.3.1 Preprocessing

Once the TDASA data has been obtained by finding the vector representations of the terms in the text format by using the TFIDF (Term Frequency–Inverse Document Frequency) weight, as mentioned in Chapter 3, the pre-processing techniques are all applied. As described in Chapter 3, these techniques include tokenizing, removing stop words, and Arabic light stemming.

4.4 Validation Method

4.4.1 X-Validation

X-validation is a nested operator which is made up of two sub-processes, including training and testing. The training sub-process is applied to the training data set, for the purposes of training the model, after which the testing sub-process is applied. The model’s performance is then evaluated through the course of the testing phase.

First of all, the data is partitioned into ‘k’ subsets, each of which are the same size. One of the subsets is set aside as the testing data set, while the remaining $k - 1$ subsets are set aside for training. The testing and training processes are iterated $k$ times, and the results are given as an aggregate or average estimate. In addition, $k$ is adjustable based on the validation parameter.

4.4.2 Split-validation

Split-validation is a nested operator made up of two sub-processes, specifically training and testing. The training sub-process is applied to the training data set for the purposes of training the model, after which the testing sub-process is applied. The model’s performance is then evaluated in the course of the testing phase.

This is where the data is partitioned into two subsets, whose sizes are adjusted
through use of the parameters. One of the subsets is set aside as the testing data set, while the other is set aside for training. The testing and training processes are carried out in a single iteration, as opposed to the X-validation which is iterated $k$ times.

Figure 4.4: Method of TASC

4.5 Feature Selection Techniques

4.5.1 Wrappers Validation

The accuracy of the predictive learning algorithm is measured more effectively using wrappers, as opposed to filters. Given that the learning algorithm considers every feature set, wrappers are rather expensive, particularly those for large databases which hold multiple features. The combination of learning algorithms with feature selections has resulted in the more prevalent use of filters as opposed to wrappers, which need to
be iterated for each learning algorithm.

4.5.2 Information Gain

The information gain procedure determines the probability of an instance on the segment border, comparing it to the probability of a segment, whereby features have a specific value (Abbasi et. al., 2008). The changes in probability are ranked, and the features with a larger change in probability are considered to be more useful. The ranking is widely used for text categorization applications, particularly in instances where there are large volumes of data being handled, thereby making it difficult to use other attribute selection methods. A reduction of the class entropy sheds more light on the attribute’s class information, in the ‘q’ process, which is referred to as information gain.

4.6 Evaluation of sentiment classification

The performance of sentiment classification is examined through the application of two indices, specifically accuracy and precision. Both of these are computed using a confusion matrix, as previously illustrated in Table 3.4.

4.7 Experimental results

The results of the experimentation, which involve the application of several pre-processing techniques, are discussed in this section. Opinion classification is assessed through use of a comparative evaluation of the review corpus.
4.7.1 Experiment-A: Comparing ML Classifiers without Pre-processing

This experimental setup was used to evaluate the accuracy of the data sets using X-validation with 10 folds. A comparison of the performance of the ML classification methods (SVM, Naïve Bayes, KNN), using the unigram feature, is required.

The training data set consists of 1,000 positive tweets and 1,000 negative tweets without pre-processing, and then the ML classifiers are tested using the 10-fold cross validation method. It is important to note that the performance measures of both the positive and negative classifiers were first calculated using the average of the 10-fold validations, and then these measures were averaged to produce the numbers presented in the tables. The results are shown in Table 4.1.

<table>
<thead>
<tr>
<th></th>
<th>SVM</th>
<th>NB</th>
<th>KNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>83.67%</td>
<td>74.17%</td>
<td>52.33%</td>
</tr>
<tr>
<td>Precision</td>
<td>79.42%</td>
<td>66.37%</td>
<td>96.83%</td>
</tr>
<tr>
<td>Recall</td>
<td>91.03%</td>
<td>98.34%</td>
<td>20.27%</td>
</tr>
<tr>
<td>F-Measure</td>
<td>84.83%</td>
<td>79.25%</td>
<td>33.52%</td>
</tr>
</tbody>
</table>

A comparison of the results from the ML classifiers reveals that the SVM has superior results, in comparison to the other classifiers in all metrics. A comparison of the top two models revealed a 9.5% improvement, in terms of accuracy, precision, recall and the F-measure. Furthermore the SVM has a more superior performance, in terms of sentiment classification, in comparison to other ML classifiers. For this reason the SVM was implemented in sentiment analysis, as a result of its benefits which include its robustness in high dimensional spaces. Any feature is considered to be relevant, while there is robustness when there is a sparse set of samples, and most text categorization problems are linearly separable (Rushdi-Saleh et al., 2011). A comparison of the
machine learning techniques shows that the SVM has a better overall performance, with respect to sentiment analysis.

However, given that 7,189 features were used, constituting a large dimension space, coupled with the noisy nature of the text derived from online reviews, this may have led to the distortion of the feature space. This may have subsequently resulted in a reduced learning rate and the over-fitting of the data, in order to come up with an appropriate model.

### 4.7.2 Experiment-B: The Impact of Pre-processing on ML Classifiers

To test the effect of pre-processing on the three machine learning algorithms, specifically SVM, NB and KNN, this experimental setup was used.

The collected tweets are subjected to three successive stages, including normalizing, stemming, and the removal of stop words. The experiment involved the application of a series of four procedures after normalization, including using bigram, trigram, a light stemmer, and stop words removal. Each stage applied the 10-fold, cross-validation method, giving an average for each class. The results of the experiment are as illustrated in Tables 4.2, 4.3, 4.4 and 4.5.

Tables 4.2, 4.3, 4.4 and 4.5 show the cumulative results of each process. It is worth noting that performance measures shown for the positive and negative classifiers, represents the average values obtained from 10-fold cross-validations.

**Table 4.2: Accuracy Results of Different Classifiers**

<table>
<thead>
<tr>
<th></th>
<th>SVM</th>
<th>NB</th>
<th>KNN</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bigram</strong></td>
<td>82.00%</td>
<td>84.67%</td>
<td>51.83%</td>
<td>72.83%</td>
</tr>
<tr>
<td><strong>Trigram</strong></td>
<td>74.17%</td>
<td>84.83%</td>
<td>52.33%</td>
<td>70.44%</td>
</tr>
<tr>
<td><strong>Stemming with Bigram</strong></td>
<td>84.67%</td>
<td>83.17%</td>
<td>59.50%</td>
<td>75.78%</td>
</tr>
<tr>
<td><strong>Stop words removal + Stemming with Bigram</strong></td>
<td>88.00%</td>
<td>85.00</td>
<td>69.17%</td>
<td>80.72%</td>
</tr>
<tr>
<td><strong>BOW</strong></td>
<td>83.67%</td>
<td>74.17%</td>
<td>52.33%</td>
<td>70.06%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>82.50%</td>
<td>82.37%</td>
<td>57.03%</td>
<td>73.97%</td>
</tr>
</tbody>
</table>
Figure 4.5: Summary of the Accuracy Results of Different Classifiers

Table 4.3: Precision Results of Different Classifiers

<table>
<thead>
<tr>
<th></th>
<th>SVM</th>
<th>NB</th>
<th>KNN</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigram</td>
<td>75.73%</td>
<td>80.12%</td>
<td>92.86%</td>
<td>82.90%</td>
</tr>
<tr>
<td>Trigram</td>
<td>66.37%</td>
<td>80.35%</td>
<td>100.00%</td>
<td>82.24%</td>
</tr>
<tr>
<td>Stemming with Bigram</td>
<td>79.60%</td>
<td>79.07%</td>
<td>96.77%</td>
<td>85.15%</td>
</tr>
<tr>
<td>Stop words removal+ Stemming with Bigram</td>
<td>87.79%</td>
<td>80.67</td>
<td>94.62%</td>
<td>87.69%</td>
</tr>
<tr>
<td>BOW</td>
<td>79.42%</td>
<td>66.37%</td>
<td>96.83%</td>
<td>80.87%</td>
</tr>
<tr>
<td>Average</td>
<td>77.78%</td>
<td>77.32%</td>
<td>96.22%</td>
<td>83.77%</td>
</tr>
</tbody>
</table>
## Table 4.4: Recall Results of Different Classifiers

<table>
<thead>
<tr>
<th></th>
<th>SVM</th>
<th>NB</th>
<th>KNN</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigram</td>
<td>84.02%</td>
<td>85.80%</td>
<td>8.25%</td>
<td>59.36%</td>
</tr>
<tr>
<td>Trigram</td>
<td>79.25%</td>
<td>85.94%</td>
<td>9.49%</td>
<td>58.23%</td>
</tr>
<tr>
<td>Stemming with Bigram</td>
<td>85.93%</td>
<td>84.34%</td>
<td>33.06%</td>
<td>67.78%</td>
</tr>
<tr>
<td>Stop words removal + Stemming with Bigram</td>
<td>88.08%</td>
<td>86.00%</td>
<td>57.08%</td>
<td>77.05%</td>
</tr>
<tr>
<td>BOW</td>
<td>84.83%</td>
<td>79.25%</td>
<td>33.52%</td>
<td>65.87%</td>
</tr>
<tr>
<td>Average</td>
<td>84.42%</td>
<td>84.27%</td>
<td>28.28%</td>
<td>65.66%</td>
</tr>
</tbody>
</table>

![Bar chart showing recall results](chart.png)

### Figure 4.7: Summary of the Recall Results of Different Classifiers

## Table 4.5: F-Measure Results of the Different Classifiers

<table>
<thead>
<tr>
<th></th>
<th>SVM</th>
<th>NB</th>
<th>KNN</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigram</td>
<td>94.35%</td>
<td>92.36%</td>
<td>4.32%</td>
<td>63.68%</td>
</tr>
<tr>
<td>Trigram</td>
<td>98.34%</td>
<td>92.36%</td>
<td>4.98%</td>
<td>65.23%</td>
</tr>
<tr>
<td>Stemming with Bigram</td>
<td>93.36%</td>
<td>90.37%</td>
<td>19.93%</td>
<td>67.89%</td>
</tr>
<tr>
<td>Stop words removal + Stemming with Bigram</td>
<td>88.37%</td>
<td>92.45</td>
<td>40.86%</td>
<td>73.89%</td>
</tr>
<tr>
<td>BOW</td>
<td>91.03%</td>
<td>98.34%</td>
<td>20.27%</td>
<td>69.88%</td>
</tr>
<tr>
<td>Average</td>
<td>93.09%</td>
<td>93.18%</td>
<td>18.07%</td>
<td>68.11%</td>
</tr>
</tbody>
</table>
4.8 Discussion

A comparison of the results of SVM, NB and KNN, indicate that the results are better when pre-processing is implemented. The percentage improvement in the accuracy of the SVM, before and after processing, is shown to be 4.3% for SVM, 10.8% for NB and 16.8% for KNN (Ibrahim, M. A., Salim, N., 2016). Similar improvements are observed for precision, recall and F-measure. This is attributed to the fact that pre-processing serves the purpose of reducing the noise present in text, thereby eliminating distortions in the feature space. In addition, the number of features has been significantly reduced from 7,189 for unigrams, to 3,220 features after pre-processing. This is attributed to the fact that the application of more pre-processing steps results in the convergence of related features, thereby minimizing the problem of over-fitting, and improving the rate of learning.

An analysis of the results following the removal of stop words, shows that there is an earnest need to investigate the developed stop words list. Table 4.2 show an improvement in performance of 3.35% in SVM and 1.83% for NB, which further confirms that either there are some stop words in the candidate text which still need to
be removed, or that some necessary words were erroneously removed.

Despite the fact that pre-processing significantly reduces the noise and lack of structure in text, the noise still has a negative effect on performance measures.

In spite of the fact that pre-processing greatly reduces noise and lack of structure, for example, once the word “وفي - loyal” is subjected to stemming it is transformed into the particle word “في - in”, which is a stop word. Therefore, it is eliminated from the candidate text during the process of removing stop words, thereby losing the sentiment in the process.

One considers the following tweets:

الشعب السوداني شعب وفي المملكة و للامة الإسلامية جميعا

بصراحه خدمة المطعم في فرع حي الروضة تعبانه لاتوجد خدمات سريعة وفي نقص للعماله امام الزبائن

Both of these tweets contain the word ‘وفي’. In the first instance, ‘وفي’ means ‘loyal’, which is a positive sentiment, whereas in the second instance ‘وفي’ means ‘and in’. Therefore, the second part erroneously increases the intensity of the first feature. Once stemming is carried out, the word becomes ‘في’, which is a stop word, and is subsequently removed. This results in a change of the overall sentiment of the tweet.

From the previous experiments, it can be concluded that the best classifier is SVM, and the best per-processing technique is the use of the light stemmer, after eliminating stop words, and the use of Bigram.

4.9 Conclusion

The classification of sentiments at the document level provides an overview of the opinion on an entity, topic or event, which has some shortcomings in terms of implementation. The majority of its application is for the purpose of acquiring additional knowledge regarding certain aspects, such as what aspects of entities are liked and disliked by customers. Document sentiment classification does not possess the capacity
to perform fine-grained tasks that require in-depth natural language processing. Therefore, document sentiment classification falls short in terms of being able to extract such details. Feature-based level classification is more suitable for obtaining a greater level of detail. This phenomenon will be discussed in the following chapters.
CHAPTER FIVE

5 PATTERN-BASED ASPECT EXTRACTION

5.1 Introduction

Identifying a product review’s aspect is crucial in terms of opinion mining, which includes extracting product entities from unstructured product reviews. A potential customer would find it tedious to read through the large database of reviews, in order to make a purchasing decision, as previously discussed in Chapter 4. The document-level representation is not entirely representative of what customers like or dislike at a granule level. Recent studies have observed that aspects or opinion words depend on nouns or noun phrases. This chapter aims to propose architecture for aspect extraction and selection. This architecture explains how an opinion aspect can be extracted from an input document, using patterns and the association rule.

Despite the large amount of research conducted in regards to opinion identification and the construction of lexicons, very little has been done in terms of opinion aspect identification, which is fundamental in aspect-based OM. A hybrid pattern algorithm has been proposed in this study, which aligns with the research’s key objective of identifying opinion aspects. In this Chapter, the methodology, algorithm and an analysis of the proposed technique are laid out.

The analysis of all aspects of a product, using feedback provided by customers in the form of reviews, is paramount for all parties including potential customers, merchants, and manufacturers. In order to obtain valuable input from these reviews, it is crucial to have an automatic aspect extraction method which is useful for opinion mining and summarization. Aspect extraction and mining helps establish a basis
through which opinion summarization can be carried out (Feldman et al., 2007).

There are several difficulties encountered in aspect extraction. First of all, a system capable of extracting aspects from target documents is also assumed to be capable of identifying accompanying opinion terms and expressions within text. This means that the identification of opinions, at both sentence or document levels, requires the identification of evaluative expressions. Secondly some aspects may not be expressly stated and are therefore only deduced based on semantics, otherwise called implicit aspects. As previously discussed, the extraction of opinions involves several tasks including pre-processing, tokenization, Parts-of-Speech tagging, noise removal, aspect selection, and classification. In cases where the opinion aspects are implicit, context dependency or distribution similarity may be employed. On the other hand, noun phrases with syntactic rules are used for explicit opinion aspects identification (Ferreira et al., 2008; Hu and Liu, 2004; Popescu and Etzioni, 2005; Somprasertsri, 2010; Turney, 2002; Yi et al., 2003).

The extraction of opinion aspects from unstructured text is closely related to entity extraction. However, this only applies if the entity is contained within subjective text. Therefore, named entity recognition is applicable within aspect identification, requiring further processing in order to establish whether the text contains an opinion. Initial studies regarding named entity extraction were conducted by Rau (1991), who proposed heuristics and hand-crafted rules as means for facilitating the extraction of company names from text. For Arabic reviews, Elarnaoty et al. (2012) proposed the use of an opinion holder and subjectivity lexicon. Arabic Named Entity Recognition (ANER), was used to extract names from documents (Liu, 2010).

The rest of this chapter is organized as follows. Section 5.2 offers an overview of aspects identification techniques. Section 5.3 describes the aspects opinion mining challenges. Section 5.4 provides a discussion of the aspect extraction and engineering process. Section 5.5 describes the proposed linguistic patterns of evaluative expressions. Section 5.6 provides an explanation of the hybrid pattern framework. Section 5.7 introduces the patterns that used for aspect candidate selection. Section 5.8 provides a discussion of the steps followed in the proposed algorithm for selecting
candidate features. Section 5.9 provides an explanation of the results achieved in the comparison. Section 5.10 gives a discussion. Lastly, Section 5.11 summarizes the chapter.

5.2 Aspects Identification Techniques

There are two main categories of approaches for aspect classification and opinion target identification, regarding unstructured text, including supervised and unsupervised approaches. Other users have employed semi-supervised approaches as well.

Supervised learning approaches make use of manually-labeled text, as a means of classifying and extracting features. Supervised approaches are characterized by good feature extraction results, which is achieved through the manual preprocessing of training data sets. However, this process is tedious, skill-oriented, requires a considerable amount of time, and is at times domain-dependent. The most commonly-supervised techniques include the Decision Tree, the K-Nearest Neighbor (KNN), the Support Vector Machine (SVM), the Neural Network, and the Naïve Bayes Classifier (Weiss et al., 2010).

Conversely, the unsupervised techniques do not use labeled data, but rather they automatically predict product features, on the basis of patterns in syntax, and on relations in semantics. A considerable amount of research has been conducted regarding this subject (Carenini et al., 2005; Gamgarn and Pattarachai, 2008; Hu and Liu, 2004; Lopez-Fernandez et al., 2009; Nasukawa and Yi, 2003; Popescu and Etzioni, 2005; Somprasertsri, 2010; Toprak et al., 2010; Wei et al., 2010; Zhuang et al., 2006).
5.2.1 The Supervised Learning Approach

The aspects are referred to in opinionated text, which can be retrieved through using supervised learning techniques which are trained and tested with data from reviews. The training data sets are manually annotated for applying in ML. It follows that sentiments and semantic aspects of a particular language are an essential part of supervised machine learning (Jin et al., 2009; Wong and Lam, 2008). Past research has delved into the linguistic features of expressions used for assessment, as well as identifying the subject (opinion aspect) of these assessments. Opinion words that have been explored in previous research, as reported by Liu (2010), include:

i) Bag-of-Words

This refers to a collection of words or phrases, along with their frequency of appearance, without considering the context in which they are used or their syntax with regards to other words contained in that sentence or document. The Term Frequency-Inverse Document Frequency (TF-IDF) models make use of bag-of-words in identifying the subject or target of opinionated text, as well as when extracting the conveyed sentiments (Nigam and Hurst, 2004; Qu et al., 2010).

ii) POS Tags

Text parsing and the categorization of words is a by-product of research in NLP. Parsing involves the categorization of words, according to the parts of speech that they belong to, such as adverbs, verbs, nouns, adjectives, and others. Particular emphasis is placed on adjectives and adverbs, as they help to distinguish and classify sentiments. The nouns give an indicator of the opinion target and sources, with the help of supervised and unsupervised approaches (Hatzivassiloglou and Wiebe, 2000; Liu et al., 2005; Popescu and Etzioni, 2005; Turney, 2002; Yi et al., 2003).
5.2.2 The Unsupervised Learning Approach

Each language has specific grammar rules and particular parts of speech, as well as grammatically-acceptable word sequences. Therefore, syntactic patterns containing word category sequences, dependency grammar, or a contextual and semantic relation, serve as an indicator of the effective machine learning techniques used in distinguishing opinions, classifying sentiments, and identifying targets of expressed opinions. Some unsupervised learning approaches make use of patterns in syntax, and the context in which words are used (Ferreira et al., 2008; Hu and Liu, 2004; Kobayashi et al., 2004; Lu et al., 2011; Toprak et al., 2010; Wei et al., 2010; Yi et al., 2003; Zhai et al., 2011). Unsupervised approaches extract aspects of opinions through two major steps, which include:

1. Candidate Selection

Unsupervised learning approaches make use of patterns, in order to distinguish different candidate aspect features through the use of syntactic relation, with the most common pattern being a noun-based pattern. The two broad patterns used in selecting candidates, used to narrow down the aspect of the expressed opinion, include:

i) The Noun-Based Pattern

The Base Noun Phrase (BNP) is often used to refer to an entity. The BNP is a sequence of nouns or adjectives (JJ), and nouns (NN). Examples are NN, NN NN, JJ NN, NN NNNN, JJ NN NN, JJ JJ NN, where NN and JJ are nouns and adjectives. Since all BNPs in a document cannot be opinion targets, the selection of all BNPs results in lower levels of precision. In order to improve the precision, the existing approaches employed restricted patterns in order to control a false positive outcome. Ferreira et al. (2008), Turney (2002) and Yi et al. (2003) have all proposed two different patterns, as an effort to restrict the patterns of noun phrases, for instance defining the Base Noun Phrase Pattern (dBNP) and the beginning Definite Base Noun Phrases (bBNP). The dBNP pattern considers the noun phrase starting with the article ‘the’ as a candidate opinion target, while the bBNP considers the subset of dBNP to be defined as a dBNP,
followed by a verb. In the previously cited papers, authors compared the results of these two types of patterns. The results showed that although the bBNP’s precision is higher than that of dBNP, the bBNP has very low recall. In addition, dBNP provides balanced precision and recall, thereby resulting in a better F-score. This research has proposed to adapt and use both the dBNP and bBNP patterns.

ii) Dependency Grammar-based Pattern

Syntactic relations have been used to identify the correlation between opinion words and the aspects referred to in the opinions (Fei et al., 2006; Nakagawa et al., 2010; Fei et al., 2004). Previous work by Qiu et al. (2011) have used dependency relations to effectively describe opinion word expansions, and to identify aspects of the conveyed opinion. Table 5.1 represents the dependency relations for nominal sentences in Arabic Language.

The N-grams word model finds a series of consecutive words, which have the ‘n’ length. The more widely used N-grams models are the unigram, bigram and trigram models, with sizes 1, 2 and 3, respectively. Larger model sizes are referred to by their respective value ‘n’, examples being ‘four grams’, ‘five grams’, and others. For example, for the phrase "لوسين اجمل مطاعم المملكة", its bigram will be as follows:

<table>
<thead>
<tr>
<th>Relation</th>
<th>Arabic Name</th>
<th>Dependency</th>
<th>Dependent → Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>adj</td>
<td>صفة</td>
<td>Adjective</td>
<td>Adjective → noun</td>
</tr>
<tr>
<td>poss</td>
<td>مضافة اليم</td>
<td>Possessive construction</td>
<td>Second noun → first noun</td>
</tr>
<tr>
<td>pred</td>
<td>مبتصا و خير</td>
<td>Predicate of a subject</td>
<td>Predicate → subject</td>
</tr>
<tr>
<td>app</td>
<td>بلد</td>
<td>Apposition</td>
<td>Second noun → first noun</td>
</tr>
<tr>
<td>spec</td>
<td>تمييز</td>
<td>Specification</td>
<td>Second noun → first noun</td>
</tr>
<tr>
<td>cpnd</td>
<td>مركب</td>
<td>Compound</td>
<td>Second number → first number</td>
</tr>
</tbody>
</table>
2. **Relevance Scoring**

Relevance scoring falls into one of two categories. The first category contains those reliant on distributional similarity (Alvarez and Lim, 2007; Bollegala et al., 2007; Chen et al., 2006; Ferreira et al., 2008; Hu and Liu, 2004; Sahami and Heilman, 2006; Wei et al., 2010; Yi et al., 2003). The second category are those reliant on pre-existing knowledge resources, for instance thesauruses, ontologies or encyclopedias (Agirre et al., 2009; Alvarez and Lim., 2007; Chen et al., 2010; Lu et al., 2010; Yang et al., 2009; Yang and Powers, 2005).

Current unsupervised approaches for opinion aspect extraction apply the distributional similarity method, which evaluates the frequency distribution of words, in order to determine the relevance of aspect words. Another popular approach is to use the semantic relatedness between terms to identify aspect words. However, this approach has not been specifically applied when determining the comparative relevance of potential opinion aspects.

5.3 **Aspect Opinion Mining Challenges**

Aspect opinion mining faces several challenges, including the use of synonyms, the use of varied sentiments for similar ratings, noisy information, and the use of implicit aspect or sentiment and comparative opinion expressions.

An aspect opinion extraction system uses product or service user reviews as inputs and outputs for a set of relevant aspects. Generally, opinions may be expressed about various subjects, including products, people, companies, events, or topics. The term ‘object’ is used to refer to an entity which contains a set of components, as well as attributes or properties. For instance, a restaurant is an object which has components, including the menu, the bill and catering. In addition, catering contains a set of attributes, which include takeaway, home-delivery, and self-service. Home-delivery has
its own set of attributes, including delivery area, and delivery delay. In describing or criticizing a product, users do not usually mention objects, even though they do describe those objects’ components and attributes. In the context of this research, an aspect is used to refer to components and attributes.

If a specific aspect ‘A’ is only found in an evaluative text ‘T’, it is referred to as an explicit aspect in T. For example, if T is “زدْ حُٞخزش يـ٤ؽ”, translating to “The size of the meal is small”, the explicit aspect A is “size of the meal”. This study only considers explicit aspects.

5.4 Aspect Extraction and Engineering

Aspect engineering is an essential task in data-driven opinion mining, which converts text into an aspect vector (Pang and Lee, 2008). This is referred to as the aspect extraction phase, which identifies the pertinent aspects of opinion mining.

There are several types of aspects used in opinion mining. These include:

5.4.1 Term Presence vs. Frequency

Term frequency is an essential aspect of the traditional approach to information retrieval and text classification. In sentiment analysis, the most important consideration is term presence, as opposed to term frequency, as proven by the study by Pang et al.(2002) for which better results were achieved through using term presence. The concept of term presence is to indicate the appearance of a term using the binary values 0 or 1,0 for present and 1 for not-present. Pang’s study points out the difference between topic-based and polarity text classification. Although the subject of an opinion may be highlighted by the repetition of specific words, or rather their high frequency, it may not necessarily represent the overall sentiment.
5.4.2 Parts-of-Speech

Parts-of-Speech (POS) information is the most commonly-used feature in opinion mining. One of the most significant reasons for using POS is that it can help clarify the meaning of words used in a sentence. Hatzivassiloglou and Wiebe (2000) proved the correlation between adjective presence and sentence subjectivity. Researchers have used adjectives as a feature, as they serve as an indication of sentiment, which is useful for constructing a sentiment lexicon (Whitelaw et. al., 2005; Kamps et. al., 2004; Abdul-Mageed et. al., 2011).

5.4.3 Term Position

The position of words in text determines the weight of words. Words which appear in the title, subtitle or abstract carry more weight than words in the body of the text (Shelke et. al., 2012). In some cases, the presence of a negative word at the end of a sentence made up of positive words, can hold the greatest influence over the overall sentiment of the sentence. An example is "ٛحُٔٓؼٌُْٖرؼٍٝخزخطٚؿخُ٤شٕحُٔٓؼٌُْٖرؼٍٝخزخطٚؿخُ٤شٍ,“ or “This restaurant is wonderful, but there is a high price for some dishes”. The sentiment classification of this example, when using the polarity lexicon only, will be neutral due to the appearance of one positive and one negative adjective. There is in fact a positive sentiment, because the first impression of the writer is “This restaurant is wonderful...”. This result will be achieved through using the word position feature. Hu and Liu (2004) used word position as a feature in opinion mining.

5.4.4 Negation

The interpretation of negation and the part of the sentence that is affected by the negation, is an essential aspect of sentiment analysis. Often times the negation may not be straight forward or contain words that clearly indicate negation, such as "لايس.ما". Research indicates that there are a variety of words which invert the polarity of an
opinion. Such words include disposition shifters, such as "وجدت الطاولة ليست نظيفة", and connectives such as "بالرغم من أن الوجهة لنجدها لكنها صغيرة". A list of negation Arabic words has been included in Appendix C.

5.4.5 Syntactic Dependency Tree Patterns

A syntax relation between a word and its dependents, presented in the form of a tree structure, is referred to as a syntax dependency. Dependency structures are used to identify semantic relationships. Here a leaf node represents a word or phrase, while an edge connects two nodes. Meng (2012) indicated that the relationship between nodes is determined by dependency grammars, with the parent word being the head in the tree structure, while the modifiers are the children. A great deal of research has been conducted in the area, in an effort to find efficient and accurate parsing tree patterns useful for sentiment analysis. Research conducted by Collins (1997) and Nakagawa et al. (2010) applied syntactic dependency trees for the purpose of sentiment analysis. Here the performance was better when compared to the use of Bag-of-Word aspects. Words, phrases or patterns are allocated certain thresholds for treatment as aspects, depending on their frequency. Syntactic dependency tree patterns are structured patterns which may appear rarely in a corpus, particularly in lengthy syntactic patterns.

5.4.6 Opinion and Emotion Words and Phrases

Attributive adjectives such as ‘good’, ‘like’, ‘ugly’, ‘wonderful’, ‘excellent’ and so on, are identified as opinion words that convey certain sentiments. Machine learning approaches build a seed list made up of these words, so as to come up with an opinion lexicon which identifies the target of expressed opinions, by looking at the closest noun phrases. This approach has been applied in the use of both supervised and unsupervised learning approaches (Zhang et al., 2008; Esuli, 2008; Pang and Lee, 2008; Somprasertsri and Lalitrojwong, 2010; Wilson, 2008).
5.4.7 Valence Shifter

A contextual valence shifter serves the purpose of flipping sentiment expressions from negative to positive, and vice versa. For example, in the sentence “The quality of the product is not good”, the valance shifter ‘not’ has changed the polarity of the opinion from positive to negative. Polanyi and Zaenen (2006) conducted an in-depth study on this subject. Later studies by Kennedy and Inkpen (2006) involved a demonstration of the effect of valence shifters on the effective classification of movie reviews. Furthermore, Longton (2008) presented an empirical analysis of lexical polarity and contextual valence shifters, as used in the classification of opinions.

5.4.8 Comparative Terms and Phrases

Some previous studies have considered the use of comparative and superlative words in sentences, as a means of identifying underlying opinions. These previous works presume that comparative and superlative words can provide comparisons between different objects or features. For example, the sentence "product X is better than product Y" presents an opinion by using a comparison between different products. Therefore, the existing work has made use of words that present a comparative opinion, serving to narrow down on the subject of the opinion (the target), and to effectively classify the opinion (Carenini et al., 2005; Feldman et al., 2007; Jindal and Liu, 2006; Xu et al., 2011).

5.5 Selection of Candidate Aspects using Linguistic Patterns

Recent research has been conducted regarding the association rule, as well as frequent itemset mining, resulting in the development of faster algorithms. This thesis has used the fast and scalable Frequent Pattern Tree algorithm, FP-Growth (Agrawal and Srikant, 1994). Association rule mining takes a sentence as a transaction, but
association rule mining is unable to factor in a word sequence, being an essential aspect of natural language texts. Therefore the pre-processing methods are used in order to find patterns for extracting aspects (N-grams). The FP-Growth module was then applied in order to generate the frequent itemsets, specifically nouns or noun phrases that appear in more than 1% (representing minimum support). Table 5.2 represents the candidate frequent aspects, collected and stored in the feature set for further processing. Figure 5.1 explains this method.

<table>
<thead>
<tr>
<th>Support</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.450</td>
<td>المطعم</td>
</tr>
<tr>
<td>0.162</td>
<td>سعر الوجبة</td>
</tr>
<tr>
<td>0.155</td>
<td>الجلسات</td>
</tr>
<tr>
<td>0.150</td>
<td>الخدمة</td>
</tr>
<tr>
<td>0.144</td>
<td>المنيو</td>
</tr>
<tr>
<td>0.141</td>
<td>الخدمات الآخرى</td>
</tr>
</tbody>
</table>

The aspect selection is based entirely on noun phrase patterns, which is the technique applied by this study.

i) **Base Noun Phrases (BNP)**

Candidate features can be identified using word patterns, consist of a combination of the noun (NN) and adjective (JJ). As an example, NN, NN NN, JJ NN, NN NNNN, JJ NN NN, and JJ JJ NN.

ii) **Definite Base Noun Phrase (dBNP)**

dBNPs refer to the noun phrases BNPs, which start with the article ‘Al ꞉’. This word pattern is used on the premise that a great deal of proper nouns begin with the article ‘Al ꞉’, thereby serving as a useful pointer for the entity extraction process.
iii) **Beginning Definite Base Noun Phrases (bBNP)**

The bBNP refers to a sequence of definite noun phrases superseded by verbs. This pattern is applied to the presumption that the noun phrase starting with the article “Al حُرُ,” and a verb, is often identified as a feature.

### 5.5.1 Identification of Aspect Boundaries for Patterns

The study by Yi et al. (2003) used BNPs, dBNPs and bBNPs to identify the subjects of expressed opinions, or product aspects, whereby noun phrases were considered to be representative of candidate aspects. The limitation of this approach is that it does not outline rules pertaining to multiple matches. For example, the pattern “battery life” can be deduced to reflect several features, including "battery life", "battery", and "life". In other words, they refer to the longevity of a battery cell, and longevity in general. A study by Ferreira et al. (2008) extended the work of Yi et al. (2003), proposing an algorithm which picks out the longest BNP patterns. For instance, in the previous example, this rule would select "battery life" as the feature.

### 5.5.2 Issues and Limitations of Existing Approaches

Issues arising from the use of unsupervised approaches of patterns selection and relevance scoring, are mostly the result of the use of base noun phrases as a pointer to the aspect of an opinion. Given that not all noun phrases hold an opinion aspect, recent studies have looked into the identification of dependency patterns for aspects identification. A review document may contain sentences conveying an opinion(s) through BNPs, but others may not. For instance, the sentences “Tao’s restaurant in Riyadh offers families breakfast and others.”, "The Cavalli Café Restaurant on the Boulevard Raval Tower has a beautiful..." do not contain opinion aspects, despite the fact that they have a base noun phrase. The sentences “The Cavalli Café Restaurant on the Boulevard Raval Tower has a beautiful..."
“seating area”, “eating area”, hold a single opinion aspect being’s eating area’, ‘جنساته’, despite having six different BNPs. It can therefore be concluded that the casual selection of BNPs results in a large false positive ratio.

Several strategies have been proposed to counter this issue, one of them being the association mining approach which works on the premise that opinion aspects are regularly mentioned in reviews. The disadvantage of this approach is that there may be words which appear frequently, but are not opinion aspects, while other words may be infrequent but will hold an opinion aspect. Another solution strategy is the use of a set of pruning rules. Pruning, however, still has the potential for improvement, including the use of pruning heuristics that cover most positive examples and exclude most negative examples of the training set. This would reduce the significant noise problem.

5.5.3 FP-growth Algorithm

Let I =\{a_1, a_2, \ldots , a_m\} be a set of items, and DB=T_1, T_2, \ldots , T_n, be a transaction database in which T_i (i ∈ [1 \ldots n]) is a transaction that contains a set of items in I. The support of a pattern A, with A being a set of items, refers to the number of transactions having A in the database. Pattern A is frequent in cases where A’s support is equal to or greater than a predefined minimum support threshold, or mnsup. The frequent pattern mining problem describes the problem of identifying the complete set of frequent patterns, when the DB and minimum support threshold have been provided.

The FP-Growth algorithm facilitates the identification of frequent item sets, without the need for candidate generation, and an additional two-step process. The first step of the process involves building a compact data structure, known as an FP-tree for item sets, which are derived from transactions that meet a user-specified minimum support. The second step is extracting frequent item sets directly from the FP-tree. In addition, only frequent item sets with a maximum of four words are considered, since a
product feature has three words at most.

5.6 Hybrid Pattern Framework

A hybrid pattern framework is comprised of a combination of linguistic patterns, made up of base noun phrases and evaluative expressions. These combined patterns are called combined base noun patterns (cBNP), which are terms selected through grammatical categorization. A number of these patterns have been incorporated into existing candidate feature selection techniques (Khairullah Khan et al., 2014). In this study, hybrid patterns comprised of both existing and novel patterns, using context and semantic features of the lexical categories of language elements, are used.

A number of factors contribute towards candidate selection, including language elements, context dependency, ambiguity, and term relevancy, with respect to the entity in question. For this reason, text patterns should be accompanied by rules ensuring that the intended objectives are met. In the next section, a discussion of the modelling framework for hybrid patterns has been discussed in sufficient detail.

The objective of the proposed hybrid approach is to improve the performance accuracy of candidate selection, in regards to opinion aspect identification. The hybrid pattern utilized in this approach is used on the premise that position-based patterns, which factor in the context of the text, have the capability of deducing the implied meaning of sentences in their natural language. In addition, studies have shown that opinion dependent contextual patterns give rise to far more accurate results, in terms of the identification of opinion components (Ferreira et al., 2008; Hu and Liu, 2004; Kobayashi et al., 2004; Toprak et al., 2010). Position-based patterns refer to sequences of lexical categories, including Adjective Noun (JJ NN), Noun Noun (NN NN), and others. There are a variety of patterns applicable in aspect extraction, as previously mentioned in Chapter 3. Unsupervised learning approaches classify candidate aspects as being either opinion aspects or non-opinion aspects. The classifier’s accuracy
depends on the patterns used to determine candidate aspects. The more restricted and specific the patterns, the more accurate the identified candidate aspects for the classifier. The validity of the proposed framework is verified through the use of empirical measures provided in Section 5.7.

The proposed hybrid pattern framework consists of two phases, including dataset pre-processing and pattern modelling. In the first phase involving the pre-processing of data sets, in order to facilitate an empirical evaluation, two tasks have to be conducted including dataset selection and data preparation. The review dataset was used in this study, as mentioned in Chapter 3, given that Modern Standard Arabic (MSA) is used instead of dialectical Arabic. The pattern analysis is based on language element categories, as presented in Table 5.3. The documents that make up the dataset are converted into POS-tagged corpuses, using a state-of-the-art language Parts-of-Speech tagger, referred to as the Stanford POS tagger.\(^\text{17}\)

The second phase of the framework entails pattern modelling, using an innovative process which operates through the utilization of language element sequences. General rules are applied when evaluating language features in cycles, applying various combinations of language elements. Each cycle is followed by a comparison of the generated candidate pattern, with the Manually-Labelled Features (MLF) being tagged in benchmark datasets. Eventually, the patterns which achieve high aspect feature ratios are selected.

The generated candidate patterns are useful for providing the candidate opinion aspect, as well as the evaluative expressions. Candidate opinion aspects are subjected to additional refining through the application of scoring techniques, which facilitate the effective identification of opinion aspects. During the course of this study and analysis, the following potential patterns were identified. The patterns singled out through the course of this study has been shown in Table 5.3.

\(^{17}\)http://www.nlp.stanford.edu/software/tagger.shtml
5.7 Patterns for Aspect Candidate selection

This section delves into the proposed linguistic patterns for candidate selection drawn from unstructured reviews. The proposed combined pattern is composed of four different patterns, two of which are inferred from previous existing work, particularly
the definite base noun phrase (dBNP) and the semantic (subjective) base noun phrase (sBNP). Meanwhile the other two patterns are novel, specifically the linking verb-based noun phrase (vBNP), and the preposition based noun phrase (iBNP). The vBNP is comprised of different sub-patterns, as discussed in the next section.

5.7.1 Linking Verb Based Patterns

These are patterns which link verbs connecting subjective adjectives with a base noun phrase. This pattern is based on the observation that, in most of the opinionated sentences, linking verbs link opinion targets with subjective adjectives. For example, the following sentences have the opinion target ‘player’, which is linked with the subjective adjectives through linking verbs. "This player has worked flawlessly. This player is most popular." This pattern shows interesting results, as explained in the subsequent sections. The constituent patterns of vBNP are also explained in the following subsection.

5.7.1.1 Noun Phrase-Verb Phrase-Adjective (NP VBJJ)

This NP VBJJ pattern binds the base noun phrase with a linking verb and a subjective adjective. The example "applications/NNS are/VBP awesome/JJ" contains NNVBJJ, which has "applications/NNs" as the opinion target, interlinked with the linking verb "are/VB", to the subjective adjective "awesome/JJ". As explained in Chapter 2, adjectives provide the best clue to the opinion word in a sentence. The NPVBJJ is based on the assumption that in most cases, opinion aspects are interlinked with subjective adjectives through linking verbs.
Figure 5.1: Method of Aspect Extraction
5.7.1.2 Noun Phrase-Verb Phrase-Adverb Adjective (NP VB RBJJ)

The *NP VB RBJJ* pattern binds the base noun phrase with the linking verb, the adverb, and the subjective adjective. The expression “commands/NNS are/VBP very/RB responsive/JJ” is an illustration of this pattern. In this example, the opinion target “commands/NNS” is interlinked with the subjective adjective “response/JJ”, using “are/VB” and the adverb “very/RB”. This pattern is based on the outcome of this study’s literature review, which shows that in most of cases, adverbs with adjectives provide the best clue to a sentence’s opinion word. Definite Base Noun Phrase (dBNP)

The dBNP was utilized by Ferreira et al. (2008), Turney(2002), and Yi et al.(2003) for extracting opinion aspects, as previously discussed in Chapter 3. This pattern has been adopted in this study, based on those previous studies wherein the patterns were analysed and further categorized into two patterns, specifically dBNP and bBNP. The bBNP is a subset of the dBNP pattern. The bBNP pattern provides greater precision compared to dBNP, but has very low recall. Therefore the dBNP is considered to have a better overall performance.

5.7.2 Preposition Based Noun Phrase (iBNP)

The iBNP pattern depends on the observations that entity-to-entity and entity-to-feature are interlinked through the preposition ‘of/IN’/‘من/في’. For example, in the sentence “المطعم في الخارجية الجلسات“، the opinion aspect is “المطعم في الخارجية الجلسات“. Although this pattern rarely occurs, it has a major impact on study results.

5.7.3 Subjective Base Noun Phrase (sBNP)

The sBNP pattern considers noun phrases to be opinionated expressions that beginning with subjective adjectives. For example, the expression "good/JJ colour/NN setting/NN, funny/JJ pictures/NN" represents the sBNP pattern. This pattern has been
used in a number of previous studies for extracting evaluative expressions (Qiu et al., 2011; Toprak et al., 2010; Zhai et al., 2011). In most cases, this pattern may be the subset of the vBNP pattern. However, in other cases, this pattern represents an opinion and in other times an opinion aspect. An illustration of this instance is shown in the sentence "Some very delicious food was offered to us in this restaurant”, “Some/DT very/RB delicious/JJ food/NN was/VBD offered/VBN to/TO us/PRP in/IN this/DT restaurant/NN”, and “طْ/VBD طوع٣ْ/NN رؼٍ/NOUN_QUANT حُٔخًٞلاص/DT NNS حُِػ٣ػس/DT JJ خعح/NN ُ٘خ/JJ ك٢/IN ٛػح/DT حُٔٓؼْ/DT NNS حُِٓغ/DT NNS اللذىة/DTJJ جدااا/IN انا/INذي/JJ اناا/IN البيت/DT /NN//NN".

5.8 Candidate Aspect Selection

This section provides a description of the candidate selection phase of the proposed architecture. The algorithm used to extract candidate features is based on the four previously-described hybrid patterns. This phase is comprised of three major steps, which include pre-processing, sentence labelling as opinionated or non-opinionated text, the extraction of evaluative expressions, and the extraction of candidate opinion-target pairs. The outcome of this phase is a list of candidate target features, which will be subjected through the next phase, involving a relevance-scoring technique.

5.8.1 Pre-processing

Given that the proposed hybrid patterns are particularly reliant on linguistic features and word positions, the data should be prepared in a proper format. Pre-processing involves two tasks, including Parts-of-Speech (POS) tagging, and sentence splitting.

POS tagging involves the assignment of grammatical categories to words, in accordance with the parts of speech they belong to, such as nouns, adjectives, and so
on. This has been illustrated in Table 5.3 through the use of an example. The input dataset is converted to POS tagged documents, as presented in the illustration in Figure 5.2. This is followed by splitting the text into sentences, after which the individual sentences are subjected to the candidate selection algorithm.

5.8.2 The Candidate Selection Algorithm

The proposed hybrid pattern-based candidate selection algorithm for this study is comprised of two major steps, each detailed in turn over the following section.

**Step 1** involves searching for components of hybrid patterns in the pre-processed input sentence. The search begins by evaluating the presence of the proposed cBNP, at which point it will be labelled as being either opinionated or non-opinionated. The algorithm goes on to check the proposed cBNP constituent patterns in priority, starting with vBNP, dBNP and iBNP, and finally reaching sBNP. In the event that the search yields a match for the vBNP, the search sequence will be terminated and no additional patterns will be searched for, for instance dBNP, iBNP and sBNP. It should be noted that all four patterns are selected based on the opinion’s context. The context is confirmed by the subjective adjectives used in these observed patterns. These subjective adjectives are, in turn, checked using the ArSenL, as previously discussed in Chapter 3. Therefore this first step is composed of the following sub-steps:

a. Searching for patterns.

b. Checking for the presence of opinions in identified patterns.

c. Extracting the opinionated patterns.
A sample of the outcome for the first step is illustrated in Table 5.4.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Example</th>
<th>Opinion Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>vBNP</td>
<td>طاقة/VRT</td>
<td>وسيلة/IN</td>
</tr>
<tr>
<td>dBPN</td>
<td>طاقة/VRT</td>
<td>جيدة/IN</td>
</tr>
<tr>
<td>vBNP</td>
<td>طاقة/VRT</td>
<td>جيدة/IN</td>
</tr>
<tr>
<td>sBNP</td>
<td>طاقة/VRT</td>
<td>جيدة/IN</td>
</tr>
<tr>
<td>iBNP</td>
<td>طاقة/VRT</td>
<td>جيدة/IN</td>
</tr>
</tbody>
</table>

**Step 2** involves generating a list of candidates features from the extracted hybrid patterns. The noun phrases identified in the opinionated patterns extracted in the first step are selected as candidate features, and then the frequency of each distinct noun is calculated. This is done with the use of a candidate selection algorithm. The algorithm goes on to plot a histogram of the appearance frequency of candidate features in the identified patterns. The algorithm applied to the candidate selection phase is shown in Figure 5.3.

### 5.8.3 Grouping Candidate Aspects

Given that multiple people use various words and phrases to describe the same aspect, the grouping of synonyms offers a way of reducing the size of the extracted aspect set. However, the majority of the previous methods do not take feature grouping into account. In some instances, a somewhat synonym grouping approach is used, which may result in many errors during feature set generation. The syntactic role of the group word is also used. The traditional Arabic grammar of *irāb* (إعراب) is used, assigning a syntactic role to each word in a sentence. The relationship between pairs of syntactic units is shown in Table 5.1, with the help of directed binary
dependencies. The table shows the syntactic relation between nouns and other words that define the first word. For instance, "ليه" is a possessive construction of "مضاف ومضاف".

The FP-Growth algorithm works on the assumption that BNPs with dependent subjective adjectives offer a better indication of the opinion target, as opposed to the use of just BNPs. The approach relies on opinionated expressions. The problem arises in identifying opinionated text, as there could be multiple noun phrases accompanied with an adjective in the same sentence. If you consider the following example, the sentence "The/DT picture/NN quality/NN is/VBZ not/RB rich/JJ in/IN colour/NN" contains two candidate BNPs, specifically ‘picture quality’ and ‘colour’, along with the one adjective ‘rich’. Even though ‘colour’ is a feature appearing in a large number of items, in this case ‘picture quality’ is identified as the opinion aspect. Likewise, for the following review, "جميل/NN مكان/NN جميل/NN/NN" is selected as the opinion aspect. This is in line with the dBNP pattern previously discussed in Section 5.5. These patterns do not apply to all instances. Considering the phrase "this DVD player is basically junk" or "الدبل/JJ جمل/NN فوركينا"، ‘player’ is the identified opinion aspect, even though it is not in accordance with dBNP pattern rules. Given that the FP-Growth base approach takes into account frequency distribution, it encounters similar problems in terms of the incorrect selection of opinion aspects, based purely on frequency, and the instances of opinion aspects which have infrequent appearances in opinionated text.

Table 5.5 shows the extracted aspects when using the association rule, without N-grams, with the solution to this problem presented in Table 5.6. Table 5.7 presents a set of aspects with a similar meaning. In order to minimize the size of the set of candidate aspects, the group of words will be "جميل/NN إيطالي/JJ جميل/NN/NN". Table 5.8 illustrates the number of frequent aspects generated for each dataset, in which column 1 lists the dataset names, column 2 lists the number of extracted aspects, column 3 gives the number of actual aspects, and column 4 details the accuracy of the frequent features generated for each product. It can be observed that the tweets dataset’s accuracy is...
relatively low, because the reviews were taken randomly, containing implicit aspects. Also the size of the review is quite small, being limited to only 140 characters. The accuracy of the review dataset is high. Because this type of review is relatively large, it can contain many aspects, and additionally the website provides a description of the main services offered by the restaurant. However, there is the problem that some frequent nouns and noun phrases, such as town names and street names, may not be real product aspects. This impacts the accuracy of the proposed model.

Table 5.5: Aspect Extraction Using Association Rules Without N-grams

<table>
<thead>
<tr>
<th>Size</th>
<th>Support</th>
<th>Item1</th>
<th>Item2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.162</td>
<td>سعر الوجبة</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.6: Aspect Extraction using Association Rules with N-grams

<table>
<thead>
<tr>
<th>Size</th>
<th>Support</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.162</td>
<td>سعر الوجبة</td>
</tr>
</tbody>
</table>

Table 5.7: Set of Aspects with the Same Meaning

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Grouped Aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>وجبة</td>
<td>وجبة</td>
</tr>
<tr>
<td>اطلاق</td>
<td></td>
</tr>
<tr>
<td>أصناف</td>
<td></td>
</tr>
<tr>
<td>اكلات</td>
<td></td>
</tr>
<tr>
<td>الطبخ</td>
<td></td>
</tr>
<tr>
<td>الجلسة</td>
<td>جلسة</td>
</tr>
<tr>
<td>المكان</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.8: Evaluation of the Extracted Aspects

<table>
<thead>
<tr>
<th>Data set</th>
<th>Extract Aspects</th>
<th>Actual Aspects</th>
<th>accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets</td>
<td>71</td>
<td>48</td>
<td>67.6%</td>
</tr>
<tr>
<td>Review</td>
<td>21</td>
<td>18</td>
<td>85.7%</td>
</tr>
</tbody>
</table>
5.9 Experiments and Results

An analysis of the different patterns used for the candidate feature selection of opinion targets has been conducted, with the objective of establishing which patterns are most effective. Given that the patterns are fundamentally based on noun phrases, the analysis commences with base noun phrases. The analysis entails an evaluation of various pattern combinations, in order to establish their comparative accuracy. The benchmark dataset previously mentioned in Chapter 3 was used as the input data for the evaluation. First and foremost, the dataset was converted to a POS tagged document, using the Stanford parser. This was followed by the application of the prototype system based on pattern extraction, using regular expressions in four different setups, with the objective of extracting candidate features.

The first experimental setup was used to extract candidate features, using BNP patterns. As previously explained, the BNP patterns are used for named entity extraction. The purpose of this experiment was to determine the following: if all BNPs were considered to be candidate features, what would be the outcome in terms of precision and recall? Although it is clear that when all BNPs are selected the text will contain all noun-based explicit features, and that its recall will be at 100%, most of the BNPs will not be opinion targets. There will therefore be a high ratio of false positive candidates, which will significantly reduce precision. Since better performance requires both balanced precision and recall, the rudimentary selecting of BNPs is not feasible for identifying candidate opinion targets.

The second experimental setup was for candidate opinion targets extraction, using dBNP patterns. As previously described in related work, dBNPs have been used by Ferreira et al.(2008), Turney(2002) and Yi et al.(2003), for the purpose of extracting candidate opinion targets from unstructured reviews. In this setup, the study analysed the results of candidate opinion targets extraction from opinionated text, using dBNP. This setup used the regular expressions (JJ/DT) *(JJ) *(NN+) *(JJ) * to extract the candidate features.
Algorithm 1: Candidate Selection

Input: POS Tagged Review (R), Opinion Lexicon (ArSenL) Output: Patterns list, Candidate Aspects list

Begin

‘Declare patterns list and Initialize as empty

‘Declare candidate feature list and Initialize as empty

For each Sentence S in R

Patterns = Ø

For each word w in S

‘Extract dBNP

If w is noun or adjective and start with “Al” Then

While w is Adjective or Noun

‘Add the word in the pattern

‘Move the pointer to right

w = Right of w

End While

If Patterns != Ø Then

‘Add pattern in pattern list

‘Reinitialize the Pattern

End If

End If

If w is adjective Then

While right of w is adjective or Noun

‘Add the word in the pattern

‘Move the pointer to right

w = Right of w

End While
If Patterns $\neq \emptyset$ and Pattern contains noun Then

‘Add pattern in pattern list

‘Reinitialize the Pattern

End If

End If

‘Extract iBNP

If $w$ is “of/IN” and left of $w$ is noun or Adverb

Then While right of $w$ is Adjective or Noun

‘Add the word in the pattern

‘Move the pointer to right

$w=$Right of $w$

End While

If Patterns $\neq \emptyset$ and Pattern contains noun Then

‘Add pattern in pattern list

‘Reinitialize the Pattern

End If

End If

‘Extract vBNP

If $w$ is verb and left of $w$ is noun and right of $w$ is opinionated adjective or adverb Then

While left word of $w$ is noun or adjective

‘Add Left of $w$ in pattern

‘Move pointer to the left word

$w=$left of $w$

End While
While Right of w is noun or adjective

‘Add Right word w in pattern
‘Move pointer to the right
word w=Right of w

End While

If Patterns !\=\ø Then

‘Add Pattern in pattern list

‘Reinitialize the Pattern

End If

End If

‘Extract sBNP

If w is adjective and left or right word of w is noun then Then

If w contains in ASL with Positive or Negative Polarity Then

While left word of w is noun

‘Add the word in pattern to the left
‘Move pointer to the leftword
w=left of w

End While

While Right of w is noun or adjective

‘Add the word in pattern to the right
‘Move pointer to the rightword
w=right of w

End While

End If
If Patterns ≠ Ø THEN
   ‘Add Pattern in pattern list
   ‘Reinitialize the Pattern

   End If
End For

End For

‘Generate list of candidate feature
For Each pattern P in pattern list
   For Each word w in P
      If w is noun or (w is adjective and not in opinion list) Then
         While w is noun or adjective
            If w is noun Then
               ‘Add in candidate features
            Else If w is adjective and w not in opinion list Then
               ‘Add in candidate feature
            End If
            If candidate ≠ Ø Then
               ‘Add candidate feature in candidate list C
               ‘Reinitialize Candidate
            End If
            ‘Move pointer to the right
            w = right of w
         End While
      End If
   End For
End For

End

Figure 5.3: Algorithm for Candidate Selection(Khan et al., 2014)
The third experimental setup was for the extraction of candidate opinion targets, using bBNP patterns, which are a subset of dBNP. As previously described in related work, bBNPs have also been tested by Ferreira et al. (2008), Turney (2002) and Yi et al. (2003), in extracting candidate opinion targets from unstructured reviews. In this setup, the researchers tested the results of candidate opinion targets extraction from the opinionated text, using dBNP. This setup used the regular expressions (JJ/DT) * (JJ) * (NN^+) * (JJ) * (VB^+) * to extract candidate features.

The final experimental setup involved a combination of four different patterns, specifically linking verb base noun phrases, definite base noun phrases, preposition-based noun phrases, and subjective adjective base noun phrases. The setup was labelled cBNP and is based on the algorithm shown in Figure 5.3, which employs the regular expressions mentioned earlier in this Section.

In each setup, the results from each pattern were compared with those of the manually tagged features, thereby establishing which of them had a True Positive (TP), False Positive (FP), True Negative (TN) or False Negative (FN). The measures of precision, recall and the F-Score for each experimental setup were computed with the help of the confusion matrix. The next sub-section describes the results obtained from the four setups.

5.9.1 Results

The results of the four experimental setups have been illustrated in Table 5.9, detailing the observed measures of precision, recall and F-Score for each pattern in each dataset, denoted by P, R and F respectively. The results provide a realistic picture of all patterns, in terms of their performance. As indicated by the values in Table 5.9, the recall is greater for BNPs due to higher true positive rate for selected opinionated text. Conversely, the false positive rate is also high. This results in low precision, and in a low F-Score for BNPs.
The bBNP has the best performance in terms of precision, if the lowest false positive rate is considered. However, if the false negative rate is too high, then the recall and F-Score are comparatively low. The true positive rate of dBNP is observed as being much higher than bBNP, while its false negative rate is low. This results in a higher F-Score than that of the bBNP’s. Overall, the cBNP pattern yields comparatively better results, as its true positive rate is higher, even with a low false negative rate.

Table 5.9: The Comparative Results of Pattern-based Candidate Selection

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Pattern</th>
<th>P (%)</th>
<th>R (%)</th>
<th>F (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter corpus</td>
<td>BNP</td>
<td>26.70</td>
<td>88.29</td>
<td>41.00</td>
</tr>
<tr>
<td></td>
<td>dBNP</td>
<td>85.88</td>
<td>67.57</td>
<td>75.63</td>
</tr>
<tr>
<td></td>
<td>bBNP</td>
<td>90.76</td>
<td>43.24</td>
<td>58.58</td>
</tr>
<tr>
<td></td>
<td>cBNP</td>
<td>87.81</td>
<td>81.98</td>
<td>84.79</td>
</tr>
<tr>
<td>Review corpus</td>
<td>BNP</td>
<td>24.91</td>
<td>82.14</td>
<td>38.23</td>
</tr>
<tr>
<td></td>
<td>dBNP</td>
<td>81.99</td>
<td>61.61</td>
<td>70.36</td>
</tr>
<tr>
<td></td>
<td>bBNP</td>
<td>90.01</td>
<td>48.21</td>
<td>62.79</td>
</tr>
<tr>
<td></td>
<td>cBNP</td>
<td>81.50</td>
<td>69.64</td>
<td>75.11</td>
</tr>
<tr>
<td>Average</td>
<td>BNP</td>
<td>25.81</td>
<td>85.22</td>
<td>39.62</td>
</tr>
<tr>
<td></td>
<td>dBNP</td>
<td>83.94</td>
<td>64.59</td>
<td>73.00</td>
</tr>
<tr>
<td></td>
<td>bBNP</td>
<td>90.39</td>
<td>45.73</td>
<td>60.69</td>
</tr>
<tr>
<td></td>
<td>cBNP</td>
<td>84.66</td>
<td>75.81</td>
<td>79.95</td>
</tr>
</tbody>
</table>

The results are made clearer through the graph in Figure 5.4, showing the difference measured against each pattern for twitter corpus, while Figure 5.5 illustrate the average measured for the two corpora.
Figure 5.4: Pattern-based Candidate Selection, for Performance Measures

Figure 5.5: Summary of Average Pattern for the Two Corpora

5.10 Discussion

This section describes the issues and limitations of existing approaches for the extraction of opinion aspects from unstructured reviews. As discussed in the previous
sections, the existing unsupervised approaches exploit linguistic patterns and frequency based-relevance scoring techniques, in order to identify opinion aspects. However, there are certain issues which relate to both patterns selection and relevance scoring, which might affect the technique’s performance. Most of the work considers base noun phrases as opinion aspects. However, all base noun phrases in text cannot be opinion aspects. Therefore the existing research work has been primarily focused on the problem of selecting dependency patterns which are used for aspect identification. For example, some sentences in a review document may not contain opinions, while other sentences may have more than one base noun phrase, with few opinion aspects. Therefore, simply selecting BNPs can lead to a large false positive ratio. The association mining approach is used to overcome this issue. It assumes that opinion aspects are frequently discussed in reviews. However, this approach suffers from two major issues, specifically aspects that are frequent, but not opinion aspects, and aspects that are infrequent, but are opinion aspects. As mentioned earlier, in order to overcome these problems, the research proposed grouping candidates, with the use of traditional Arabic grammar of īrāb (إعراب) as a pruning technique. However, the performance has been improved through the pruning rules.

5.11 Summary

This chapter provides a comprehensive analysis pattern-based candidate selection for opinion aspect extraction from unstructured reviews. Previous studies have shown that various associations of base noun phrases can be applied to feature identification. Given that not all noun phrases can be considered features, certain patterns and rules are applied in order to extract target aspects. In this study, several patterns were proposed as means for restricting noun phrase extraction for the identification of candidate aspects.

This study evaluates the results obtained using the existing patterns associated with base noun phrases. Finally, the hybrid patterns (cBNP) were proposed, made up of
a combination of existing patterns novel features for candidate selection. The proposed patterns are partially based on semantic relations and context dependency. The experimental results indicate an average improvement of 6.95\% in the F-Score, which proves that the proposed hybrid pattern has better performance compared to existing patterns.

Beside the significant improvements in the accuracy of opinion aspect identification, the problem of infrequent aspect identification has not been completely addressed due to dependency on a threshold value. The frequency measure also affects accuracy, as a review can have multiple topics under discussion. Since the existing techniques do not consider the text’s contextual features, they fail to extract implicit features. This is clear from the fact that some semantic features significantly improve the results of association mining. Furthermore, semantic relations and context aware patterns can be potentially employed for implicit opinion target identification.
CHAPTER SIX

6 ONTOLOGY-BASED CANDIDATE FEATURES SELECTION

6.1 Introduction

Ontology is a clear, detailed reference model used in the application domains, with the aim of being a source of knowledge and concepts regarding certain domains, in a form understood by developers and computer systems (Gruber, 1993). An ontology illustrates the domain concepts and the relationships between them, and offers meaningful machine-readable descriptions of the digital content (Guarino, 1995). Ontology is particularly useful for annotating documents with metadata, ensuring improved information consistency, extraction, reasoning and sharing, thereby facilitating application interoperability. Using ontology in opinion mining has several advantages, including the structuring and extraction of features (Pang & Lee, 2008).

This study proposes innovative candidate features selection for aspect extraction, due to its exemplary performance in Information Extraction (IE), and its ability to utilize new Semantic Web-guided solutions with a better performance than the traditional natural language processing and sentiment analysis techniques. The goals to be achieved through using the proposed methodology are firstly to improve aspect-based sentiment analysis through the application of ontologies during the aspect selection stage, and secondly to provide a new vector analysis-based method which is applicable to sentiment analysis. This proposed ontology has been test, and subjected to a series of tests run for real-world restaurant reviews. It has shown exemplary performance, when compared to existing conventional approaches.

The feature-based opinion mining of object reviews is challenging, due to the
varying semantics or meanings of the expressed opinions, as well as the diversity of the features and sub-features used to describe products, and the polarity of opinion words (Hu and Liu, 2004). Recent research has looked into new semantic web technologies, as well as domain-dependent corpora, used for the purpose of feature-based opinion mining (Carenini and Pauls, 2006). Ding and Liu (2007) have asserted that the now-stable semantic web technology could serve as a useful addition to traditional opinion mining. Ontologies are used to represent and structure knowledge in a standard way, on the semantic web. The underlying formal semantics facilitate the automatic information processing and knowledge derivation, using semantic reasons.

In this proposed work, ontology is defined as being the formal and explicit specification of a shared conceptualization, used for structured knowledge representation, and in terms used for a particular field, in this case restaurants (Hu and Liu, 2004). In this study, Web Ontology Language (OWL), which is a W3C standard for ontologies’ representation in the semantic web, represents concepts and features within the restaurant domain. The key contribution to the framework proposed in this study is the classification of Arabic opinions as being either positive or negative, in other words the determining of polarity. In order to form an opinion about an entity, peoples’ points of view about its features have to be extracted. For instance, the features or attributes in restaurants include the menu, and meals such as breakfast, lunch, and others. Research conducted in the past, in regards to this domain has centred on text in the English language. Meanwhile, little research has been conducted in regards to Arabic text. The limited research in Arabic has been attributed to:

- The use of varying forms of the Arabic dialect in forums and blogs, which poses a difficulty in opinion mining, using pre-processing tools constructed for the modern standard Arabic language.
- The difficulty in determining the polarity of text, due to limited datasets and the Arabic dictionary (lexicon), containing words that communicate points of view, or opinions.

The rest of this chapter has been organized in the following way. Section 6.2 offers a description of the document preparation process. Section 6.3 provides a
description of ontology development. Section 6.4 contains an in-depth description of features and opinion aspects extraction. Section 6.5 lays out the results of the conducted experiments, along with a discussion. Finally, a conclusion and summary, are both described in Section 6.6.

### 6.2 Document Preparation

To demonstrate the application of proposed methods, a domain containing some features has been made available on a website with Arabic language reviews. Accordingly a dataset has been selected from restaurant domains, specifically a review dataset from Trip Advisor (2016) and Qaym (2016), containing reviews of restaurants and their features, such as menu، service، meal، and others. A corpus consisting of 2,000 reviews has been collected, as described in Chapter 3. The polarity of the review documents was identified as being an equal number of positive and negative sentiments.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Number of positive reviews</th>
<th>Number of negative reviews</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restaurant</td>
<td>500</td>
<td>920</td>
<td>Qaym(2016)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>80</td>
<td>Trip</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>1,000</td>
<td>Total</td>
</tr>
</tbody>
</table>

### 6.2.1 Pre-processing

Pre-processing is a necessary step in the proposed method. The pre-processing techniques discussed in section 3.3.5 are also conducted when POS tagging is implemented.

Arabic POS tagger, in the Stanford POS-Tagger form, is used to classify words in terms of categories of grammar, being nouns, verbs, pronouns, and others (Stanford
NLP Group, 2013). It uses the rule-based approach, utilizing rules created by linguists specifically for the POS tagging approach (El Hadj, Al-Sughayeir, and Al-Ansari, 2009). For example, “الأمر مطعم هو الربى الذهبي”, meaning “The most beautiful restaurant is the golden shrimp”, is tagged using POS in the following manner:

```
الأمر/JJR مطعم/NN هو/PRP الربى/DTNN الذهبي/DTJJ.
```

6.3 Building the Ontology

Domain ontology refers to concepts, being either a product or an organization, their attributes, their relationships, and the relationship constraints between them. The purpose for ontologies within feature-based opinion mining is feature identification, using common terminologies and definitions of relationships between the terms, within a specific domain (Hatzivassiloglou and McKeown, 1997).

6.3.1 Ontology Development

The use of an ontology is imperative within feature-based opinion mining, particularly for feature identification within the restaurant domain, (for this study). This section details processes undertaken to build the ontology, which could be implemented through either of two approaches. These can include specifically constructing a new ontology, or modifying an existing ontology through extension and adaption, before using it.

In order to build the ontology, ConceptNet (Rob Speer, 2016) was used as a knowledge resource for automatically constructing an independent domain-specific ontology tree for product reviews. ConceptNet relations have an inherent structure, helping construct an ontology tree from the resource. The sample ontology for the ‘restaurant’ domain, using the ConceptNet database demonstrated in Figure 6.1, can be manually translated into Arabic. The next step expands the ontology by merging each node in the ontology with synonyms words in the Arabic Language, using the Arabic WordNet database (Princeton University/global wordnet, 2014). The benefit of using WordNet is that it provides better coverage of domain-specific features within the
The pseudocode used to construct the automatic ontology tree has been illustrated in the algorithm in Figure 6.4. The algorithm uses a recursive function to build the ontology tree. It takes the domain name (root), and a number of levels of the ontology tree, as input parameters. The get_features function uses an SQL query to return a list of features from the ConceptNet database, with a subclass of the root name parameter seen in Table 6.2. The get_synonyms function also uses an SQL query to return Arabic synonyms from the Arabic WordNet database, for the feature parameters. Finally, the function returns an ontology tree for a specified domain.

The construction of the restaurant ontology has been achieved done using Protégé (Knublauch et. al.,2004) with OWL 2. Given that a portion of the concepts and relations did not suit the Arabic restaurant domain, and the fact that there is limited knowledge on the domain, while also considering the ontology hierarchy, this was factored in during refinement.
6.3.2 Ontology Refinement

In this study, the researchers opted to extend and adapt the existing ontology to suit the study’s requirements, due to limited knowledge about ontology construction. The extension was carried out manually, with the help of two annotators, to whom two sets of 450 reviews were assigned. The annotator’s task was to pinpoint the relevant concepts, and generate a list of them, after which the lists were combined and duplicities were eliminated. Once this was done, the concepts were then aligned with the existing ontology. At this point the concepts with similar meanings were classified as synonyms, and were placed into the synonyms dictionary, while one was manually aligned to its corresponding place. Figures 6.2 and 6.3 give a snapshot of the proposed ontology for the restaurant domain, containing 230 classes, 10 of which are top-level classes.

Table 6.2: Concepts and Relations in the ConceptNet Database, for the Restaurant Domain

<table>
<thead>
<tr>
<th>Start</th>
<th>Relation</th>
<th>End</th>
<th>Weigh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restaurant</td>
<td>UsedFor</td>
<td>eat</td>
<td>1.0</td>
</tr>
<tr>
<td>Restaurant</td>
<td>Has A</td>
<td>seats</td>
<td>1.0</td>
</tr>
<tr>
<td>Menu</td>
<td>RelatedTo</td>
<td>restaurant</td>
<td>1.1</td>
</tr>
<tr>
<td>Tasty food</td>
<td>AtLocation</td>
<td>restaurant</td>
<td>1.0</td>
</tr>
<tr>
<td>Restaurant</td>
<td>CapableOf</td>
<td>serve dinner</td>
<td>1.0</td>
</tr>
</tbody>
</table>

6.4 Features Extraction

The constructed ontology from the previous step is then applied to the extraction of product features, which in turn are used to express opinions. To identify the feature term, all noun terms have been extracted from the review. The Stanford Parser tool (Stanford NLP Group, 2013) was used to parse each review, and then assigned a POS tag for each word. The parser tool utilizes the rule-based approach for assigning POS tags.
(El Hadj et al., 2009). An example of a POS-tagged review is “مطعم جدة، مطعم جدة، مطعم جدة، مطعم جدة، مطعم جدة/NN”. Here “مطعم جدة” is a noun term which maybe relevant to features. This is followed by the extraction of noun terms, and the identification of corresponding features in two steps, with the help of the domain ontology.

Figure 6.2: Restaurant Ontology Class Hierarchical

Figure 6.3: Data Property Hierarchical of the Restaurant Ontology
First of all, the concepts relating to the domain are identified with the help of semantic information in the ontology. The relevant features are extracted through a comparison of nouns in the review, the concepts contained in the ontology, and all irrelevant features such as ‘ جدا’, which can be extracted. In the case where the nouns are not present in the ontology, the following step is carried out.

Secondly, the nouns without matches in the ontology are compared with those in the synonyms dictionary. If no match exists, the noun is then disregarded.

An illustration of the identification of features from a review has been provided in Table 6.3.

Table 6.3: Examples of Manual Labelling

<table>
<thead>
<tr>
<th>Arabic Tweet</th>
<th>English Translation</th>
<th>Features</th>
<th>Opinion</th>
<th>Polarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>مطعم سيزرل هاوس يوجد في ارتفاع شارع التحليه اسمه بالله من ارقي وأفضل المطاعم وخدمة راعته وتعامل ممتاز</td>
<td>The Sizzler House Restaurant is located in Riyadh city Tahlia Street, I swear by God it is one of the finest and best restaurants, and the service is Wonderful and has excellent treatment</td>
<td>المطاعم</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>راعية المتغيرة رداسين رقم</td>
<td></td>
<td>مطاعم</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>ارقي وأفضل المطاعم وخدمة راعية ممتاز</td>
<td></td>
<td>راعية</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>المطاعم وخدمة راعية ممتاز</td>
<td></td>
<td>ممتاز</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>في مطعم عدننا اسمه فوزز فاردن وصوله للناس من المدخل راحته وكان سهوره غالي اطلالتة جميلة لكن كلاه خيام الله يديم النعمة</td>
<td>There is a restaurant in our area called Fairouz Qardan, a link to the sky from the praise of his journey and its price is expensive, its views are beautiful, but it is food by the grace of God.</td>
<td>سعر</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>اطلالتة جميلة</td>
<td></td>
<td>خيام</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>مطعم سيزرل هاوس يوجد في ارتفاع شارع التحليه اسمه بالله من ارقي وأفضل المطاعم وخدمة راعية وتعامل ممتاز</td>
<td>The Sizzler House Restaurant is located in Riyadh city Tahlia Street, I swear by God it is one of the finest and best restaurants, and the service is Wonderful and has excellent treatment</td>
<td>/OntologyTree</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>راعية المتغيرة رداسين رقم</td>
<td></td>
<td>مطاعم</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>ارقي وأفضل المطاعم وخدمة راعية ممتاز</td>
<td></td>
<td>راعية</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>المطاعم وخدمة راعية ممتاز</td>
<td></td>
<td>ممتاز</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>في مطعم عدننا اسمه فوزز فاردن وصوله للناس من المدخل راحته وكان سهوره غالي اطلالتة جميلة لكن كلاه خيام الله يديم النعمة</td>
<td>There is a restaurant in our area called Fairouz Qardan, a link to the sky from the praise of his journey and its price is expensive, its views are beautiful, but it is food by the grace of God.</td>
<td>سعر</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>اطلالتة جميلة</td>
<td></td>
<td>خيام</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Algorithm 6.1: Algorithm the Creation of the 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
p_rootName //A root node of the tree that created recursively.
No_of_level 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 featuresList do
    Root.Add (feature); // append a node to the root

    synonymsList = Get_synonyms(feature); //return the synonyms for these feature
    For each synonym synonymsList 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 = 'HasA'
    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';
Union
Select start into V_list from ConceptNetDatabase where end = root and rel = 'CapableOf';
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 has 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;
}

---

**Figure 6.4: Algorithm for Creation Ontology Tree**

**6.4.1 Extract the Opinion Aspect**

After identifying the aspects in the review, the next step is to obtain opinion words related to the specific aspect. The polarity of the aspect can be obtained by identifying the opinion word related to its aspect. Opinion words may be adjectives, verb or nouns, such as "عَظِيم" (excellent), "أَحب" (love), "ممتاز" (prefer), and "لَيس حُسن" (not good). To determine the opinion word related to a certain feature, we used the following rules which are well suited for dealing with the dialectical Arabic language, and which are an effective approach to opinion detection for specific features. Figure 6.4 shows some example of the rules used in the proposed method. The first rule 6.1 is used to check the noun token which is followed by an opinion word in the adjectives category, and example being "الوجبة صغيرة". On the other hand, the second rule 6.2 is used to check opinion words in a verb category, followed by the first noun category, with an example being "أحب وجبات ماكدونالدز".
6.4.2 Polarity Identification

It is imperative that a list of opinion words is prepared, in order to facilitate the effective polarity identification. In this sense, opinion words hold positive or negative views or sentiments. The ArSenL (a large lexicon) was used in this study, from which the features extracted in the previous section can be classified.

The opinion orientation for each feature was identified through the prior-extracted features and opinion words, resulting in tuples of features and corresponding polarities. The tuples are made up of words related to the feature, which can be gathered through several ways. To confirm the validity of the proposed solution, it has been evaluated using the following three measures:

- N-grams Before, which involves obtaining the N-grams words before the feature in the user's review.
- N-grams After, which involves obtaining the N-grams words after the feature in the user's review.
- N-grams Around, which involves obtaining the N-grams words before and after the feature in the user's review.

N-grams refer to the number of words close to the feature, which are required for the purpose of determining polarity. If a negative word appears in a sentence it flips, or reverses, its polarity. The proposed method is capable of distinguishing negation words like ‘لا’, at which point it flips the orientation of the opinion. For instance, the sentence, “لا احب هذا النوع من الاطباق” “I do not like this kind of dishes” contains the word ‘لا’, which results in the conclusion that the orientation of the sentence is a negative one, even though ‘like احب’ is a positive word.

A review's global polarity is determined based on the majority-polarized features. If the majority of the features have a positive polarity, then the global polarity is positive. Likewise, it is negative, if most of the features are negatively polarized. Otherwise, the review is said to have neutral global polarity.
If (Token.category=="NN")
If (TokenNext1.category==JJ | | TokenNext1.category==DTJJ)
Then
    candidateFeature = Token;
    candidateOpinion= TokenNext1;

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

Rule 6.1
Rule 6.2

Figure 6.5: Rules Used to Detect the Adjective and Verb Opinion Words Related to the Feature

6.5 Experimental Results

The proposed technique have been evaluated using the benchmark twitter dataset, which contains 909 tweets, as described in Chapter 3. A random selection of 100 tweets was made for testing purposes, including 50 positive tweets and 50 negative tweets. The features were manually labelled in each review. The manual tagging was conducted through an approach involving two annotators reading the reviews, identifying the features, and determining the corresponding polarities. The polarities were labelled as 1 for a positive polarity, and -1 for a negative polarity. An illustration of the manual tagging of several phrases has been shown in Table 6.3. As a result of this phase, 187 aspects have been extracted and used as a baseline. Eventually the resulting manual results and output from the ontology were compared. Using the ontology, 152 aspects were extracted, 141 of which are correct aspects, which results in the output shown in Table 6.4. This table also gives the result of applying patterns, as detailed in Chapter 5, on the same tweet set. The comparison results have been made clearer through the graph showing different measures against the applying ontology and
patterns, as shown in Figure 6.5.

Table 6.4: Extracted Aspects Resulting from the use of Ontology and Patterns

<table>
<thead>
<tr>
<th>Method</th>
<th># of Aspects</th>
<th>Accuracy</th>
<th>Recall</th>
<th>Precision</th>
<th>F-mature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct by ontology</td>
<td>141</td>
<td>75.4%</td>
<td>75.4%</td>
<td>92.8%</td>
<td>83.2%</td>
</tr>
<tr>
<td>Patterns</td>
<td>126</td>
<td>67.4%</td>
<td>67.4%</td>
<td>96.9%</td>
<td>79.5%</td>
</tr>
</tbody>
</table>

In the course of the experiment, by extracting the set of explicit aspects, the researchers simply projected the lexical component of the ontology in the review. Aspects with noun tags were extracted, after which they were matched to the ontology, in order to narrow down the product aspects. Afterwards the opinion words which corresponded to the features were extracted from the previous step. Then the Arabic sentiment lexicon was used to determine the polarity of the opinion words. Eventually the overall polarity review was determined. Table 6.5 shows the confusion matrix used for the ontology classification approach. The rules shown in Figure 6.4 have been used to detect the adjective and verb opinion words related to the feature. The number of sentiments that were correctly classified by the system as positive were 43, while there were 39 negative, with an accuracy of 86.0% and 78.0% respectively. The results showed a satisfactory level of accuracy, as shown in Table 6.5, confirming that the polarity can be effectively established using ontology. However, due to the limited amount of knowledge in the area of ontologies, the field in general, and with large numbers of literal expressions, some aspects have been undetectable. When using ontology, many terms in the review can be missed due to a reduced coverage of all linguistic realizations of concepts and properties in a given domain. In order to overcome this limitation, linking features to opinion expressions has been used to partially resolve this problem. On the other hand, ontology properties are used to extract implicit features, which define relations between ontology concepts. For example, the property ‘look at’ serves to link the ‘customer’ and ‘design’ concepts. In the sentence “we eat too much”, the property ‘eat’ of the ontology
links ‘customer’ and ‘food’, allowing the system to determine that ‘too much’ refers to ‘food’. Examples of extracted aspects using ontology have been listed in Table 6.6.

**Table 6.5: Results for Sentiment Analysis**

<table>
<thead>
<tr>
<th>Method</th>
<th>Ontology</th>
<th>Baseline</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>43</td>
<td>50</td>
<td>86.0%</td>
</tr>
<tr>
<td>Negative</td>
<td>39</td>
<td>50</td>
<td>78.0%</td>
</tr>
</tbody>
</table>

**Table 6.6: Examples of the Extracted Aspects Using Ontology**

<table>
<thead>
<tr>
<th>Review</th>
<th>Aspects</th>
<th>Opinion words</th>
<th>Linguistic concepts</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>“really good restaurant with excellent food”</td>
<td>Restaurant, Food</td>
<td>Really good, Excellent</td>
<td>Eating place, food</td>
<td>Restaurant, food</td>
</tr>
<tr>
<td>“Very good restaurant but very expensive, I recommend pizzas and ice cream”</td>
<td>Restaurant, Pizzas, ice cream</td>
<td>Very good, expensive, Recommend</td>
<td>Meals, Price</td>
<td>Restaurant</td>
</tr>
<tr>
<td>“acceptable prices”</td>
<td>Prices</td>
<td>Acceptable</td>
<td>Price</td>
<td>Restaurant</td>
</tr>
<tr>
<td>“old fashion restaurant”</td>
<td>Restaurant</td>
<td>Old fashion</td>
<td>Restaurant</td>
<td>Restaurant</td>
</tr>
</tbody>
</table>

**Figure 6.6: Summary of Aspects Results Using Ontology and Patterns**

**Performance Measures**

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>100.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend:  
- **Blue Bar**: Ontology 
- **Purple Bar**: Patterns
6.6 SUMMARY AND CONCLUSION

Firstly, the researchers constructed the restaurant domain ontology using various restaurant descriptions and reviews. They then extracted customer reviews of the restaurant and stored each review as a flat text file. The preprocessing techniques were then performed on each review, as detailed in Chapter 3. Secondly, the domain ontology was used to identify only the restaurant aspects from the reviews. Finally, the sentiments orientation associated with each aspect were extracted, and polarity was assigned to each aspect based on ArSenL, which was constructed in Chapter 3. By summing up the sentiment scores associated with each aspect, a total score could be attained for each aspect. Finally, by summing up the scores of each aspect, a score could be attained for the restaurant. The main purpose of the ontology was to improve the process of aspect identification. The ontology contains concepts, attributes and relationships. Experimental results have demonstrated that the proposed approach not only leads to noticeable improvement over either patterns or association rule mining, but that it is also capable of extracting implicit aspects.
7 HYBRID ARABIC SENTIMENT ANALYZER

(HASA)

7.1 Introduction

A great deal of previous research work regarding aspect-based opinion mining has focused on frequency-based approaches, as detailed in Chapter 2. The approaches identify aspects that should be filtered out at a later stage. Relation-based approaches, on the other hand, make use of aspect-sentiment relationships which are useful in syntactic relations, particularly between aspects and sentiments. This study utilizes both approaches to come up with an Arabic Sentiment Analyzer, used for identifying aspects, and for defining semantic orientations using the opinion lexicon created in the previous stage.

These approaches have been effectively merged using predefined syntactic filtering patterns. Syntactic patterns, however, are language and text-specific, due to the differences in grammatical structure (see Appendix F).

This research study proposes a Hybrid Arabic Sentiment Analyzer (HASA), which mines opinions obtained from online restaurant customer reviews, and then summarizes them. The inputs processed by the sentiment analyser are the reviewed texts, while the processed output is a set of aspects, along with their polarity.

The Arabic sentiment analyser first segments the review into sentences, and then finds the nouns and noun phrases with the greatest frequency in the segments. By filtering frequent noun phrases, it is possible to mine a set of opinion patterns from the given text (the review). In addition, the Hybrid Arabic Sentiment Analyzer determines
the polarity of each aspect, depending on the Arabic Sentiment Lexicon (ArSenL) as outlined in Chapter 3.

This chapter is then organized in the following way. Section 7.2 presents the Hybrid Arabic Sentiment Analyzer (HASA). Section 7.3 presents the experimental results of the evaluation of a dataset from two corpora. Section 7.4 provides a discussion of experimental results. Finally, Section 7.5 offers a summary of the chapter.

### 7.2 Hybrid Arabic Sentiment Analyzer (HASA)

Most of the primary work on aspect-based opinion mining involves frequency-based approaches. They provide a set of candidate words, given that some words could represent aspects while others do not, and therefore they still require filtering in order to extract the aspects. Relation-based approaches make use of aspect-sentiment relationships, in order to pinpoint aspects referred to in the sentiments. Some of these relationships serve as a syntactic relation between aspects and sentiments. Therefore the Hybrid Arabic Sentiment Analyzer (HASA) has been proposed, which takes advantage of both approaches, in order to identify aspects and define semantic orientations using the Arabic Sentiment Lexicon (ArSenL), as constructed in Chapter 3.

The two approaches can be merged through using predefined syntactic filtering patterns specifically designed for the language, given that every language has a unique grammatical structure and syntax, such as that of MSA. Figure 7.1 describes the proposed Hybrid Arabic Sentiment Analyzer (HASA), which is capable of mining and summarizing opinions obtained from customer reviews. This takes review texts as the input, while the output is a set of aspects, combined with their polarity. It first divides the review into segments, and then uses the segments as transactions for finding frequent nouns or noun phrases, and consequently filtering them to identify aspects. Filtering the frequent noun phrases is undertaken through the syntactic relation to
group synonym aspects.

![Diagram of the Hybrid Arabic Sentiment Analyzer (HASA)](image)

Figure 7.1: Hybrid Arabic Sentiment Analyzer (HASA)

7.2.1 The Preprocessing Step

The pre-processing techniques previously discussed in Chapter 3 are applied to each sentence, and then the vector representations of the terms are retrieved from the textual representations, through considering term occurrences.

Customer reviews of products might be in an incorrect syntactic form, sentence
fragments or short phrases, and punctuations may be missing. The presence of adjectives in a sentence usually means that the sentence is subjective, and contains opinions (Liu, 2007).

A review sentence containing multiple aspects, one of the pertinent issues in aspect-based opinion mining, is split into units which hold only singular aspects. This issue is tackled using an aspect segmentation model. For example, the aspect segmentation model can be used to segment the review sentence “مطعم هادي و انيق” into two single feature units, specifically “مطعم هادي” and “مطعم انيق”.

The first course of action is the treatment of single-feature segmentation, by taking dependency relations for the nominal sentence, with adjective (صفة) and noun (موضوع). Table 5.1 shows the relation between an adjective and the noun it describes, together with the dependencies that link the pairs of nominals (predicate and apposition). The compound relation serves the purpose of forming numbers which are derived from single digit words.

### 7.2.2 Candidate Aspect Selection

Unsupervised learning approaches classify candidate aspects as being either opinion or non-opinion aspects. The classifier’s accuracy depends on the patterns used to determine the candidate aspects. The more restricted and specific the patterns are, the more accurate the identified candidate aspects are for the classifier. Combined patterns are proposed, which are composed of four different patterns, the definite base noun phrase (dBNP), and a semantic (subjective) base noun phrase (sBNP), the linking verb-based noun phrase (vBNP), and the preposition-based noun phrase (iBNP).
7.2.3 Aspects Extraction

The customer interest aspects are extracting with the use of a tool, which is capable of discovering frequent and infrequent patterns. In the context of this study, an item set is a phrase or set of words which appear together. Correlations are identified among the items in the set, through the use of association rules and ontology concepts with relations. These relationships are based on the data items’ co-occurrence, rather than on the data’s inherent properties, similar to functional dependencies.

7.2.4 Factors Effect Opinion Polarity

This section presents important factors in the proposed method, improving the performance of opinion mining. These factors are negation and intensifiers. Negation is important, as it has an effect on opinion words. Due to the use of negation, such as "لا أعشق", "لا أحب ماكدونالدز", "I don’t like McDonald’s", the opinion’s polarity is reversed. For example, "يُحب", "لا أعشق ماكدونالدز" (I like McDonald’s), and "لا أعشق ماكدونالدز" (I don’t like McDonald’s). Intensifiers refer to particles which intensify the strength of the polarity, such as "، جداا" in the example "، جميل جداا". Words like "، جميل" in the previous example intensify the strength of the opinion word "، جميل", which has a positive polarity, and accordingly receives the value +1 and the intensifier "، جداا" (very), thereby increasing the value of the adjective to +2, in order to emphasize the positive polarity in this sentence. As a result, the researchers manually formed a list of 72 words, including "، أوي", "، بشهدة", "، بدون شك", "، رهيب", "، موت", "، تماما", "، خالص", "، كثير", "، جدا", and others. These lists will be added to the polarity of opinion words. A list of negation and intensifier words has been provided in Appendix E.
7.2.5 Determining the Review’s Overall Polarity (OP)

Once the aspects have been extracted from the review, these are then matched with the ontology. The level of ontology, where it is located, determines the importance of the feature. The aspects appearing at higher levels in the ontology, close to the root, are considered to be more important than the lower-level ones. In addition, opinion words related to the higher-level feature are identified through the use of the opinion extraction process, while polarity is retrieved from the ArSenL lexicon, as detailed in Chapter 3. Eventually, the review’s overall polarity is based on the summation of the opinion polarity, multiplied by the height of the ontology for each feature, with respect to the existence of Negation and Intensifiers, as mentioned in the previous section. The following general formula has been proposed, in order to determine the overall polarity of a review.

\[
\text{Overall Polarity (OP)} = \sum \text{feature polarity} \times \text{height of ontology}
\]

7.2.6 Sentiment Analysis and Summary Generation

Upon completing the previously-discussed steps, the next action is the generation of a novel information summarizing method. This is based on the NLP technique, having the advantage of being straightforward or direct, and is comprised of the following five steps.

- Fetching the review from the directory.
- Applying the pre-processing to reviews.
- Fragmenting the reviews in sentences, depending on the adjectives used.
- Fetching an aspect from the aspect list.
- Assigning a weight to each feature mentioned in the sentences, as based on the opinion lexicon, which contains both positive and negative words for MSA and Dialect, and also for ontology.
- Noting the appearance of negation words in a sentence, such
as they reverse the expressed opinion.

- Obtaining the summation of the positive/negative weight of the opinion polarity for each aspect.
- Obtaining the summation of the positive or negative sentences of each aspect, in order to attain an overall text summary.

### 7.3 Experiment Results

**Experiment 1** Sentiment classification using the Semantic Orientation (SO) approach, without pre-processing. The objective of this experiment is to investigate the SO classifier’s performance. The experiment is conducted through the Twitter dataset, with sentiment words that have not been subjected to any pre-processing. The results of the classification experiment are presented in Table 7.1.

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td>72.5</td>
<td>65.3</td>
<td>68.9</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td>76.8</td>
<td>71.4</td>
<td>74.1</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td>72.5</td>
<td>65.3</td>
<td>68.9</td>
</tr>
<tr>
<td><strong>F-measure</strong></td>
<td>74.6</td>
<td>68.2</td>
<td>71.4</td>
</tr>
</tbody>
</table>

The results in Table 7.1 show that a large number of tweets have been incorrectly classified. This observation is the result of the following three reasons, specifically including 1) the absence of sentiment words, 2) the sentiment words used in the tweet that is not in the list, and 3) the sentiment word presented in the list as written in a form differing from ones in the list, such as using Hamza(ُ,١) and Yaa(٢,٧), for example ‘أجمل جميل’ - most beautiful’. Despite having similar meanings, the inclusion of prefixes and suffixes converts them into two distinctly-different words. One solution to this problem is to look out for the inflected forms of the sentiment words in tweets, which might offer an indication of the semantic orientation.
An alternative solution to the problem is the pre-processing of tweets and sentiment words, so as to effectively extract the sentiment words, while also classifying the tweets. The results show that the performance of the classified tweets almost matches that of the ML approach (Chapter 4), attributed to the lack of processing (lack of structure) and the noise in text contained in tweets, which negatively affects the classifier’s performance.

**Experiment 2** The impact of pre-processing on semantic orientation classifier. The objective of this experiment is to test the effect of pre-processing on the SO performance. Three experiments were conducted using pre-processed tweets and sentiment words, with one experiment being undertaken at each phase. Prior to the experiment, stop words were eliminated to ensure ease of classification, although they did not affect the performance. In the first experiment, the tweets and sentiment words were normalized in order to serve the purpose of transforming the text into a common form, so as to therefore maintain its consistency while reading (Table 3.6). In the second experiment, the normalized tweets and sentiment words have also been stemmed. The effect of the removal of stop words on performance has not been tested, as there is no evidence of intersection between sentiment words and stop words. This has consequently results in no effect on SO performance.

The results of the experiment, regarding the effect of pre-processing on SO classifier performance, have been shown in Tables 7.2, 7.3 and 7.4.

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td>72.8</td>
<td>65.8</td>
<td>69.3</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td>76.7</td>
<td>71.1</td>
<td>73.9</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td>72.8</td>
<td>65.8</td>
<td>69.3</td>
</tr>
<tr>
<td><strong>F-measure</strong></td>
<td>74.7</td>
<td>68.3</td>
<td>71.5</td>
</tr>
</tbody>
</table>
Tables 7.2, 7.3 and 7.4 show the performance results for the binary classifiers at each pre-processing stage. Tables 7.3 and 7.4 show results following the application of two stemmers, specifically the normal stemmer and the light stemmer.

The experiment regarding the effect of pre-processing on SO performance shows a 6.6% improvement in accuracy and recall, a 2.6% improvement in precision, and a 4.7% improvement in F-measure. SO performance is only affected by the form of the sentiment words. Once pre-processing was carried out, the sentiment words were transformed into a form similar to that of the lists, which thereby facilitated easier extraction. Sentiment words form a significantly small portion of the tweet, and not all tweets have sentiment words. Therefore, the construction of more comprehensive lists of sentiment words can contribute towards improved performance.

An analysis of Tables 7.3 and 7.4 has revealed almost similar results from both stemmers. This is somewhat expected, as the stemming of the sentiment words is the same, given that there are fewer dialect-specific words.

A comparison of the negative and positive binary classifiers has revealed that a positive classifier’s performance improves the negative classifier’s performance. Once the stemmer is applied, the performance of the two classifiers draws closer to almost

---

**Table 7.3: SO Results Using a Normal Stemmer**

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>76.0</td>
<td>75.8</td>
<td>75.9</td>
</tr>
<tr>
<td>Precision</td>
<td>76.1</td>
<td>77.0</td>
<td>76.5</td>
</tr>
<tr>
<td>Recall</td>
<td>76.0</td>
<td>75.8</td>
<td>75.9</td>
</tr>
<tr>
<td>F-measure</td>
<td>76.0</td>
<td>76.4</td>
<td>76.2</td>
</tr>
</tbody>
</table>

**Table 7.4: SO Results Using a Light Stemmer**

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>75.3</td>
<td>75.5</td>
<td>75.4</td>
</tr>
<tr>
<td>Precision</td>
<td>75.8</td>
<td>76.3</td>
<td>76.0</td>
</tr>
<tr>
<td>Recall</td>
<td>75.3</td>
<td>75.5</td>
<td>75.4</td>
</tr>
<tr>
<td>F-measure</td>
<td>75.5</td>
<td>75.9</td>
<td>75.7</td>
</tr>
</tbody>
</table>
similar performance measures. This is attributed to the fact that positive tweets contain less noise, when compared to negative tweets. Thereby, the minimal processing involving normalization, results in the achievement of best results.

**Experiment 3** use of CRF. The objective of this experiment is conducted to evaluate the effectiveness of HASA model CRF uses a discriminative undirected probabilistic graphical model. They were first introduced by Lafferty et al. (2001) for the task of labeling sequential data for speech recognition tasks. It is used to model known relationships between observations and then construct consistent interpretation. It is widely used in sequence labeling problems, i.e. Natural Language Processing such as part of speech tagging, Named Entity Recognition (NER) (Stanford NER by Finkel et al.,2005), language identification (Samih and Maier, 2016), and Information Extraction (IE). It was also used in other problems such as biological sequencing, image and video labeling, and image recognition. Similarly, Linear-chain CRF has been applied to Aspect extraction tasks because the problem of finding aspects in a sentence can be viewed as a sequence labeling problem (Jakob and Gurevych, 2010).

The 10-fold cross validation CRF is applied to Twitter dataset, for the task of aspect extraction as a supervised method. CRF relies on a defined set of features that are fed to the model during training. We used the same sets of features for HASA. MSA Arabic NLP tools were used to create the set of features to be used for Arabic CRF models. More precisely, we employed Stanford Core NLP (Manning et al., 2014) to derive the set of features to be used in our CRF model. Those set of features are defined as fallowed:

- **Part of Speech tags (POS):** a set of POS tags applied through the use of part of speech tagger. We employed Stanford POS tagger from the Stanford NLP Core suit to extract POS tags.
- **Noun Around (NA):** the closest noun phrase to sentiment word (Jakob & Gurevych, 2010).
- **Sentiment words (SW):** the sentiment words are identified by the model and they are the words that holds opinion regarding the aspect being reviewed.
We applied 10-fold cross validation CRF using a combination of these features. Table 7.5 shows the result of building CRF model using various features.

Table 7.5: Results of Using CRF

<table>
<thead>
<tr>
<th></th>
<th>F-measure</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter dataset</td>
<td>71.1</td>
<td>86.7</td>
<td>67.5</td>
</tr>
<tr>
<td>Review dataset</td>
<td>73.6</td>
<td>85.5</td>
<td>70.0</td>
</tr>
</tbody>
</table>

Since the performance of CRF depends highly in the features extracted and most of those features extracted using NLP tools. This shows the important role that NLP plays in extracting those features. The lack of dialect tools and better sentence structure contributed to the lower performance of the CRF method. On the other hand, review dataset yield better performance than Twitter dataset. We contribute the better performance to the amount of the data. The review dataset is a larger set than Twitter dataset which provide a better training set. Also, review dataset are longer than Twitter dataset which provide a better training and probably provide a better probability of features per sentence.

7.4 Discussion of Results

The proposed technique was evaluated in order to see how effective it is in identifying product aspects from a set of corpora, which constitutes collected tweets and customer reviews written in Arabic. Each review is a short text. In these Arabic scripts, some of the letters of the alphabet were normalized, particularly ones with multiple forms. The duplicities are eliminated, and some of the wrongly-spelt words are corrected.

We found out that our proposed model performed better than the CRF model. We also found out that the CRF method relies on training and a set of features that are considered the basis for the model. Most of the features that we used are simple and can be obtained using simple NLP tools. The main drawback of CRF method is that it
requires a dataset with labels and the larger the dataset the better the model performance as is the case with most supervised methods. The availability of a dataset specifically labeled for this task might be a challenge for many low resource languages, thus our proposed mode is preferable in this case. Figure 7.2 shows an example of tweet polarity identification, Figure 7.3 shows an example of target identification. Figure 7.4 provides an illustration of the summary for the feature ‘السّر’. Figure 7.5 shows that the proposed technique handles the negation word.

7.5 Conclusion

This chapter provides a description of the proposed set of techniques for mining and summarizing product reviews, constructed on the basis of data mining and natural language processing methods. The main objective highlighted in the study is the construction of HASA, which serves the purpose of providing a feature-based summary of large volumes of customer reviews written in Arabic. The proposed techniques are subjected to experiments, whose results show satisfactory results, thereby proving the effectiveness of the proposed techniques. The problem solved in the course of this study is expected to become a growing concern, given that the number of people making online purchases and posting reviews continues to grow. The summary of reviews is valuable to customers, as well as to product manufacturers.
Figure 7.3: Tweet Target Identification

Figure 7.4: An Example Summary of the Feature ‘السعر’

Figure 7.5: Handling the Negation Word
CHAPTER EIGHT

8 CONCLUSIONS AND FUTURE WORK

8.1 Conclusion

The widespread growth of the social web has had a profound effect on the creation of various research interest areas. One particular interest is the possibility of acquiring valuable information regarding customer’s views and opinions of products and services, which is beneficial both to the field of information retrieval, and to computational linguistics. This information acquisition process is referred to as opinion mining, or sentiment analysis. A great deal of sentiment analysis research at the document, sentence and feature levels has been conducted for the English language, but very little has been done in the Arabic language. In this thesis, the researchers proposed an aspect level opinion mining classification used to detect the polarity of Arabic opinion reviews. Furthermore, the proposed approach has been combined with ontology information, in order to offer better opinion mining classification performance.

In the course of the research, an innovative aspect-based opinion mining methodology has been proposed which uses a hybrid approach. The proposed methodology is comprised of four stages, including i) the construction of a noun phrase pattern, ii) the construction of domain ontology (specifically for restaurants), iii) using the pattern and ontology for aspect identification, and iv) assigning the polarity to each aspect, based on the Arabic Sentiment Lexicon (ArSenL).
A hybrid approach has been proposed in this study, combining semantic relation and syntactic sequence. Semantic relation involves using adjectives which have polarity, while syntactic sequence involves using patterns.

Ontology is a concept model that provides a description of the system at semantic and knowledge levels. The purpose of ontology is to gain an understanding of concepts in a domain, and the relationships between them, so as to realize the knowledge of the domain. Opinion mining, which is a demanding problem, involves evaluating opinions contained in review texts sent by users, concerning various features, in order to establish their positive, negative or neutral polarity, and the strength of these opinions. This research study focuses on using an ontology to mine restaurant reviews, since an ontology facilitates the acquisition of knowledge within in a specified domain, in a form that is well understood by both developers and computer systems.

First of all, an evaluation of the effect of text pre-processing on classification accuracy has been undertaken through experimentation. The evaluation proceeded through three stages. The first stage was a pre-processing mechanism, involving normalization, stemming and the elimination of stop words, for tweets and reviews, which was recommended prior to its use in noise reduction and in reducing unstructured text. The second stage was the recommendation of a machine learning approach, which incorporates the evaluations of features to be applied. The third stage was the recommendation of an SO approach, which involved the construction of a sentiment lexicon which extracts sentiment words and calculates their orientation, so as to facilitate the classification of sentiment tweets in Arabic, and to provide an easy approach for handling negation.

In accordance with the proposed approach, the research study kicked off with the construction of feature vectors, involving the exclusive use of unigrams as the features from the TDASA dataset serving as the training data. The data was then applied to the SVM, NB and KNN classifiers, so as to be able to identify the classifier with the most accurate results. It was observed that the application of a single feature SVM, the unigram frequency, produces the best classifier. The study went on to show
the effect of pre-processing on machine learning performance, when applied on the same dataset. The result showed that pre-processing improved the accuracy of the overall classifier.

Secondly, pattern and ontology were used to improve the process of aspect identification. Ontology contains concepts, attributes and relationships. In due course of this research, data mining and natural language processing techniques relevant to each phase of the research were studied. Firstly, the reviews and tweets were crawled and stored separately in a text file. Each text file was pre-processed and stored in a structured format. These structured reviews were then used to extract only those restaurant aspects that were present in the ontology. In the next stage of this research, the sentiments adjacent to the aspects were extracted.

Thirdly, HASA incorporated frequency, as well as relation-based approaches in the identification of relevant aspects and opinion classification. HASA identified the relationships between aspects and sentiments, through the opinion features that have been mined from the reviews. Afterwards, the ontology, together with the mined patterns, was used to eliminate the non-aspects from the frequent noun phrases. A novel technique was applied in grouping synonymous aspects. HASA was used to establish the strength of an opinion feature, through its classification according to ArSenL, and the corresponding summary generation. Upon evaluating the results, the conclusion drawn was that the combined use of the frequency approach, together with the relation-based approach, resulted in a greater-improved aspect extraction accuracy. Moreover, a method for detecting negation words was also incorporated into the proposed approach, so as to effectively handle instances in which negation sentiments were included in review texts.

The work carried out, as presented in this thesis, has been based on the patterns and ontology approach to Arabic sentiment analysis. The opinion corpuses, restaurant ontology and HASA framework, could all be used in the future to investigate new methodologies and resources for sentiment analysis.

Therefore, this research study’s main contribution is the proposal of aspect-based
sentiment analysis for the Arabic language, which involves extracting aspects from Arabic text contained in a blog, review, tweet, for example, and then determining its orientation as being one that contains either overall positive or negative sentiments. This classification is based on a new feature set, comprised of a combination of machine learning and SO features, patterns and ontology for aspect selection, and coupled with an easy approach to negation detection. The implementation of the proposed approach exhibits better performance measures, compared with other aspect-based sentiment classification systems which use ML or SO approaches.

8.2 Contributions

This thesis introduces the aspect-based sentiment analysis of Arabic tweets, and enhances the technique that uses association rules in general patterns and ontology. This study’s contributions can be grouped as follows:

1. Designing a new aspect-based sentiment classification model for Arabic tweets, based on association rules mining.

The Association Mining approach for product features extraction was first used by Hu and Liu (2004), who extracted frequent features in their work, using the association rule mining technique (Agrawal and Srikant, 1994). The algorithm’s earliest implementation was in market basket analysis, whereby the level of dependency of a sold item was measured against another item. Using this analogy, Hu and Liu (2004) assumed that words in a sentence are the sold items, while the association between the terms can be used to predict features and opinions. This method was found to be useful for feature extraction. Wei et al. (2010) later extended this approach for application in semantic-based pruning, for the refinement of frequent features and the identification of infrequent features. This refined approach showed improved results, in terms of opinion target identification.
2. Enhancing the new aspect-based sentiment classification model, using noun phrase patterns.

3. The design of a novel fuzzy domain ontology for restaurants, consisting of concepts and attributes associated with taxonomic and non-taxonomic relationships between them.

4. Using the domain ontology during the aspect selection stage in opinion mining, and extracting the sentiments associated with the aspects.

5. Enhancing the new aspect-based sentiment classification model, using patterns.

6. Combining the use of these approaches to enhance the overall model accuracy.

8.3 Future Work

The work carried out during the course of this study is neither optimal nor conclusive, as scientific advancements are being made continuously and progressively. Given that the Arabic sentiment analysis research field is still in its infancy, there are numerous potential areas that could be still expanded upon. These improvements could commence in the first stage through the introduction of a new dialect Arabic sentiment corpus and could culminate in the application of one domain within the prediction of the polarity of another domain. There are several alternatives in which the system may be extended, some of these which include:

- An exploration of the effect that including hash tags (مطاعم#), positive emoticons(,:) and negative emoticons(;) into the training data and in data collection, has on performance accuracy. The study by Kouloumpis et al. (2011) states that both collection methods have proven useful.

- Given that there are several Arabic dialects with unique vocabularies and structures, a potential research direction would entail the construction of a comprehensive list of positive, negative and negation words for each of these dialects. An example could be "مرده،نور،أو،نور،حسناً،سالفتي".

- Since a great deal of research work conducted in this field only extracts explicit aspects, another possible research direction could be the
identification of implicit aspects. However, a review may contain numerous types of implicit aspect expressions, which may pose a challenge. The most commonly-used expressions are adjectives and adverbs, since the intrinsic nature of adjectives is to describe attributes or properties that belong to entities. For example, ‘>Lorem ipsum’

Moreover, intrinsic aspects may be verbs, which further adds to the complexity. Although some previous research has been conducted in this area, there is still room for further work.

- The development of a finer ontology, which will contribute towards improved accuracy in feature extraction.
- Using the constructed ontology to determine important features highlighted in an expressed opinion, and then going on to identify which features are more important than others. Feature importance can be captured using the height or level of a feature node within the ontology tree.
- Generating a summary of sentiment scores for each aspect and using these scores to rate the restaurant.
- Applying the proposed approach to another domain, such as Smartphone reviews, and evaluating its consequential performance.
REFERENCES


Esuli, A. and Sebastiani, F., (2005, October). Determining the semantic orientation of terms through
gloss classification. In *Proceedings of the 14th ACM international conference on Information and knowledge management* (pp. 617-624). ACM.


Hatzivassiloglou, V. and McKeown, K.R., (1997, July). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics* (pp. 191...


Appendices

Appendix A: Samples of tweets and reviews used in classification and manual annotation

I. Samples of Positive Tweets and Reviews:

<table>
<thead>
<tr>
<th>Review</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. من أفضل مطاعم بالرياض جلسات رائعة وواكل لذيذ ونظيف انسجح به وبشده إذا كنت من يحبون الأكل الارمني والمصاوي هذا على مودك ومزاجك لاالليلة فقط.</td>
<td>Review dataset (Trip Advisor&amp;Qaym)</td>
</tr>
<tr>
<td>2. المطعم رائع جدا وهادي وجلساته رايق جدا والاسعار مقبوله نوعا ما والخدمة ممتازة جدا وموقعه أيضا هادي جدا.</td>
<td></td>
</tr>
<tr>
<td>3. من افخم المطاعم اليابانية في المملكة مطعم نوروزمي يقدم تشكيلة مميزة من الوجبات الشهية منيو مطعم نوروزمي في السعودية.</td>
<td></td>
</tr>
<tr>
<td>1. دجاج روز ماري في مطعم سيزر هاوس المطعم رائع والخدة ممتازة والاكل جميل.</td>
<td>Twitter dataset</td>
</tr>
<tr>
<td>2. مطعم ابتكر على ذوقك طعم جديد ورائع لجر الاجوس من مالكلصم البرجر على كيفك ابتداء من الخزى وحتى الصوص.</td>
<td></td>
</tr>
<tr>
<td>3. مطعم عناج من الفطير المميزة عندهم حجازية دجاج و بالأخص ليبانيز فطيرة الجرجر جدا لذيذة.</td>
<td></td>
</tr>
</tbody>
</table>
## II. Samples of Negative Tweets and Reviews:

<table>
<thead>
<tr>
<th>Review</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. بسراحي مطعم سيئ وغالي أنا شخصياً دفعت 100 ريال حق العشاء وطلعت تعشيت بمكان ثاني لأن أكلهم سيئ جداً .. وطبعا جو المطعم حلو والمكان فخم.</td>
<td>Review dataset (TripAdvisor &amp; Qaym)</td>
</tr>
<tr>
<td>2. مطعم هياط من الدرجة الأولى دفعنا 100 ريال لـ 4 أشخاص وطلعنا جوعاين الاطبق الصغير ناسب الحجم الياباني.</td>
<td></td>
</tr>
<tr>
<td>3. الثمن مقابل الطعام غير متكافئ .. خيب التوقعات تعاملهم جاف .. بعض الفلبينيين جنسين في التعامل .. بعض الاطبقات قدمت باردة .. طبق الكباب بالكرز غالي.</td>
<td></td>
</tr>
<tr>
<td>1. أكيد هذا مطعم فايف فايز الجديد طعم شين وغلاء سعر ولا حممه بعد تجربته لن تعاد. 2. أبل بيز اللي بالظهرا مطعم تحت السواء .. يفشل قسم. 3. كنتاكي اسووه مطعم جربته بحيائي لا خدمة ولا طعم ولا أسعار</td>
<td>Twitter dataset</td>
</tr>
</tbody>
</table>
### III. Samples of Manually-annotated Tweets and Reviews

<table>
<thead>
<tr>
<th>Tweet</th>
<th>Manual Labelling</th>
<th>Polarity</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>مطعم سبرام هاس يوجد في الرياض عامود للتحليه اسم بابدي من ارقى وافضل المطاعم وخدمة رائعة وتعامل ممتاز</td>
<td>ارقي المطاعم</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>في مطعم عدناء اسمه فيرور فين العسل وصلو للسايرا من الدجاج رائحة وكأن سعر جللي اطالتها جميلة لكن أصلي خيال</td>
<td>سعر عالي</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>اتخدم هذا مطعم فايز الجديد تهذم شين وغلاء سعر ولا زحمه بعد تجريه لن تعا</td>
<td>طعم شين</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>رخت له مره وحده اجرب عربي ويعود (ساعة انتظار) لمعن شين والسعر غالي كنت متوقع انه مطعم زي الاولاد اشارمر</td>
<td>السعر غالي</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>مطعم فطار فارس أسعار مرتوبة ومع ذلك فيه منتجات مضوية شين وقوي عين</td>
<td>أسعار مرتبة ومع 1 منتج مضوية</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>أضعه بنبيك زيادة في الأسعار بشكل غيري مع نفس في جودة الطبخ يعني نأكل كيلو السمك ب250 ريال</td>
<td>أسعار غالية</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>أكبر مطعم حرامي في الأسعار صحن صغيرة رز ب300 ريال اتوقع لو طلب ذهب ما يوصل بهذا السعر ؟</td>
<td>السعر الكمية</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>القدر المسمك غالي ونفر الز</td>
<td>أسعار المسمك غالي</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>مطعم مجلس الحسيني زود على أسعار المبالغ فيها جودة سنة جدا فيلم عدة بسيطة</td>
<td>سعر الجودة</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>للاسف رز بياو كان المطعم سبأ سيء للحم بارد وفاصلي لم يستطيع أكله وسعر مرتغتر وليست بسأطابق صحن ذهبي مع منتج بسيط</td>
<td>السعر</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>مطعم أبو كمال في المعلطبات شيز توافر محمر زيادة عيدا مجمون مرتغتر وصوص بالصين إيه شي سيء البطاطس واللومبوتو سعر مرتغتر</td>
<td>السعر</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>السعر</td>
<td>المطعام</td>
<td>النظام الغذائي</td>
<td>BEGIN Table</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>سعر متوسط 1</td>
<td>طلبات</td>
<td>4 طلبات</td>
<td>كاتلوني وملعقة متنازٍن، سوشي جيداً، كملاريمي س٢ء، قلي س٢ء</td>
</tr>
<tr>
<td>المطعم</td>
<td>المطعم دارين رقم قدمه الا حلوة</td>
<td>ببطاطس ببطاطس وأسعار حلوه جدا</td>
<td>التغطية</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------</td>
<td>---------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>مطعم الفرنسا</td>
<td>المطعم رقم قدمه الا</td>
<td>ببطاطس ببطاطس وأسعار حلوه جدا</td>
<td>التغطية</td>
</tr>
<tr>
<td>المطعم 1</td>
<td>المطعم رقم قدمه الا</td>
<td>ببطاطس ببطاطس وأسعار حلوه جدا</td>
<td>التغطية</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>المطعم</th>
<th>المطعم دارين رقم قدمه الا حلوة</th>
<th>ببطاطس ببطاطس وأسعار حلوه جدا</th>
<th>التغطية</th>
<th>السعر</th>
</tr>
</thead>
<tbody>
<tr>
<td>مطعم الفرنسا</td>
<td>المطعم رقم قدمه الا</td>
<td>ببطاطس ببطاطس وأسعار حلوه جدا</td>
<td>التغطية</td>
<td>السعر</td>
</tr>
<tr>
<td>المطعم 1</td>
<td>المطعم رقم قدمه الا</td>
<td>ببطاطس ببطاطس وأسعار حلوه جدا</td>
<td>التغطية</td>
<td>السعر</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>المطعم</th>
<th>المطعم دارين رقم قدمه الا حلوة</th>
<th>ببطاطس ببطاطس وأسعار حلوه جدا</th>
<th>التغطية</th>
<th>السعر</th>
</tr>
</thead>
<tbody>
<tr>
<td>مطعم الفرنسا</td>
<td>المطعم رقم قدمه الا</td>
<td>ببطاطس ببطاطس وأسعار حلوه جدا</td>
<td>التغطية</td>
<td>السعر</td>
</tr>
<tr>
<td>المطعم 1</td>
<td>المطعم رقم قدمه الا</td>
<td>ببطاطس ببطاطس وأسعار حلوه جدا</td>
<td>التغطية</td>
<td>السعر</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>المطعم</th>
<th>المطعم دارين رقم قدمه الا حلوة</th>
<th>ببطاطس ببطاطس وأسعار حلوه جدا</th>
<th>التغطية</th>
<th>السعر</th>
</tr>
</thead>
<tbody>
<tr>
<td>مطعم الفرنسا</td>
<td>المطعم رقم قدمه الا</td>
<td>ببطاطس ببطاطس وأسعار حلوه جدا</td>
<td>التغطية</td>
<td>السعر</td>
</tr>
<tr>
<td>المطعم 1</td>
<td>المطعم رقم قدمه الا</td>
<td>ببطاطس ببطاطس وأسعار حلوه جدا</td>
<td>التغطية</td>
<td>السعر</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>المطعم</th>
<th>المطعم دارين رقم قدمه الا حلوة</th>
<th>ببطاطس ببطاطس وأسعار حلوه جدا</th>
<th>التغطية</th>
<th>السعر</th>
</tr>
</thead>
<tbody>
<tr>
<td>مطعم الفرنسا</td>
<td>المطعم رقم قدمه الا</td>
<td>ببطاطس ببطاطس وأسعار حلوه جدا</td>
<td>التغطية</td>
<td>السعر</td>
</tr>
<tr>
<td>المطعم 1</td>
<td>المطعم رقم قدمه الا</td>
<td>ببطاطس ببطاطس وأسعار حلوه جدا</td>
<td>التغطية</td>
<td>السعر</td>
</tr>
<tr>
<td>المطعم</td>
<td>المطعم مطعماً في لوول هاير ماركت من الأد مطاعن الهندية وإستغزه وخصه وجلساه جمله</td>
<td>السعر</td>
<td>الشريحة</td>
<td>الملمع</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------------------------------------------</td>
<td>------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>لا حسبت اليوم زيادة استعاز مطعماً بيل بيز الجيل وسأله من متي الزيادة قابلو من منهل بيق له</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>روح مطعماًجالاجليحية أكلات شعبية وبغرر مطعماً مقول والمحليفي شعي وحلج بوس</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>المدينة المنورة؟ مقاطعة مطعماً البانسي لزيادة الاستعاز. رجل لا يخفيف الله نزلت اساعر</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>و بالماظيarti بيطمراجع جزين بيجده وحنكت بطمع بيتزا وأساعر خياليه</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>مهاني الوزير لا حق مطيع بيتزا رفع اساعر التوصيل من ريانال إلى خمسة ريالات علما انتملت كلي ابلغ ورفض بلاغي بجدة إنها خدمه!</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>أسعار مرفعة مطعماً سيء، كميمضاج قليله بخدمة سهية .. بااختصار مطعماً جمع ويثومن عن أكبر قدر من الربح بمعه</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>كنتيكي أسوي مطعماً جرينغة لا خدمه ولا طعم ولا أسعار</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>مطعماً مستهلكة وأساعرها خياليه</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>مطعماً ريدان</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>مطعماً الفزم</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>وقفو التيمون و البصة مع</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

الكل: 1000
<table>
<thead>
<tr>
<th>الطلب</th>
<th>السعر</th>
</tr>
</thead>
<tbody>
<tr>
<td>الطلب 1</td>
<td>1 نون.</td>
</tr>
<tr>
<td>الاسم</td>
<td>المطلوب</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>المطعم</td>
<td>السعر</td>
</tr>
<tr>
<td>المطعم</td>
<td>السعر</td>
</tr>
<tr>
<td>المطعم</td>
<td>السعر</td>
</tr>
</tbody>
</table>

**ملاحظة:** لا يمكنني قراءة النص العربي بشكل طبيعي. إذا كنت بحاجة إلى مساعدة في شيء آخر، فأخبرني بذلك.
Appendix B: Samples from the List of Stop Words

<table>
<thead>
<tr>
<th>كمان</th>
<th>كله</th>
<th>وفيها</th>
<th>اه</th>
<th>خالص</th>
</tr>
</thead>
<tbody>
<tr>
<td>على</td>
<td>ان</td>
<td>مع</td>
<td>عن</td>
<td>يا</td>
</tr>
<tr>
<td>الكل</td>
<td>اللي</td>
<td>ما</td>
<td>علي</td>
<td>هو</td>
</tr>
<tr>
<td>دة</td>
<td>دي</td>
<td>ده</td>
<td>بس</td>
<td>بعد</td>
</tr>
<tr>
<td>ونائ</td>
<td>مكشش</td>
<td>برده</td>
<td>احد</td>
<td>حيد</td>
</tr>
<tr>
<td>حضرته</td>
<td>عليه</td>
<td>بعض</td>
<td>انا</td>
<td>الذي</td>
</tr>
<tr>
<td>زيك</td>
<td>مقدم</td>
<td>بنش</td>
<td>زيك</td>
<td>زيك</td>
</tr>
<tr>
<td>لانها</td>
<td>دولة</td>
<td>انت</td>
<td>قد</td>
<td>حيث</td>
</tr>
<tr>
<td>بس</td>
<td>انت</td>
<td>قد</td>
<td>حيث</td>
<td>حيث</td>
</tr>
</tbody>
</table>

Appendix C: Samples from the Lists of Sentiment Words

<table>
<thead>
<tr>
<th>Positive Sentiment Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>حبيب</td>
</tr>
<tr>
<td>التزم</td>
</tr>
<tr>
<td>عبد</td>
</tr>
<tr>
<td>افرج</td>
</tr>
</tbody>
</table>

207
### Negative Sentiment Words

<table>
<thead>
<tr>
<th>الفاضلي</th>
<th>غالي</th>
<th>تحلم</th>
<th>زحجه</th>
<th>غلاء</th>
<th>غالية</th>
<th>نفص</th>
<th>خياس</th>
<th>بخيل</th>
<th>عجر</th>
<th>العيب</th>
<th>زيدة</th>
<th>الطمع</th>
<th>تنغشون</th>
<th>سي</th>
<th>بسج</th>
<th>قدم</th>
<th>كرهك</th>
<th>خساره</th>
<th>فاشل</th>
<th>زناخه</th>
</tr>
</thead>
<tbody>
<tr>
<td>واع</td>
<td>يتسم</td>
<td>طدم</td>
<td>نوخ</td>
<td>نسو</td>
<td>عام</td>
<td>غبيان</td>
<td>شيّن</td>
<td>جتم انقل</td>
<td>ينقرم</td>
<td>ينترمزون</td>
<td>ينبق</td>
<td>ينبق</td>
<td>ينبق</td>
<td>ينبق</td>
<td>ينبق</td>
<td>ينبق</td>
<td>ينبق</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>اتخمت</td>
<td>اختقم</td>
<td>اتخمار</td>
<td>اتخمار</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td>اتخمر</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Appendix D: 1. Samples from the List of Negation Words

<table>
<thead>
<tr>
<th>الكلمة</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>لا</td>
<td>1</td>
</tr>
<tr>
<td>ما</td>
<td>2</td>
</tr>
<tr>
<td>بلا</td>
<td>3</td>
</tr>
<tr>
<td>دون</td>
<td>4</td>
</tr>
<tr>
<td>غير</td>
<td>5</td>
</tr>
<tr>
<td>عدم</td>
<td>6</td>
</tr>
<tr>
<td>لسنا</td>
<td>7</td>
</tr>
<tr>
<td>بغير</td>
<td>8</td>
</tr>
<tr>
<td>لات</td>
<td>9</td>
</tr>
<tr>
<td>عديم</td>
<td>10</td>
</tr>
<tr>
<td>أن</td>
<td>11</td>
</tr>
<tr>
<td>لما</td>
<td>12</td>
</tr>
<tr>
<td>لم</td>
<td>13</td>
</tr>
<tr>
<td>أن</td>
<td>14</td>
</tr>
<tr>
<td>ليست</td>
<td>15</td>
</tr>
<tr>
<td>ليس</td>
<td>16</td>
</tr>
<tr>
<td>لست</td>
<td>17</td>
</tr>
</tbody>
</table>
2. Sample from the List of Intensifier Words

<table>
<thead>
<tr>
<th>#</th>
<th>الكلمة</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>جدا&quot;</td>
</tr>
<tr>
<td>2.</td>
<td>كثيرا</td>
</tr>
<tr>
<td>3.</td>
<td>خاصص</td>
</tr>
<tr>
<td>4.</td>
<td>تماما</td>
</tr>
<tr>
<td>5.</td>
<td>موت</td>
</tr>
<tr>
<td>6.</td>
<td>رهيب</td>
</tr>
<tr>
<td>7.</td>
<td>بدون شك</td>
</tr>
<tr>
<td>8.</td>
<td>بشدة</td>
</tr>
<tr>
<td>9.</td>
<td>أو</td>
</tr>
<tr>
<td>10.</td>
<td>مره</td>
</tr>
</tbody>
</table>

Appendix E:

The traditional Arabic grammar of *iʿrāb* (إعراب) assigns a syntactic role to each word in a sentence. Pairs of syntactic units are related through directed binary dependencies. In the Arabic language, these relations are represented as directed edges on dependency graphs. The following tables list dependencies which are used to relate morphological segments, words, phrases and clauses (Rahamatallah, L, et al., 2015).

**Table F.1: Dependency Relations for Nominal**

<table>
<thead>
<tr>
<th>Relation</th>
<th>Arabic Name</th>
<th>Dependency</th>
<th>Dependent → Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adj</td>
<td>صفة</td>
<td>Adjective</td>
<td>adjective → noun</td>
</tr>
<tr>
<td>poss</td>
<td>مضاف إليه</td>
<td>Possessive construction</td>
<td>second noun → first noun</td>
</tr>
<tr>
<td>pred</td>
<td>ميدا و خير</td>
<td>Predicate of a subject</td>
<td>predicate → subject</td>
</tr>
<tr>
<td>app</td>
<td>بدل</td>
<td>Apposition</td>
<td>second noun → first noun</td>
</tr>
<tr>
<td>spec</td>
<td>تميز</td>
<td>Specification</td>
<td>second noun → first noun</td>
</tr>
<tr>
<td>cpnd</td>
<td>مركب</td>
<td>Compound</td>
<td>second number → first number</td>
</tr>
</tbody>
</table>
### Table F.2: Dependency Relations for Verbs

<table>
<thead>
<tr>
<th>Relation</th>
<th>Arabic Name</th>
<th>Dependency</th>
<th>Dependent → Head</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subj</strong></td>
<td>فاعل</td>
<td>Subject of a verb</td>
<td>subject → verb</td>
</tr>
<tr>
<td><strong>Pass</strong></td>
<td>نائب فاعل</td>
<td>Passive verb subject representative</td>
<td>subject representative → verb</td>
</tr>
<tr>
<td><strong>Obj</strong></td>
<td>معول به</td>
<td>Object of a verb</td>
<td>object → verb</td>
</tr>
<tr>
<td><strong>Subjx</strong></td>
<td>اسم كان</td>
<td>Subject of a special verb or particle</td>
<td>subject → verb or particle</td>
</tr>
<tr>
<td><strong>Predx</strong></td>
<td>خبر كان</td>
<td>Predicate of a special verb or particle</td>
<td>predicate → verb or particle</td>
</tr>
<tr>
<td><strong>impv</strong></td>
<td>أمر</td>
<td>Imperative</td>
<td>imperfect verb → imperative particle</td>
</tr>
<tr>
<td><strong>imrs</strong></td>
<td>جواب أمر</td>
<td>Imperative result</td>
<td>result → imperative verb</td>
</tr>
<tr>
<td><strong>pro</strong></td>
<td>نهي</td>
<td>Prohibition</td>
<td>imperfect verb → prohibitive particle</td>
</tr>
</tbody>
</table>

### Table F.3: Dependency Relations for Phrases and Clauses

<table>
<thead>
<tr>
<th>Relation</th>
<th>Arabic Name</th>
<th>Dependency</th>
<th>Dependent → Head</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>gen</strong></td>
<td>جار و مجرور</td>
<td>Preposition phrase</td>
<td>preposition → noun</td>
</tr>
<tr>
<td><strong>link</strong></td>
<td>منطق</td>
<td>PP attachment</td>
<td>PP phrase → verb or noun</td>
</tr>
<tr>
<td><strong>conj</strong></td>
<td>معروف</td>
<td>Coordinating conjunction</td>
<td>second phrase → first phrase</td>
</tr>
<tr>
<td><strong>sub</strong></td>
<td>مصلة</td>
<td>Subordinate clause</td>
<td>subordinate clause → particle</td>
</tr>
<tr>
<td><strong>cond</strong></td>
<td>شرط</td>
<td>Condition</td>
<td>condition → conditional particle</td>
</tr>
<tr>
<td><strong>rslt</strong></td>
<td>جواب شرط</td>
<td>Result</td>
<td>result → conditional particle</td>
</tr>
</tbody>
</table>
Table F.4: Dependency Relations for Adverbial Expressions

<table>
<thead>
<tr>
<th>Relation</th>
<th>Arabic Name</th>
<th>Dependency</th>
<th>Dependent → Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>circ</td>
<td>حال</td>
<td>Circumstantial accusative</td>
<td>accusative → verb</td>
</tr>
<tr>
<td>cog</td>
<td>مفعل مطلق</td>
<td>Cognate accusative</td>
<td>accusative → verb</td>
</tr>
<tr>
<td>prp</td>
<td>المفعول لأجله</td>
<td>Accusative of purpose</td>
<td>accusative → verb</td>
</tr>
<tr>
<td>com</td>
<td>المفعول معه</td>
<td>Comitative object</td>
<td>accusative → verb</td>
</tr>
</tbody>
</table>

Table F.5: Dependency Relations for Particles

<table>
<thead>
<tr>
<th>Relation</th>
<th>Arabic Name</th>
<th>Dependency</th>
<th>Dependent → Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>emph</td>
<td>توكيد</td>
<td>Emphasis</td>
<td>verb → emphatic particle</td>
</tr>
<tr>
<td>intg</td>
<td>استفسام</td>
<td>Interrogation</td>
<td>verb → interrogative particle</td>
</tr>
<tr>
<td>neg</td>
<td>نفي</td>
<td>Negation</td>
<td>imperfect verb → negative particle</td>
</tr>
<tr>
<td>fut</td>
<td>استقبال</td>
<td>Future</td>
<td>imperfect verb → future particle</td>
</tr>
<tr>
<td>voc</td>
<td>منادي</td>
<td>Vocative</td>
<td>noun → vocative particle</td>
</tr>
<tr>
<td>exp</td>
<td>مستثنين</td>
<td>Exceptive</td>
<td>noun → exceptive particle</td>
</tr>
<tr>
<td>res</td>
<td>حصر</td>
<td>Restriction</td>
<td>noun → restriction particle</td>
</tr>
<tr>
<td>avr</td>
<td>ردع</td>
<td>Aversion</td>
<td>dependent → aversion particle</td>
</tr>
<tr>
<td>cert</td>
<td>حقيق</td>
<td>Certainty</td>
<td>dependent → particle of certainty</td>
</tr>
<tr>
<td>ret</td>
<td>اضراب</td>
<td>Retraction</td>
<td>dependent → retraction particle</td>
</tr>
<tr>
<td>prev</td>
<td>كاف</td>
<td>Preventive</td>
<td>preventive particle → accusative particle</td>
</tr>
<tr>
<td>ans</td>
<td>جواب</td>
<td>Answer</td>
<td>dependent → answer particle</td>
</tr>
<tr>
<td>inc</td>
<td>ابتداء</td>
<td>Inceptive</td>
<td>dependent → inceptive particle</td>
</tr>
<tr>
<td>sur</td>
<td>فجأة</td>
<td>Surprise</td>
<td>dependent → surprise particle</td>
</tr>
<tr>
<td>sup</td>
<td>زائد</td>
<td>Supplemental</td>
<td>dependent → supplemental particle</td>
</tr>
<tr>
<td>exh</td>
<td>حثتفض</td>
<td>Exhortation</td>
<td>dependent → exhortation particle</td>
</tr>
<tr>
<td>exl</td>
<td>تفصيل</td>
<td>Explanation</td>
<td>dependent → explanation particle</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Arabic Word</td>
<td>English Word</td>
<td>Role</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>--------------</td>
<td>------</td>
</tr>
<tr>
<td>eq</td>
<td>تسوية</td>
<td>Equalization</td>
<td>verb → equalization particle</td>
</tr>
<tr>
<td>caus</td>
<td>سببية</td>
<td>Cause</td>
<td>imperfect verb → particle of cause</td>
</tr>
<tr>
<td>amd</td>
<td>استدراك</td>
<td>Amendment</td>
<td>dependent → amendment particle</td>
</tr>
<tr>
<td>int</td>
<td>تفسير</td>
<td>Interpretation</td>
<td>dependent → particle of interpretation</td>
</tr>
</tbody>
</table>