Aspect-Based Sentiment Analysis for Government Smart Applications Customers' Reviews

تحليل المشاعر المستند على السمات الخاصة بتعليقات المتعاملين على تطبيقات الحكومية الذكية

by

OMAR HARB ABDELKARIM ALQARYOUTI

A dissertation submitted in fulfilment of the requirements for the degree of

INFORMATICS (KNOWLEDGE AND DATA MANAGEMENT)
at
The British University in Dubai

Prof. Khaled Shaalan
July 2017

DECLARATION

I warrant that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.

I understand that a copy of my research will be deposited in the University Library for permanent retention.

I hereby agree that the material mentioned above for which I am author and copyright holder may be copied and distributed by The British University in Dubai for the purposes of research, private study or education and that The British University in Dubai may recover from purchasers the costs incurred in such copying and distribution, where appropriate.

I understand that The British University in Dubai may make a digital copy available in the institutional repository.

I understand that I may apply to the University to retain the right to withhold or to restrict access to my thesis for a period which shall not normally exceed four calendar years from the congregation at which the degree is conferred, the length of the period to be specified in the application, together with the precise reasons for making that application.

____

Signature of the student

Abstract

Nowadays, sharing opinions has been made easier with the evolvement of Web 2.0. People can share their opinions on their daily activities and consider others' opinions to decide whether to buy a product or install an app or use a service. Therefore, the public opinion on the web has become a norm in the modern world. Government agencies and business owners are keen to understand the publics' opinions towards their services and products. This is a key input for these organizations decision making process in terms of understanding the customers' needs in order to enhance the product or improve the service or introduce new features. This dissertation presents a holistic review on a variety of recent articles that commences with a background on Sentiment Analysis (SA) as well as it touches on numerous SA techniques, issues, challenges and real-life applications with focus on governmental services and smart apps. In this study, the government smart applications aspects that can be used in aspect-based SA were defined based on written standards with emphasis on customer experience as an important aspect. The proposed aspects include User Interface, User Experience, Functionality and Performance, Security, as well as Support and Updates. For studying SA of government smart applications customers' reviews, a novel domain-specific annotated dataset has been constructed. It involves government apps in the United Arab Emirates (UAE) as well as its corresponding aspects terms and opinion lexicons. This was done with the help of a proposed Government Apps Reviews Sentiment Analyser (GARSA) which is a responsive web tool that we have developed in order to facilitate the annotation process in a flexible, organized, efficient and tracked manner. Aspect-based SA is considered as one of the challenging tasks in SA. In this regard, an integrated lexicon and rule-based approach was employed to extract explicit and implicit aspects and their sentiment classification. This model utilized the manually generated lexicons in this dissertation with hybrid rules to handle some of the key challenges in aspect-based SA in particular and SA in general. This approach reported high performance results through an integrated lexicon and rule-based model. The approach confirmed that integrating sentiment and aspects lexicons with various rules settings that handle various challenges in SA such as handling negation, intensification, downtoners, repeated characters and special cases of negation-opinion rules outperformed the lexicon baseline and other rules combinations.

Keywords: Aspect-Based Sentiment Analysis; App Store Reviews; Aspect Extraction; Building Dataset; Sentiment Classification; Government Mobile Apps;

ملخص

في الوقت الحاضر، أصبح تبادل الآراء أسهل مع تطور ويب 2.0. يمكن للناس تبادل آرائهم حول أنشطتهم اليومية والنظر في آراء الآخرين لاتخاذ القرار المناسب لشراء منتج أو تثبيت التطبيق أو استخدام الخدمة. لذلك، أصبح الرأي العام على شبكة الإنترنت عادة في العالم الحديث. تحرص الجهات الحكومية وأصحاب الأعمال على فهم آراء الجمهور تجاه خدماتهم ومنتجاتهم. هذا هو المدخل الرئيسي لعملية اتخاذ قرارات هذه المنظمات من حيث فهم احتياجات العملاء من أجل تعزيز المنتج أو تحسين الخدمة أو إدخال ميزات جديدة. تقدم هذه الأطروحة استعراضا شموليا على مجموعة متنوعة من المقالات الحديثة التي تبدأ بإعطاء خلفية عن تحليل المشاعر فضلا عن أنها تتطرق إلى العديد من التقنيات في تحليل المشاعر والقضايا والتحديات والتطبيقات من واقع الحياة مع التركيز على الخدمات الحكومية والتطبيقات الذكية. في هذه الدراسة، تم تعريف سمات التطبيقات الذكية الحكومية التي يمكن استخدامها في تحليل المشاعر المعتمد على الخصائص بناء على معايير مكتوبة ومعرفة مع التركيز على تجربة المتعاملين باعتبارها جانبا هاما. وتشمل الجوانب المقترحة على خمس سمات وهي واجهة التطبيق للمستخدم، وتجربة المستخدم، والوظائف والأداء، والسرية، والدعم والتحديثات. بالإضافة إلى ذلك، تم إنتاج حزمة بيانات مفهرسة وفريدة بالإضافة إلى معجم الآراء والمصطلحات الخاصة بخواص التطبيقات والخدمات. كما يشار أن حزمة البيانات تتخصص بنطاق الخدمات والتطبيقات الحكومية والمتعلقة بعدد من التطبيقات الحكومية في دولة الإمارات العربية المتحدة. يذكر أنه قد تم ذلك بمساعدة الأداة التي تم بناؤها وهي محلل التعليقات الخاص بالتطبيقات الحكومية (جارسا) والذي يعد أداة ويب سريعة الاستجابة تم تطويرها لتسهيل عملية التعليقات التوضيحية بطريقة مرنة ومنظمة وفعالة وقابلة للمتابعة. ويعتبر تحليل المشاعر المعتمد على بالخصائص والميزات واحدة من المهام الصعبة في تحليل المشاعر. وفي هذا الصدد، تم استخدام نهج متكامل قائم على المعجم والقواعد لاستخراج الميزات والخصائص الصريحة والضمنية فضلا عن تصنيف المشاعر لهذه الجوانب. استخدم هذا النموذج المعجم الذي تم إنشاؤه يدويا في هذه الأطروحة مع القواعد الهجينة للتعامل مع بعض التحديات الرئيسية في تحليل المشاعر للسمات على وجه الخصوص وتحليل المشاعر بشكل عام. وقدم هذا النهج نتائج عالية الأداء من خلال نموذج معجمي متكامل يستند إلى قواعد معرفة. كما أكد هذا النهج أن دمج المفردات المعجمية لكل من مفردات الآراء والسمات مع مختلف القواعد الخاصة والتي تتعامل مع مختلف التحديات في تحليل المشاعر مثل التعامل مع النفي والتشديد والتخفيف والحروف المتكررة والحالات الخاصة من قواعد النفي و الرأي تفوق في الأداء على خط الأساس المعجمي.

Dedication

I dedicate my dissertation work to my parents, my sisters, my brothers, and my whole family. I would like to thank them for their love, unlimited support, and for their prayers. I also dedicate this dissertation to my friends who have supported me throughout the process. I will always appreciate all they have done.

I also dedicate it for my beloved wife and motivator. A special feeling of gratitude to her for making this work possible as well as her complete support.

Acknowledgement

At the beginning, I would like to thank Allah for the countless blessings and for providing me with the strength and patience to complete this dissertation. I hope that I have provided something useful that could contribute for science and society.

I would like to express my sincere gratitude to my supervisor Prof. Khaled Shaalan for the continuous support of my MSc study and related research, for his patience, motivation, and immense knowledge. His guidance helped during the time of research and writing of this dissertation. I could not have imagined having a better supervisor and mentor for my MSc study.

Special thanks go to my beloved brother Dr. Marwan Alqaryouti who thankfully supported my work by providing linguistic advices and guidance in formulating the team of linguistic specialists that consists of English students who thankfully assisted in the manual annotation process for building the dataset.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviations	**Description**
SA	Sentiment Analysis
OM	Opinion Mining
GARSA	Government Apps Reviews Sentiment Analyser
UAE	United Arab Emirates
SVM	Support Vector Machine
NB	Naïve Bayes
TF/IDF	Term Frequency / Inverted Document Frequency
NLP	Natural Language Processing
CL	Computational Linguistics
ML	Machine Learning
SO	Semantic Orientation
IT	Information Technology
POS	Part of Speech
LDA	Latent Dirichlet Allotiment
AKL	Automated Knowledge
UFL-LDA	Unified Fine-grained LDA
FL	Figurative Language
ASP	Answer Set Programming
IMDB	Internet Movie Database
IG	Information Gain
CHI	Chi-square

MI	Mutual Information
DF	Document Frequency
DP	Double Propagation
IAC	Implicit Aspect Clues
SWEM	Smart Website Excellence Model
CNN	Convolutional Neural Network

1. Chapter One: Introduction

Sentiment Analysis (SA), also known as Opinion Mining (OM), is an important task that combines Natural Language Processing (NLP) along with rule-based methods in order to computationally identify subjectivity in textual sources. This can help people in taking decisions through understanding the publics' thoughts and feelings towards a certain topic. This chapter points out the emergence of social media and smart apps and the rapid growth and reachability across the globe. Furthermore, how the government and private sectors are considering these factors in understanding their consumers. This section also introduces how Web 2.0 has changed the way the people interact as well as defines the SA as a trustworthy source that helps in taking decisions in various areas.

1.1. Social Media and Smart Apps

Nowadays, the social media and web are gaining increasing attention from both government and private sectors. Social media has become the contemporary trend in the industry of Information Technology (IT) and performs as powerful community for sharing content (Hasbullah et al. 2016; Al Suwaidi, Soomro & Shaalan 2016). Also, Alqaryouti, Alqudah and Shaalan (2016) stated that social network analysis is one of the important tools and techniques for sharing knowledge. The importance of the social media comes from internet population that has reached to 3.4 billion users and the massive time that the people spend on their smart devices as well as their computers. In 2015, Deloitte's Mobile Consumer Survey showed that 81% of respondents check their mobile phones within one hour from the time they wake up and 50% of them within five minutes. Furthermore, about 72% of the respondents said that they check their phones within one hour before going to bed every night. As per Zephoria Inc. (2016), the average time that the Facebook user spends with each visit is twenty minutes. Additionally, the Facebook users post 510 comments, update 293 thousand statuses and upload 136 thousand photos every minute. Similarly, Twitter users post over 350 thousand tweets per minute according to Internet Live Stats (2016).

Fu et al. (2013) have pointed out that the increasing proliferation of smart devices has led to rapid evolution of the app stores. Thus, the smart devices engagement level for users has massively increased and; hence, led to attract various sectors to take into consideration the users necessities and understand their applications viewpoints. According to the statistics portal Statista.com (2016), the number of smart apps offered in the top two app stores, Google Play Store and Apple App Store, have reached more than 4.2 million applications as of June, 2016. Figure 1 illustrates the number of applications in the well-known app stores. Statista.com has also published a forecast for the number of smart devices applications downloads from 2009 to 2016. In this forecast, the number of applications downloads in 2009 was around 2.52 billion and has reached to around 224.8 billion in 2016. These statistics show how people are engaged to their smart devices so they can stay connected to the outside world.

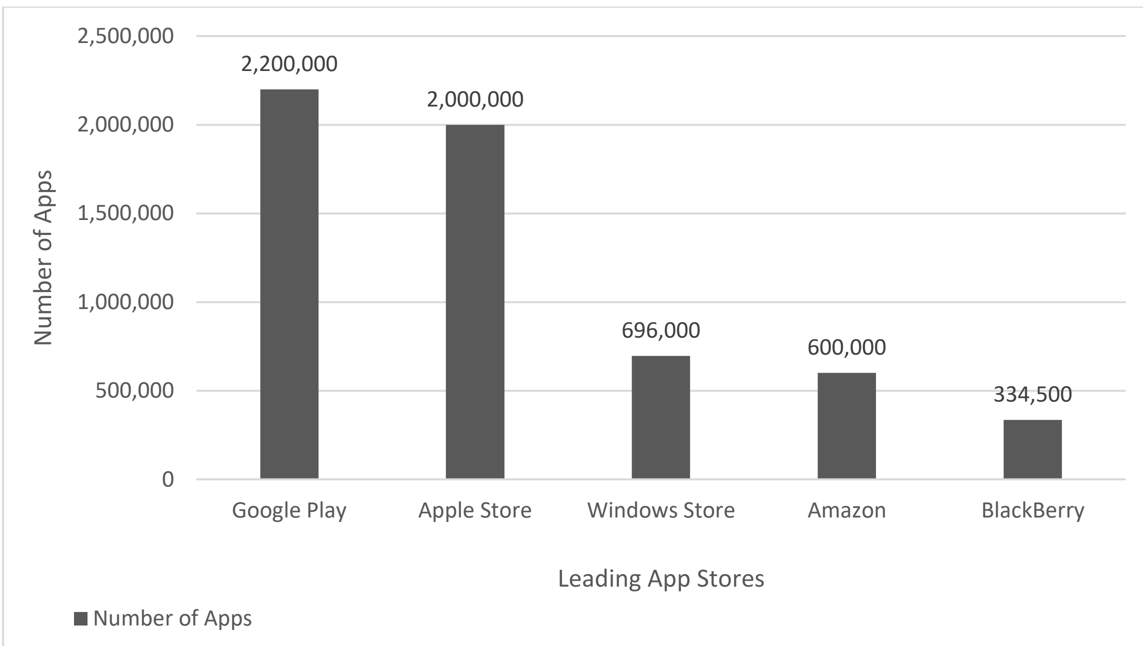

Figure 1: *Number of Apps in Leading App Stores*

In the last decade, both governmental and private sector organizations have steadily directed their focus to the social media in order to study their customers' feedbacks and needs due to its huge reach. For instance, a post on product page on Facebook with thousands of likes or followers reaches them momently and they can express their opinion or feedback towards this post. The follower opinion or feedback may also reach the friends of the followers and so on. This shows how a post may rapidly spread and how people interacts through it. Similarly, as explained by Asur and Huberman (2010), when a Twitter's user posts a status update, also known as tweet, it is made available on the user's profile page as well as on the user's followers feeds page. These tweets can be circulated by followers throughout the Twitter community using the retweet functionality.

Social media websites and applications allow users to interact with each other and stay connected to the outside world, express their feelings, communicate with their relatives and friends, track their role model's daily activities, follow their favourite news pages, stay tuned with the latest technologies and products, find out the latest offered services from government departments and more (Salloum et. al 2017). The users can speak out their opinions and thoughts through writing a review on an app or service or website, posting comments to certain news article, election, football match result, product, etc. This huge number of posts on various social channels such as Facebook, Twitter, Google Plus, YouTube, Instagram, LinkedIn as well as the apps reviews clearly illustrate how information is rapidly growing and generating massive amount of textual information database. These textual information are mostly of unstructured nature as it lakes of proper spelling, grammar, structure and uses special abbreviations that social network users usually understand. This highlights numerous ambiguities such as lexical, syntactic and semantics which complicates the automatic data extraction from such sources.

1.2. Web 2.0 and Dynamic Content Generation

In 2001, the field of social networking experienced a breakthrough, following the introduction of Web 2.0. The online technology developed communication between parties across the world due to increased interactivity and better channels of communication. Web 2.0 is the second phase of internet evolution and has transformed the web content from static (read only) to dynamic content (read-write). Furthermore, it provides greater collaboration and interaction between users compared to previous version of web. Previously, the users were only able to view or download the desired content, while in web 2.0 users are able to interact, share inputs and content through variety of channels such as blogs, peer-to-peer networks, forums, social networks, microblogging, and wikis, among others. This led to enhance the efficiency of social media activities and communication between users (O'Reilly 2007). Zabin and Jefferies (2008) pointed out the high impact of the Web 2.0 blast on companies and organizations. With Web 2.0, consumers are influenced by the opinions of others on a service or product whether it is positive or negative, so the companies can react accordingly to resolve any consequences and maintain the company's image.

1.3. Government Attention

With the advance of technology and the dramatic increase of smart phones users, the provisioning of public sector services has evolved from the need of face-to-face interaction into offering powerful electronic and smart channels for providing services to the public. Gil-Garcia and Martinez-Moyano (2007) stated that with these new trends, governments started adopting latest smart information and technological tools in their businesses and operations. Moreover, Barile and Polese (2010) pointed out that the focus of government organizations business has changed towards offering a competitive innovative and quality integrated services as a result of globalization that produced a connected world in all aspects. Additionally, Gil-Garcia, Helbig and Ojo (2014) described the usage of smart government as the joined activities between the developing technologies and innovative strategies in order to produce resilient environment.

An example that illustrates the government leadership attention to adopting smart technologies is when His Highness Shaikh Mohammad bin Rashid Al Maktoum, Vice-President and Prime Minister of the UAE and Ruler of Dubai, announced the launch of the "Smart Government" and directed all parties for immediate commencement of all steps required to transform eGovernment services into smart services. According to his highness' vision, all members of the community can benefit from the government services through smart mobile phones around the clock (GulfNews 2013a). In this regard, Shaikh Hamdan bin Mohammad bin Rashid Al Maktoum, Crown Prince of Dubai and Chairman of Dubai Executive Council said:

> "We are looking forward to being a centre of government work where excellence is no longer our goal, but creating pioneering work through a paradigm shift in Dubai Government services at all levels in line with the directives of the UAE's supreme leadership and making unprecedented historical achievements for the public service in a manner that satisfies the public and in harmony with the requirements of the new stage" (GulfNews 2013b).

This smart transformation highlights the importance of keeping pace with the rapid development as well as the needs and expectations of citizens helping them achieve their happiness.

In the following year, in October 2014, His Highness Shaikh Mohammad bin Rashid, has launched an exceptional initiative to measure the publics' happiness with government services offered to them so they can observe and monitor the consumers' happiness towards the government services. The consumers are now able to rate their happiness and send their feedback and comments regarding the services. This will help in the process of development and improvement of these services and as a result it will reflect on the society happiness as well as increase the competitiveness among governmental departments towards being global leaders in all areas (KhaleejTimes 2014). This unique initiative shows the importance of analysing and understanding the comments, feedbacks and inputs from customers' perspectives against the governmental services as a major factor for future improvements that exceed their expectations.

Government and business organizations invest massive amount of resources to capture their customers' feedbacks and opinions. SA is very important for organizations as they tend to understand the level of their services from the customers' point of view through capturing and understanding opinions yet take it into consideration in their services development and improvements. Whereas, individuals consider the others opinions to take decision in buying a product or investing in a business. A study on Malaysian government by Hasbullah et. al. (2016) pointed out the significance SA in social media as a driver for understanding the citizens needs and wants closely. Thus, government services will be customized and prioritized to meet the citizens' expectations rather than depending on the capability level of that government organization.

1.4. Sentiment Analysis

SA is defined as the computational study that analyses the individuals' attitudes, excitements, emotions, expressions, viewpoints and opinions towards certain entity. SA is one of the most extensive application of text analytics. SA focuses on analysing sentiment of variety of text sources, such as government services, smart applications, corporate surveys, product reviews, movie reviews, restaurant reviews, among others. SA is used to extract the customers' opinion from their text reviews about a service or product. Public and private organizations are keen to discover the publics' opinions about their services and products (Siqueira & Barros 2010). SA field is grapping the researchers' attention and excitement because of the extending real-world business and social applications. Finding out peoples' opinions is considered as a major factor for people to take better decisions. It is important to discover what people according to their experience about particular product or service. As stated in section 1.2, web 2.0 has enabled the users to express their opinions and allow the user generated content to rapidly grow on the internet.

Opinions in social networks play an integral role in day to day lives. The comments and reviews in social networks and web are considered as one of the most trustworthy and powerful sources of information. It is a fact that analysing social media network posts and web content is very important (Ray & Shaalan 2015). SA can help in understanding the people thoughts and feelings towards certain topic, deciding of a movie review, finding out the people thoughts of a candidate, predicting stock markets, extracting aspects or attributes about services or products as well as automatically deciding the important aspects and summarizing the various positive and negative reviews according to these aspects or attributes, etc.

1.5. Research Motivation

As stated in the previous sections, the extensive momently populated amount of information has led to the growing attention by both public and private sectors to study and consider the customers' feedback in their services and products continuous development. Government organizations are looking for customer centric approaches to find the consumers opinions toward their services. Conversely, customers try to find out the existing users' opinions about a particular service before using it. "What do other people think about it?" is an important question that always comes in mind before taking any decision.

Prior to this study, a survey was conducted to gain insights into peoples' considerations towards using government mobile applications and to find out what affects their decision in downloading and using a government mobile application. The survey results showed the significant impact of star rating of the application on the decision of downloading or using it. 63.9% of the participants stated that their decision was affected by the rating of the application.

And even a higher percentage (67%) of the participants confirmed that the app reviews influenced their decision.

The aim of this dissertation is to come up with an integrated novel aspect-based model for the government apps. This model may lead to more in-depth understanding of the customers' needs and expectations through looking into the app reviews from different angles in addition to previous models of assessing and evaluating the government apps based on written standards.

Additionally, the research aims to produce a new domain specific annotated dataset that involves government apps in the UAE as well as its corresponding aspects terms and opinion lexicons. This process will allow researchers to apply various techniques in analysing customers' sentiments and extracting the various aspects.

1.6. Research Questions

This dissertation attempts to answer the following research questions:

- **Research Question 1**: What are the government smart applications aspects that can be used in aspect-based SA?
- **Research Question 2**: How can we produce a domain-specific annotated sentiment dataset and lexicon that targets the government mobile apps?
- **Research Question 3**: How can we improve the performance of aspect extraction and sentiment classification through an integrated lexicon and rule-based model?

1.7. Research Methodology

Prior to this dissertation, a survey was conducted to gather information about customers' concerns about the usage of government applications. The survey aims to attain insights into peoples' considerations towards using government mobile applications. The analysis of the survey clarifies the overall interest to verify the quality of the government mobile app and gains the trust before taking the decision to install and use it. This verification is achieved by either looking at the star rating of the app or by analysing the users' reviews towards it.

This dissertation follows a defined methodology to answer the research questions. In this research, the review of various areas in SA was considered to attain an in-depth understanding of basics, SA approaches, challenges, and the government viewpoints towards their customers. This helps to formulate a comprehensive and holistic view in the various SA areas that eases the application of desired processes in later stages. As summarized in Figure 2, the research methodology comprises of two phases with several steps in each one. These phases are discussed in details in chapters three and four. The first phase is concerned with defining aspects of the dataset and using it to annotate the dataset. The steps of this phase can be summarized in the following steps:

1. **Identify** the target government applications as source of customers' reviews.
2. **Collect** the reviews for the selected target applications.
3. **Define** the aspects that need to be extracted and classified in aspect-based SA in the government mobile applications domain.
4. **Analyse and Filter** the reviews that will be the scope for this study.
5. **Annotate and Review** the reviews with the help of GARSA tool to extract aspects and opinion words from the reviews. This includes the automatic assignment of sentiment rating to the extracted opinion words using SentiWordNet.
6. **Verify** the annotated reviews and extracted aspects and opinions as well as the assigned sentiment polarities in order to produce the dataset and corresponding lexicons.

The second phase is concerned with building an integrated lexicon and rule-based model for aspect extraction and opinion classification. The steps of this phase can be summarized in the following steps:

7. **Pre-processing** for the source reviews.
8. **Extract** explicit and implicit aspects.
9. **Assign** the scoring aspect sentiments.
10. **Aggregate** and summarize the results.

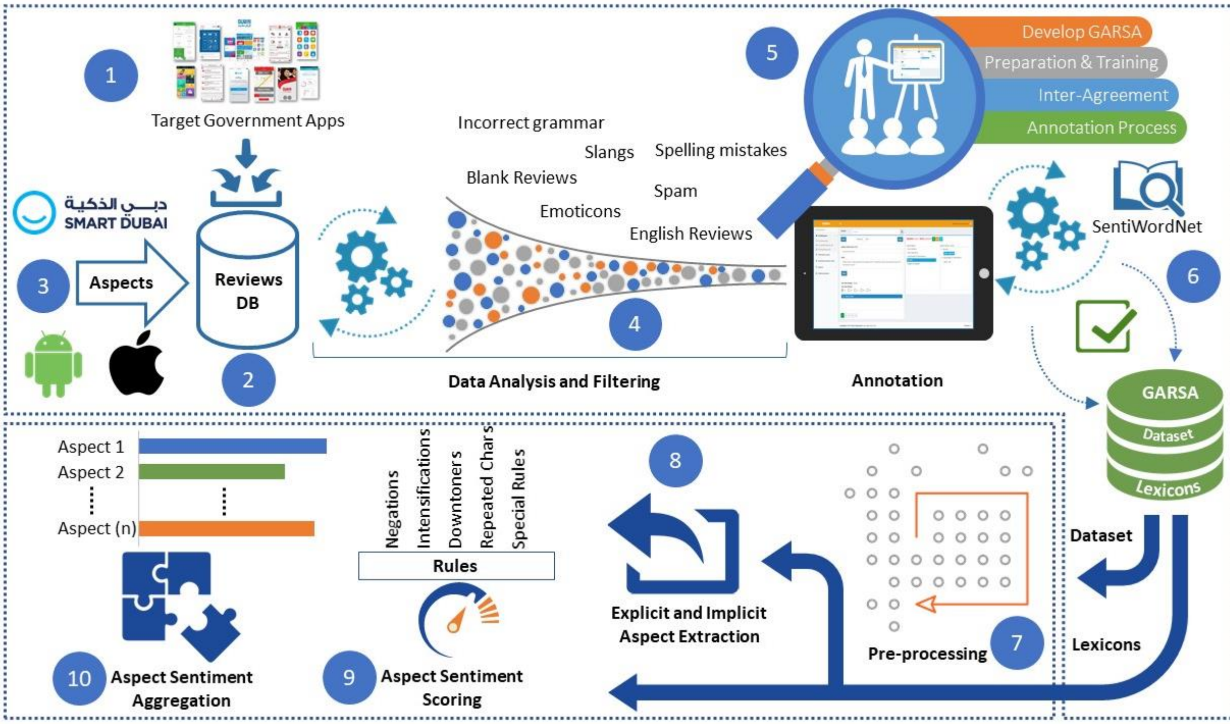

Figure 2: *Research Methodology Framework*

1.8. Dissertation Structure

This dissertation is divided into six chapters. The next chapter is the literature review which contains credible resources to determine the findings of previous studies on SA. These resources are used to develop information about basic terms in sentiment analysis and its applications. The literature review provides a background information on SA that covers the SA basics and levels of sentiment classification. Additionally, the literature review dives into the SA approaches and techniques details touching on how and when each of the techniques can be used to fulfil SA tasks. The dissertation then discusses the previous work done on aspect-based SA field as well as constructing resources for SA and presents the significant contributions in improving various tasks. Lastly, it discusses the government smart applications explosion and its importance to increase the customers' happiness with its major features to come up with powerful and highly ranked applications.

The third chapter demonstrates the work that has been done in order to produce a unique dataset that consists of government mobile apps domain aspects and opinion words. Additionally, it explains the methodology that was carried out to assign the sentiment scores to opinion words and build the desired lexicons. Chapter four summarizes the methodology used to build an integrated lexicon and rule-based SA model to extract government apps reviews aspects, classify the corresponding sentiments, aggregate the extracted results and summarize them. In this chapter, the performance evaluation results are illustrated for both aspect extraction and sentiment classification. After that, the answers for the research questions are provided in chapter five. The last chapter summarizes all parts of the dissertation and provides details information about the challenges and potential areas of improvements. Additionally, it discusses the future prospects in the government apps domain in the various areas in SA.

2. Chapter Two: Literature Review

Due to the developed information technology today, majority of businesses and government institutions make their decisions based on the data collected through social media platforms. It is thus crucial to enhance the accurate collection and analysis of such data because it has great social, political and economic implications.

The dissertation presents the various areas of SA. It gives a comprehensive background of the SA that helps in providing a solid background about SA. This also illustrates a review of various areas in SA was considered to attain an in-depth understanding of basics, SA approaches, challenges, and the government viewpoints towards their customers. This helps in formulating a comprehensive and holistic view in the various SA areas that eases the application of desired processes in later stages.

2.1. Background

2.1.1. Opinions in Sentiment Analysis

An opinion can be defined as a position or a standpoint that one has concerning a particular issue, informed by their beliefs and attitudes. Khan et al. (2014) defined opinions as private statements expressed by individuals, arguing that they represent their sentiments, ideas and beliefs about a specific topic. The authors explained that every opinion has a source, mainly referred to as the opinion holder. The other aspect of opinions is the object or the topic that leads the opinion holder to express their opinion. The last component of opinions as explained by the authors is the opinion holder's view, which is influenced by their sentiments about the object. Himmat and Salim (2014) stated that on opinion sentiment, they have defined an opinion as quadruple of four essential components as follows: (*t, s, h, dt*), where *t* represents the target of the sentiment or opinion, *s* represents the sentiment towards the target, *h* represents the opinion holder and *t* represents the date-time stamp of the opinion. The authors explained that the opinion holder has a sentiment on the topic, which determines the sentiment they associate through their view on a particular date and time. Similarly, the study by Liu (2012) provided important insights about the definition of opinions. The authors stated that an opinion has a target, aspect or feature, its holder, their sentiment and the time when it was expressed. Liu and Zhang emphasized the importance of considering the time when an opinion was expressed because the sentiments of people about particular topics are greatly influenced by the conditions during a given time. Additionally, Liu and Zhang (2012) pointed out that, when studying sentiment analysis and opinion mining, the main focus is on opinions that express sentiments, whether positive or negative. Figure 3 illustrates an example of a sample review with quintuple sentiment components extracted on one of the government apps.

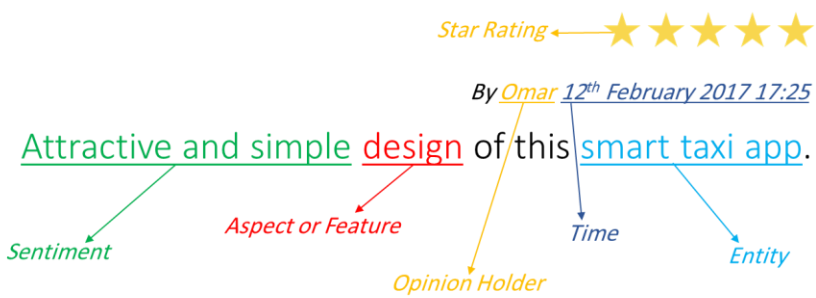

Figure 3: *Sentiment Quintuple Example of a Government App Review*

In sentiment analysis and opinion mining, one must understand the varying meanings of opinions and facts. According to Pang and Lee (2008), the primary distinction between opinions and facts is that the former is based on an individual's viewpoint while the latter is based on statistics. This implies that while one can verify the accuracy of facts, the same cannot apply for opinion. This is mainly because opinions are based on peoples' views and one can simply not be honest about their experience with something. However, it is important to keep in mind that even sentences without sentiment words can indicate opinions. For example, the sentence "This application takes long time to load" implies a negative opinion about the service. This is an example of an implicit opinion, which is an objective statement that usually expresses a fact (Liu 2012). Objective sentences may or may not contain sentiments. Another type of opinions is the explicit opinion, which is a subjective statement. For example, the sentence "This application is great" is an explicit opinion (Liu 2012). Determining whether the sentence is objective or subjective, and whether it has a sentiment or not is called subjectivity classification. This is a challenge facing researchers in the field of opinion mining (Liu & Zhang 2012).

An example to illustrate the definition of opinion can be the following review of an application:

"(1) I have been using this transportation app for more than three months. (2) This app is useful when it works fine. (3) Sometimes, the app freezes and keeps logging me out. (4) However, on my wife's phone, the app works smoothly. (5) Also, my wife thinks that the design of the app is attractive. (6) She thinks it's the best transportation app available."

There are several opinions in this review. The sentence (1) is just a fact and does not contain an opinion. Sentences (2), (4) and (5) express positive opinions. Sentence (3) expresses a negative opinion. The target of the sentences (2), (3), (4) and (6) is the app in general. While the target of sentence (5) is the design of the app. The holder of the opinion in the sentences (2) and (3) is the author of the review, while the holder of the opinion in the sentences (4), (5) and (6) is the "wife" of the author. Finally, the sentences (2), (3), (4) and (5) are implicit opinions, while sentence (6) is an explicit one.

The sentiment polarity (or orientation) represents the extent to which it is positive, negative or neutral. Positive opinions are characterized by the presence of words that have positive sentiments such as "*good*", "*awesome*" and "*great*". Negative opinions have more negative or negation phrases or words such as "*bad*", "*slow*" and "*useless*". Polarity of opinions can also be classified as neutral when the sentence does not contain an opinion or contains unrelated words (Ganeshbhai & Shah 2015). In the above example, sentences (2), (4), (5) and (6) are positive opinions that contain the positive keywords "*useful*", "*smoothly*", "*attractive*" and "*best*". On the other hand, sentence (3) is a negative opinion that contains the keyword "*freezes*".

2.1.2. The Levels of Classification

There are three different levels on where sentiments can be analysed; document level, sentence level, and aspect level. At document level, the whole review is considered as a basic information unit and is then classified into positive, negative or neutral. This process is known as document-level sentiment classification. However, for the document-level sentiment classification to be meaningful, it is essential to assume that the opinion holder and target are similar in the whole document. Similarly, sentences are regarded as short documents. Sentiment classification at sentence-level usually starts with subjectivity classification by classifying sentences into objective or subjective sentences. Then, sentences are categorized into positive, negative, or neutral opinions. Since document and sentence-based classification lack opinion target identification and the sentiment assigned to that target, they are considered insufficient for applications. At document-level, the sentiment is determined for the document as a whole, which clarifies the sentiment associated to the target. In this case, the opinion holder might have a positive opinion regarding the target (entity) but might not be satisfied with all the “aspects” of that target. To extract such information, aspect-based classification is used. Aspect-based analysis covers both entities and aspects. It does so by decomposing the entity into aspects (aspect extraction), then classify each aspect sentiment into positive, negative or neutral (aspect sentiment classification), and finally, summarize the results of the previous steps (Asghar et. al 2014).

2.1.3. Sentiment Analysis Considerations

As discussed in section 2.1.1, a sentence can be either objective or subjective. Objective sentences contain facts. These sentences can imply opinions which are called “implicit opinion” since they are implied from the desirable or undesirable facts (Pang & Lee 2008). On the other hand, subjective opinions can be in different forms such as opinions, accusations, desires, beliefs, scepticisms, and assumptions. However, not all subjective sentences necessary contains sentiments. For example, the sentence “I would like to download an application to help me pay the bills” is a subjective sentence that does not contain any sentiment.

Emotions are defined as the subjective feeling and thoughts of a person. Opinions are linked to the intensity of certain emotions such as joy and anger. It is essential to note that opinions and emotions are not equal. Opinions evaluation can be classified into two types: rational and emotional. Rational evaluation come from reasoning and solid beliefs while emotional evaluation come from emotional reactions connected to the person’s state of mind. Thus, rational opinions do not express emotions while emotional sentences may not contain an opinion at all (Liu 2012).

Another aspect in sentiment analysis that can be taken into consideration is the perspective in which an opinion can be looked at; the author (opinion holder) or the reader of the opinion. Researchers either ignore this aspect or presume the standing point to be the consumers in general (Liu 2012). For example, the sentence “This application loads slower on an iPhone than on an Android device” can be considered a negative opinion for iPhone users and a positive opinion for Android users.

2.1.4. Data Preparation for Sentiment Analysis

SA adopts several techniques and tasks from Natural Language Processing (NLP) and Computational Linguistics (CL). Due to the nature of the data being unstructured, various pre-processing tasks that can be decided according to the several factors such as domain, data source, SA task, language among others. The most common pre-processing steps are: Tokenization, POS tagging, stemming, normalization, filtering and removing stop words.

Tokenization is the essential process that precedes any text processing tasks. In this process, the text is split into segments that are called tokens. The tokens represent linguistic units of words, numbers, dates, punctuations, symbols, or other meaningful units. The tokenization typically takes place on the word level using tokenizers that relies on the white space characters as the simplest approach (for example: space and line break) or punctuation characters. The punctuation has reported some issues that may affect the tokenization performance such as the word clitics (for example: They're that represents They are) and multi token words (Abu Dhabi that represents a city in UAE).

POS tagging which is also known as lexical tags or morphological classes is the process of assigning the parts of speech to each word such as nouns, verbs, adjectives, adverbs, etc. This process provides a breakdown for the document structure as well as information about the words and words that are likely to occur near it.

Stemming is part of morphological analysis that represents the process of removing grammatical markings reduces the word to its root. For example: flexibility is stemmed to flexible and buggy is stemmed to bug.

Stop words are the words that are used very commonly and do not have any characteristics do not have impact to change the meaning of the text context. filtering these words has impact on the performance of the text processing tasks.

Further to the mentioned pre-processing tasks, text normalization is considered as one of the important pre-processing transformations in which the text is made consistent. The normalization may include converting text into lower or upper case, removing common words, removing stop words, error corrections and removing diacritics among other tasks.

2.1.5. Feature Extraction and Selection

Feature extraction for SA is important to figure out the significance of each feature type for SA. There are various types of feature extraction approaches that were used in the literature. The simplest approaches are using the unigram and bigram features. Unigram represents the bag-of-words as features that are extracted by removing unwanted spaces or characters. However, the bigram represents the features of every two adjacent words in the text.

Wawre and Deshmukh (2016) have adopted the standard features according to the frequency in which the features appear in the review in order to prepare the training data. Dang, Zhang and Chen (2010) have utilized content free, content specific and sentiment features in generating the features. Content free features has combined over 250 features that includes lexical, function words and punctuation marks. However, bigrams and unigrams were used to generate the content specific features with larger set of features. The third feature generation is the sentiment features method. This method was utilized in order to extract the adjectives, adverbs and verbs part of speech (POS) using Stanford POS tagger. Whereas, names were excluded as it mostly relies on context.

Feature selection is the process of selecting a set of appropriate features that can be used in model construction. Feature selection methods are important in order to produce simplified models, reduce the time required for training, avoid the curse of dimensionality and reduce overfitting that result in improving generalization. There were variety of feature selection methods in the literature such as Information Gain (IG) and Mutual Information (MI) (Manning, Raghvan & Schutze 2008). In addition to IG and MI, Agarwal and Mittal (2013) pointed out other feature selection methods which are chi-square (CHI) and document frequency (DF). The authors summarized that DF feature selection is used to construct the feature vector by calculating the most frequent terms that appear in the corpus.

2.2. Sentiment Analysis Approaches

SA has gained increasing attention by researchers around the world as it is considered as the base for decision making process for variety of applications. SA tends to identify the subjectivity or objectivity of certain text. Additionally, it tends to identify the polarity of this particular subjective text (Liu & Zhang 2012). SA techniques have been employed by several models that can be classified into two main approaches namely machine learning (ML) approaches for SA

(Pang & Lee 2008) and semantic orientation (SO) approaches for SA (Dang, Zhang & Chen 2010). Moreover, Medhat, Hassan and Korashy (2014) stated a third approach that can be also utilized in SA which is the hybrid approach that combines both ML and SO approaches. The hybrid approach is very common when sentiment lexicons are adopted with other techniques. Many researches over the globe have been performed to perceive sentiment from a given text (Liu 2012). Thus, there is still an enormous area for improving existing SA models and even more, to come up with new models. Figure 4 summarizes the SA approaches and techniques.

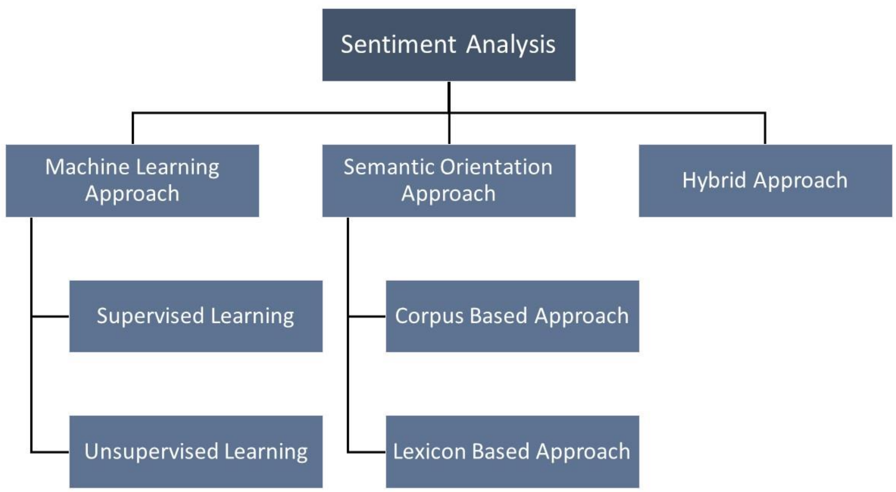

Figure 4: *Sentiment Analysis Approaches and Techniques*

ML approaches depend on the well-known algorithms being able to classify and resolve text into its corresponding classes. These approaches make use of syntactic as well as linguistic features such as term frequency, part of speech (POS), syntactic dependency and negations. The initial phase of ML approaches for SA is pre-processing the text followed by feature selection, feature weighting and finally applying the desired ML algorithm. There are two main ML techniques for text automated classification: Supervised ML and Unsupervised ML. In addition to these two approaches, there are reinforcement learning and semi-supervised learning approaches.

Supervised learning represents the ML algorithms that are trained on annotated sample of the data. This approach learns from the previously available data and is able to predict future unknown data. The main objective of the supervised learning is to produce a trained model that can predict the class of the newly unknown testing data samples.

On the other hand, the unsupervised learning represents ML algorithms that do not require annotated data to produce the trained model. The main objective is to extract significant patterns from existing date based on the data features, similarity and other data attributes without training on previously annotated dataset.

Wawre and Deshmukh (2016) proposed a model to classify movie textual reviews according to the sentiment presence in the context comparing the performance between two ML algorithms SVM and NB. The Internet Movies Database (IMDB) review dataset was used in this research. This dataset comprises of 1400 processed files that are divided according to their classification as "positive" and "negative" sentiments.

Prior to the data acquisition, the authors have performed text pre-processing to obtain the separated reviews with its assigned polarity. Then, they have converted the reviews into lowercase, removed control characters, numeric values and punctuations. After that, SVM and NB classification algorithms were adopted to evaluate the document based sentiment analysis. To utilize these algorithms, the authors have prepared the training data according to standard features denoting the frequency in which the features appear in the review. The results reported a better performance for NB over SVM giving higher accuracy of 65.57% and 45.71% repeatedly.

Weitzel, Prati, and Aguiar (2016) conducted experiments to address two of the main Figurative Language (FL) domains namely; irony and sarcasm in order to figure if the sentiment classification performance is affected by the FL. The authors have used a dataset that consists of 100,000 tweets in English language of equal number of sarcastic and non-sarcastic tweets. These tweets have been verified and accordingly the authors removed the tweets with private accounts. This ended up with around 82,000 tweets. After that, the Weitzel, Prati, and Aguiar followed the data collection with the data pre-processing. This includes tokenization using tweetokenize package built to extract tweets from twitter posts. This package consists of capabilities that can easily extract tweets features such as URLs, phone numbers and user names. These features were then replaced with general tags. The authors have removed the hashtags from the tweets as it could be an indicator for irony within the tweet. Furthermore, the researchers have adopted Gensim software package to replace slang expressions with the related meaning found in the dictionary whenever a token match is found. Then, both machine learning and lexicon based approaches have been used to assess the sarcastic identification performance. In order to perform these experiments, the authors have used scikit-learn package to apply both SVM and Logistic Regression algorithms. The TF-IDF representation as well as TF-IDF representation that is combined with the distributed vector representation vectors (Word2Vec available in Genism Package) were used as input for these algorithms. The results reported similar results for both SVM and Logistic Regression in TF-IDF and the combined TF-IDF and Word2Vec representations with around 81% accuracy with very slight differences in each combination. The authors have applied the Apriori algorithm considering the items as the words in the tweets in order to determine the words that negatively impact the performance of the approaches. Furthermore, the lexical based methods reported are poor results compared to ML methods being reasonably able to identify irony and sarcasm in twitter posts.

The semantic orientation of an opinion or sentiment word states that the word has a positive, negative or neutral. For instance, words such as "great", "excellent" and "delighted" denotes a positive semantic orientation. However, words such as "depressed", "bad" and "terrible" denotes a negative semantic orientation. According to literature, the semantic orientation approaches can be divided into two main approaches; corpus based and lexicon based approaches (Dang, Zhang & Chen 2010; Liu and Zhang 2012; Agarwal, Mittal & Sharma 2014).

There are three steps for SA using semantic orientation approaches as shown in Figure 5 namely; extracting sentiment features, determine the semantic polarity and aggregating the orientation for all features (Asghar et. al 2014).

Figure 5: *Phases in Semantic Orientation Approaches*

The first semantic orientation approach is the corpus based approach. In this approach, the polarity is calculated according to the co-occurrence syntactic patterns of the opinion words with respect to other seed words being positive or negative in a corpus (Agarwal, Mittal & Sharma 2014). Numerous methods described in the literature to determine the polarity of the opinion words. Hatzivassiloglou & McKeown (1997) proposed a novel approach for identifying the semantic orientation of words. The approach entails the development of a seed of opinion words of adjectives that express the polarity of phrases in documents. Moreover, it uses additional constraints such as conjunctions to identify additional opinion words of adjectives. For example, in this app review, “This app is complicated and slow”, if “complicated” is considered as negative, then it can be indicated that “slow” is also negative. The advantage of this approach is that it assists in resolving the problem of identifying the orientation of the opinion words according to the context.

Meanwhile, the lexicon based approaches which are also known as dictionary or knowledge based approaches employ the available lexicons such as WordNet that was developed by Miller (1995) and SentiWordNet that was developed by Esuli and Sebastiani (2006). One of the approaches that was used to build dictionaries starts with manual gathering of a slight set of opinion words that have its orientations known. Then, the set is expanded by finding and extracting the words synonyms and antonyms from WordNet. This process is done through iterations that ends when it does not return any new words (Kim & Hovy 2004; Hu & Liu 2004).

The hybrid approaches usually combine both ML and SO approaches. The hybrid approach is very common when sentiment lexicons are utilized with other techniques. Dang, Zhang and Chen (2010) have proposed enhanced method for sentiment classification by combining both ML and SO approaches. This lexicon enhanced approach has gone through three steps; acquiring the data, generating features and performance evaluation. The authors have selected the Support Vector Machine (SVM) classifier with 10-fold cross validation to evaluate the performance using the various generated feature sets. The authors stated that their experiment results outperformed for all products when the features using the three methods were combined and selected using Information Gain (IG) with average accuracy of 81.36%, precision of 81.66%, recall of 81.27% and F-measure of 81.51%.

2.3. Aspect-based Sentiment Analysis

As described in 2.1.2, aspect-based SA is one of the levels of sentiment classification. Similar to Asghar et. al (2014), Ganeshbhai and Shah (2015) confirmed that aspect-based phases SA comprises of aspect or feature extraction, sentiment polarity prediction and classification, and sentiment aggregation.

Aspect extraction phase represents the identifications of the review aspects are identified through the consumers comments in order to extract the desired aspects or features. It is essential to determine the entity to which the aspect belongs (Liu & Zhang 2012). In the following example: "*Dubai Now application design is amazing*", "*design*" represents an aspect of "*Dubai Now*" entity. And therefore, the sentiment reflects that aspect in particular. Next, the polarity prediction and classification takes place to decide if the aspect sentiment polarity denotes positive, negative or neutral orientation as well as its strength level. In the previous example, the word "*amazing*" denotes a positive sentiment polarity towards the "*design*" aspect. The last step is to summarize the results according to the extracted aspects and its corresponding classified polarity. Liu, Hu and Cheng (2005) demonstrated a prototype that allows the users to visualize to compare the pros and cons of products according to each product aspects. On the other hand, Carenini, Cheung and Pauls (2013) produced a traditional text-based aggregated opinions summary that provided an overall brief of peoples' thoughts about products. The drawback of the Cheung and Pauls's approach is that it only produces qualitative summary, compared with the quantitative summaries which provide an easy option to the users that is concise, analytical and graphically represented.

The aspects in SA can be classified into explicit and implicit aspects. In explicit aspects, the aspects terms in subjective sentences that clearly represent the opinion target. Whereas, in implicit aspects, the aspects terms are not clearly representing the opinion target and it is represented by the semantic of the sentence that represents the aspect implicitly (Hu and Liu 2004); (Poria et al. 2014); (Poria, Cambria & Gelbukh 2016); (Liu & Zhang 2012). For instance, in the previous sample review: "*Dubai Now application design is amazing*", the "*design*" denotes an explicit aspect with positive sentiment orientation "*amazing*". In contrast, in the following sample review: "*The application works as expected*", there is no clear explicit term that denotes explicit aspect. But, "*as expected*" implicitly denotes a positive sentiment orientation towards the application "Functionality and Performance" feature.

Aspect extraction is essential SA task that aims to extract sentiment targets from opinionated context. For example, the follow sample mobile app review: "*This app has an easy interface*" it aims to extract "*easy*" as the opinion in this sentence and its target "*interface*" which represents the aspect. Aspects are product or service attributes or characteristics or features. These aspects are very important in variety of SA applications (Liu & Zhang 2012).

The products and mobile smart applications features or aspects play an important role in SA. Therefore, aspects extraction in SA field is now becoming an active area of research as it is the most vital task in the aspect-based level. Sarawagi (2008) pointed out that aspect extraction is considered as part of the information extraction in which the target is to extract structured and useful information from various unstructured sources of textual data.

Researchers have reported numerous approaches and methodologies in order to extract aspects from textual resources. For instance, Mukherjee and Liu (2012) proposed a semi-supervised solution to cluster similar aspects in the same category by allowing the user to feed seed words into some defined categorized aspects. After that, the two proposed statistical models namely Seeded and Sentiment model (SAS), and improved SAS Maximum Entropy (ME-SAS) extract the aspects and cluster them into categories. While, Xianghua et al. (2013) suggested an unsupervised approach to determine the aspects and sentiments in Chinese social reviews. The authors have used Latent Dirichlet Allotiment (LDA) in social review to identify multi-aspect global topics. Then extract the local topic and the related sentiment according to a sliding window through the review text. The LDA trained model was used to identify the local topic aspect. However, to identify the associated sentiment to this particular aspect, Xianghua et al. have used HowNet lexicon. Their approach has achieved an improved accuracy in terms of SA as well as good topic partitioning. This also helped in identifying both sentiment associated and the multi-aspects topics at the same time. Chen, Mukherjee and Liu (2014) improved the LDA model and presented the Automated Knowledge LDA (AKL) which is a fully automatic approach that can use existing domain independent data to learn prior knowledge and identify new aspects. AKL approach was able to produce aspects and resolve issues related to wrong knowledge by adopting and enhancing the Gibbs sampler method. Similarly, Wang et al. (2014) used the LDA model to propose two semi-supervised topic models for extracting product aspects. Their approach used a dataset that is publicly available to extract aspects and its corresponding terms that are used to group the reviews (i.e. Topic Modelling). The first model is the Fine-grained Labelled LDA (FL-LDA) which was used with products reviews seeding aspects that assists in finding their related words. On the other hand, the second model is Unified Fine-grained Labelled-LDA (UFL-LDA) which was used to extract high-frequency words from unlabelled documents

Liu et al. (2013) used Answer Set Programming (ASP) to apply syntactical approach based aspect extraction. They have used a set of simple logic programming rules that resulted in a faster and more precise implementation. In their paper, they adopted a syntactical method to determine the general words using WordNet as well as double propagation (DP) approach to extend the set of the defined rules. The DP approach extracts aspects and opinion words through a set of opinion words seeds that will result into more aspects and opinion words. In later work, Liu et al. (2015) integrated both supervised and unsupervised domain independent automatic rule-based methods to improve DP. In their research, the DP method assumes that the opinion words will always have target. Therefore, there is a syntactic relation between the opinion word and the target within the same sentence. So, opinion words can be identified through the extracted aspects and in the same way, the aspects can be identified through the extracted opinion words. Moreover, both opinion words and aspects can be used to determine any further aspects and opinion words. This approach was able to select effective part of all rules with a better performance than the whole set.

Poria et al. (2014) asserted an unsupervised rule-based approach to obtain both explicit aspects as well as implicit aspects clues (IAC) from product reviews. While explicit aspects are extracted by identifying the words representing the opinion and the entity, implicit aspects should be implied. In the proposed approach, the authors used the IACs to identify the IACs in the review and plot them to the aspects they are signifying depending on common-sense knowledge and on the dependency structure of the sentence using WordNet and SenticNet. Samha, Li and Zhang (2014) developed a framework to extract aspects and opinions from reviews as well as group them into categories and summarize the results. The authors used frequent POS tags and rules in addition to opinion lexicon to identify aspects and opinion words.

In a more recent work, Poria, Cambria and Gelbukh (2016) claimed that they have conducted the first deep learning attempt in SA to extract aspects and to overcome the previous approaches that required many features in addition to the challenge of manual creation of linguistic patterns that relies on syntactical correctness. The researchers performed seven-layer deep convolutional neural network (CNN) in order to identify each word in the subjective sentence and mark whether it is aspect word or not. Also, as part of their study, the authors built integrated mutual linguistic patterns with the deep CNN. On the other hand, Lui et al. (2016) proposed an unsupervised recommendation approach based on semantic similarity and aspect associations. Their learner approach conducted various supervised sentiment classification processes in order to learn knowledge from previously conducted tasks such as Neural Network and Back Propagation. Table 1 below summarizes the various aspect-based approaches followed in literature along with its performance.

Reference	**Domain**	**Approach**	**Performance**
(Poria et al. 2014)	Product Reviews (Laptops)	Rule-based using common sense knowledge (Unsupervised)	Precision: 82.15% Recall: 84.32%
	Restaurant Reviews	Rule-based using common sense knowledge (Unsupervised)	Precision: 85.21% Recall: 88.15%
(Liu et al. 2013)	Product Reviews	Syntactical approach using Logic Programming and Double Propagation	Precision: 88% Recall: 89% F-Measure: 87%
(Xianghua et al. 2013)	Chinese Social Reviews	Unsupervised approach using LDA and HowNet lexicon	Accuracy: 91.23%
(Mukherjee & Liu 2012)	Hotel Reviews	Semi Supervised method using the SAS and ME-SAS	Precision: 88%
Chen, Mukherjee & Liu (2014)	Product Reviews	AKL (Automated Knowledge LDA)	Precision: 95%
(Samha, Li & Zhang 2014)	Product Reviews	Combined Rule-based and Lexicon-based	Precision: 99% Recall: 64% F-Measure: 77%
(Wang et al. 2014)	Product Reviews	Semi Supervised method using Fine-grained Labelled LDA (FL-LDA) and Unified (FL-LDA)	Precision: 92%
(Lui et al. 2015)	Product Reviews	Syntactical rule-based approach using supervised and unsupervised learning methods.	Precision: 84.35% Recall: 90.79% F-Measure: 87.44%
(Poria, Cambria & Gelbukh 2016)	Product Reviews	Deep convolutional learning approach and linguistic patterns	Precision: 89.6% Recall: 86% F-Measure: 87.6%

(Lui et al. 2016)	Product Reviews	Unsupervised approach using aspect semantic similarity and associations	Precision: 90.5% Recall: 95.4% F-Measure: 87%

Table 1: *Summary of the Aspect-Based Approaches*

2.4. Constructing Resources

In the field of SA, it is crucial to discover the suitable resources in order to conduct the desired experiments. In many cases, researchers use known publicly available datasets (Poria et al. 2014; Liu et al. 2013; Xianghua et al. 2013; Mukherjee & Liu 2012; Poria, Cambria & Gelbukh 2016). In other cases, it could be difficult to find a dataset in specific domain or in specific area. Tan and Wu (2011) confirmed that domain dependent sentiment dictionaries are vital to evaluating SA tasks. Joshi, Bhattacharyya and Ahire (2017) confirmed that the annotated dataset comprises of two main components; textual document and its labels. The textual document may consist manually annotated reviews with its corresponding sentiment orientation called labels. But, there are various challenges in the area of constructing SA resources. For instance, Montoyo, MartíNez-Barco and Balahur (2012) reported some of these challenges such as dealing with ambiguous words, building multi-language resources, dealing with words granularity in what ways the opinions can be expressed (as words, as sentences or as phrases), and dealing variations in opinion expressions in the different data sources (mobile app reviews, blogs, newspapers, product reviews, etc.). Joshi, Bhattacharyya and Ahire (2017) declared that the sentiment resources such as lexicons and datasets are essential for any SA system and represent the core knowledge-base for its application.

Researchers have described various approaches to construct resources in SA area. Tan and Wu (2011) developed a domain oriented approach based on random walks model to build sentiment lexicon that conveys four kinds of relationship between opinion words and documents. The authors approach utilized the PageRank and HITS graph-based concepts to represent the reviews as graphs and words as nodes. After that, to perform a random walk over the nodes in order to rank the nodes based on probabilities. The idea considered that the word polarity is discovered based on the adjacent words polarities. Likewise, the word polarity is related to the overall polarity of the document. Tan and Wu conducted their experiments on three manually labelled Chinese electronics, stock and hotels datasets and they achieved improved performance for automated sentiment lexicon construction.

Dang, Zhang and Chen (2010) have proposed enhanced method for sentiment classification by combining both ML and SO approaches based on lexicon enhanced approach. The authors have collected product reviews from various online resources and merged with the publicly available Blitzer's dataset. Each review is assigned with a rating from 1 to 5 stars considering 1 as the lowest rate and 5 the highest. The authors utilized content free, content specific and sentiment features in generating the features. Then, the word sentiment scores were calculated using SentiWordNet through determining the average for the positive, negative and objective polarity scores for similar POS tags. The researchers approach has considered 0.5 as a midpoint for objective polarity. The approach compares if the objective polarity score is greater than this midpoint value then, the word in its sense is considered as objective and then it will be excluded from the set. Otherwise, the approach will evaluate if the positive polarity score is greater than the negative polarity score for the same word in which it will be considered in this sense as a positive sentiment feature. On the other hand, if the positive polarity score is less than or equal to the negative polarity score then, the word will be added to the negative sentiment feature set. Finally, when the word positive polarity is equal to the negative polarity in the same tense then, the word will be excluded from the features set.

Steinberger et al. (2012) demonstrated a semi-automatic methodology in order to generate multi-language sentiment dictionaries. In their work, they started with gathering subjective terms in two languages which are English and Spanish. And then these two languages are automatically translated using Google translator into a third language by the intersection between translations. This technique is called triangulation. The outcome lists of triangulations are then passed to a manual process of filtering and expanding the lists to obtain broader coverage.

Boldrini et al. (2012) created a multilingual resource called EmotiBlog through developing a fine-grained annotation framework for classifying subjectivity at all levels. The researchers collected the blog posts that are covering three topics in English, Italian and Spanish languages. An inter-annotator agreement was measured in order to verify and ensure the clarity for annotators. Moreover, the authors aimed to measure the effect of the EmotiBlog components through set of

ML experiments and concluded that the whole components have advantageous impact on the system.

Robaldo and Di Caro (2013) developed an XML-based model to tag reviews that conveys opinions towards certain objects. The authors formulated a structured standard formalism called OpinionMining-ML that would contribute in resolving issues related to emotion-based claims in the information retrieval arena as well as assisting and guiding annotators in tagging reviews in order to build comprehensive datasets. Furthermore, it provides an important option for researchers to evaluate SA existing approaches.

2.5. Issues and Challenges

SA aims to determine the attitude, excitement, emotion, expression, viewpoint and opinions towards a certain entity. However, the opinion is usually influenced by the writer or author state of mind and emotional attitude. Thus, the field of SA is faced with various challenges that can affect the accuracy of obtaining appropriate sentiment orientation. Wawre and Deshmukh (2016) stated that one of the challenges in SA is having an opinion word as positive in one context and as negative in another context. Likewise, Weitzel, Prati & Aguiar (2016) pointed out that people may express their opinions differently based on the situation and the opinion words can have different orientation. Boldrini et al. (2012) reported the diversity of challenges in SA field such as having blended text formats, having an extensive range of data sources and domains, the use informal language, multilingual resources, lack of correct grammar and spelling, the use of slangs, and the massive growth of information.

Weitzel, Prati and Aguiar (2016) also pointed out other challenges. Coreference resolution is meant to identify the reference of the words that points to the certain entity such as pronouns. For example: the pronoun "*it's*" is considered anaphora and refers to "*application*" in the following review "*I can't download the application. it's not working.*". On the other hand, word sense disambiguation (WSD) is concerned in identifying the words with vague meaning according to its context. Meanwhile, producing newly optimized parsing through having better and broad NLP understanding more than just parse sentences into their POS and fetch for it in dictionaries. Another common challenge is the negation handling which is concerned about inverting the orientation of opinion words. For example, the positive polarity related to the word "good" would normally be changed into a negative polarity for the “not good” phrase. Negation discovery is considered as a complex task as negation does not only deals with syntactic negation (not and nor) but also deals with linguistic patterns such as prefixes (un- and dis-) or suffixes (less). Similarly, Wawre and Deshmukh (2016) highlighted that handling negation basically changes the meaning and the opinion orientation.

Blanco and Moldovan (2014) considered identifying the negation scope as the most important task in negation handling. The negation aspects such as morphological and syntactic are common in NLP. The authors suggested that further negaters should be considered other than the typical ones (not) with proper description of their scope, and the way it can be computationally determined. It is important to decide the impact of the negation words on the opinion words polarity. One approach is called switch negation which consists of reversing the polarity of the opinion word. For example, an opinion word “*good*” with positive polarity and “+2” sentiment strength, the polarity of “*not good*” will be reversed to negative with “-2” sentiment strength. Another approach is called shift negation which consists of shifting the negated word with a scale of specific threshold without reversing the polarity of the original word. This method is advised to be adopted in cases where there are highly positive or negative opinion words. For instance, an opinion word “*perfect*” with positive polarity and “+5” sentiment strength, the polarity of “*not perfect*” will be shifted to “+2” sentiment strength.

Xu et al. (2011) highlighted that the reviews contain considerable amount of comparative opinions. The users may express their opinions through conducting a comparison between various entities instead of direct and precise opinion towards the entity or its features. For instance, “*The previous version of the app was much better*”, at a first glance, the review seems to have positive polarity while, if the review is deeply analysed, it can be concluded that the user has a negative polarity towards the new version of the app compared to the previous version. These comparative opinions are a significant source to identify the product’s or service’s strengths and weaknesses so that it reflects positively on the future design and features of the product or service.

Taboada et al. (2011) described the challenge of dealing with intensifiers and downtoners in SA tasks. An Intensifier can be defined as the word that causes an increase of the sentiment score in case of positive opinions and a decrease of the sentiment score in case of negative opinions such as “*Extremely*” intensifier in “*Extremely great app*” review. Whereas, a downtoner can be defined as the word that slightly reduces the sentiment score in case of positive opinions and slightly increases the sentiment score in case of negative opinions such as “*quite*” downtoner in “*quite good app*” review. The authors stated that not all intensifiers intensify at similar level. The sentiment score of the opinion word being intensified affects the intensity of the intensifier. Thus, Taboada et al. proposed a solution to use multiplication with the intensifier percentage weight instead of addition and subtraction. For instance, 100% is assigned to “*most*”, 25% is assigned to “*really*”, 15% is assigned to “*very*”, -30% is assigned to “*somewhat*” and so on. In the same way, Kundi et al. (2014) treated the repeated characters as intensification problem and suggested a solution that whenever two or more consequent characters are found to improve the word sentiment scoring for both positive and negative orientation. For example, in the following review: “*Baaaaaaad app*”, assume that the score of “*bad*” is equal to “-2” then the sentiment score due to the repeated characters could be “-4”.

Sharma, Nigam and Jain (2014) revealed that there are instances when surveys involve sarcasm. These are sentiments that show contempt, but they cannot be classified under the negative sentiments. Weitzel, Prati, & Aguiar (2016) concluded that FL is one of the most difficult challenges in SA field. FL has various examples such as irony, sarcasm, analogy and ambiguity. The FL can be defined as extending the language meaning according to the author’s perspective by deviating from the original meaning of the words.

2.6. Emergence of Smart Government

The growing demand for public services has led to the emergence of Electronic Government (eGovernment) and Smart Government, hence the government's interest in diversifying the channels of providing these services such as face-to-face services by visiting customers to government departments, by telephone, via the website, via smart phones, through self-service kiosks, via SMS and others. Smart channels are considered now as one of the most important channels which greatly facilitated the public services and accelerated their transactions. For instance, Gil-Garcia and Martinez-Moyano (2007) stated that public sector is utilizing information technologies in their business as part of the day-to-day operations. Gil-Garcia and Martinez-Moyano also explained that the governments adoption of e-government to the great benefits by increasing the confidence of customers in these services in addition to the demonstrated transparency and reliability. Moreover, Gil-Garcia, Helbig and Ojo (2014) described the increasing prominence on all services impact on business constructing a competitive innovative services model. The researchers also highlighted the globalization impact bringing up integrated customer centric solutions through combining products and services to keep up with the connected world in all aspects.

The concept of e-government has qualified the public sector to capture, process and report outcomes based on the populated data efficiently which improved their decision making. But, the continuous evolvement in smart technologies, better informed consumers, and globalization have created potential opportunities to overcome various challenges (Harsh & Ichalkaranje 2015). Smart government is meant to describe the processes of the combined growing technologies and innovative strategies to attain resilient government framework that helps in making effective and quick decisions (Gil-Garcia, Helbig & Ojo 2014).

The governments have started to take the perception of e-government a step forward by understanding the power of existing data that can be the influential tool for services improvement, well integrated service experience, effectively engage customers, maintain policies and implement solutions for transforming their institutes into being a ‘smart government’. Thus, the citizens have been granted to engage with government organizations in a new way due to the massive growth in the development technologies, social media, mobile apps, big data analytics, open data and open innovation (Harsh & Ichalkaranje 2015). Gil-Garcia, Helbig and Ojo (2014) pointed out that mobile apps are a great example of the combined innovation and emerging technologies.

2.7. Smart Applications Standards and Guidelines

As stated in section 1.3, the emerging technologies has led to dramatic increase of smart phones users and to major changes in the way of government services provisioning through well designed and powerful mobile apps. According to Gartner (2017), the most widely used smart phones operating systems are iOS (developed by Apple) and Android (developed by Google Inc.), constituting of 99.6% of the total market share. Apple and Google Inc. have set standards and guidelines to be met for producing high quality smart applications that meet the high expectations of the consumers.

For instance, Apple (2017a) considered three primary themes distinguish them from their competitors in order to satisfy the high quality and functionality expectations for potential consumers. First, the "Clarity" theme throughout the whole app requires having a readable text, proper font size and colours, simple and attractive graphics and icons, emphasis on improved functionality, and suitable interface structure that highlight important app content and attract the users. Second, the "Deference" theme that emphasizes on that the content is considered as the most important part. It ensures that the app should have an attractive and elegant interface that helps users easily understand the content and interact with it. Moreover, ensure that the app content covers the complete mobile screen. On the other hand, the app should be simple with minimal use of unnecessary design elements such as drop shadows, blur items, bezels and gradients. Third, the "Depth" theme that suggests having unique visual layers and genuine motion that reflects the app hierarchy, show vivacity, and easy for users to understand. Additionally, offer the sense of depth when the users explore the content and practices the touch features and discovering app functions without missing the context.

Apple (2017a), have also defined six principals that should be taken into consideration to make the most of the impact and reach of the mobile app through its identity. These principals are aesthetic integrity, consistency, direct manipulation, feedback, metaphors and user control. The "Aesthetic Integrity" denotes the app's artistic appearance and behaviour and how it integrates with its functionalities. Meanwhile, the "Consistency" is meant to satisfy peoples' expectations through combining both app features and behaviours presents overall ordinary standard features, user interface elements, font styles as well as unified icons and terminologies. In the same way, the "Direct Manipulation" enhances the user's experiences by seeing immediate and clear responses for their interactions such as when the user experiences the instant response upon rotating the device screen. Another principle is the "Feedback", which admits the user's actions and keep them engaged and informed by using interactive design elements such as status indicators for long processes and feedback animations and sounds. Subsequently, the "Metaphors" are the familiar experience for users when they promptly learn through the app's objects and actions by the physical interaction with the touch screen. Finally, the "User Control" principle ensures that the users are in control on the app. The app should have an interactive and predictable design in a way that keeps a balance between allowing users to take control and avoiding the negative consequences by making the users are aware that they are in control.

Apple (2017b) has defined a set of guidelines to be followed before releasing any app and deploying it to the app store. These guidelines will ensure a higher level of satisfaction that the users experience. First, the “Safety” guideline is meant to guarantee that the app does not contain any harmful content as well as the app will not cause any damage to the smart phone and therefore, people will be confident about the app. Second, the “Performance” guideline defines several essential components that should be available prior to releasing the app. These components include the completeness of the app functionality, the accuracy of the app’s content, software requirements, and hardware computability. Third, in the “Design” recommendation, Apple emphasises on having apps that are simple, attractive, beneficial, and easy to use. Finally, the “Legal” guideline stresses on the user’s privacy in a way to ensure that the user’s collected and stored data is protected according to the user’s agreement.

Similarly, Android (2017) has proposed a set of quality guidelines that should be taken into consideration when developing the app and should be tested against them before the app release. The first quality guideline is “Visual Design and User Interaction” in which a standard design that follows a consistent design patterns, clear navigation structure, and timely and efficient notifications are taken into consideration. In the “Functionality” guideline, the app should function as expected. For example, the app should only request permissions to data that is relevant and needed to perform well. Additionally, the application should support screen rotation when possible while preserving the functionality and efficiently utilizing the whole screen. Also, the app should maintain and restore the app state without any unintentional data loss. The next guideline is “Compatibility, Performance and Stability”. This guideline guarantee that apps compatibility and stability is provided. Moreover, the app should deliver a prompt user feedback such as status indicator whenever any app loading process consumes more time than expected. Additionally, the apps various graphics elements should maintain high visual quality, visibility and clarity for all possible screens. Another guideline is concerned about “Security” and how sensitive data is carefully dealt with as well as how the users’ policies are defined in order to protect their privacy.

2.8. Dubai Government Smart Websites Excellence Model (SWEM)

Public sectors around the world have recognized the significance of having comprehensive websites and mobile apps in order to deliver their services to citizens. In the same way, Dubai government has been concerned about providing their services to public through online and smart channels while keeping pace on new trends and best practices.

Smart Dubai Government (2017) developed the SWEM, which is a combined broad set of guidelines that aim to increase the level of customers' happiness with government online and smart services. The main objective of SWEM is to define a clear set of rules and recommendations that all government departments should comply with and adopt in their online and smart channels. These guidelines are strongly influenced by publics' needs and expectations and aim to attain customers' happiness. Furthermore, SWEM aims to increase the electronic and smart channels adoption and usage by producing a high quality, functional, user friendly and easy to use government online and smart portals. SWEM main focus is to have a unified experience through all government entities while allowing a room of customization and innovation as needed by each entity to ensure customers' needs and expectations are met.

SWEM consists of four core elements with various guidelines available under each one as shown in Figure 6. These elements are Accessibility, Usability & Design, Content and Functionality. Firstly, since government services are used by diverse segments of citizens, the "Accessibility" component is essential to ensure the services are customizable to meet all of their needs. For example, the mobile app should include multilingual options, easy access for disabilities, clear structure to reach the service, and compatibility with different devices. Then, as it is important to deliver a unified perception to the entities' customers with unique branding through all channels, the "Usability & Design" component is very crucial in order to reveal the entity's brand as part of Dubai government as well as its own image. This element covers many aspects including the use of simple and consistent user interface graphical as well as form elements, proper navigation structure, the content visibility, the utilization of smart phone touch features, clear and prompt notification alerts, and unambiguous user interface functionalities. The third element is concerned about "Content". As per SWEM, the content should be up-to-date, precise, relevant, sufficient and readable in order to satisfy the needs and expectations of the customers.

Besides, in order to provide channels of interaction with customers, an effective facility to engage customers as well as collect their commendations and recommendations should be offered. For instance, the apps may include live chat, experience rating, feedback, suggestions and complaints, contact forms. Additionally, social media can be integrated as it is a vital feature that fosters the customers interaction and engagement through social media networks. Finally, the "Functionality" element states that the service should perform as indicated by the government department and according to the customers' expectations. Similarly, the service should be free of errors and should provide responses to the users in a timely manner and according to the expectations.

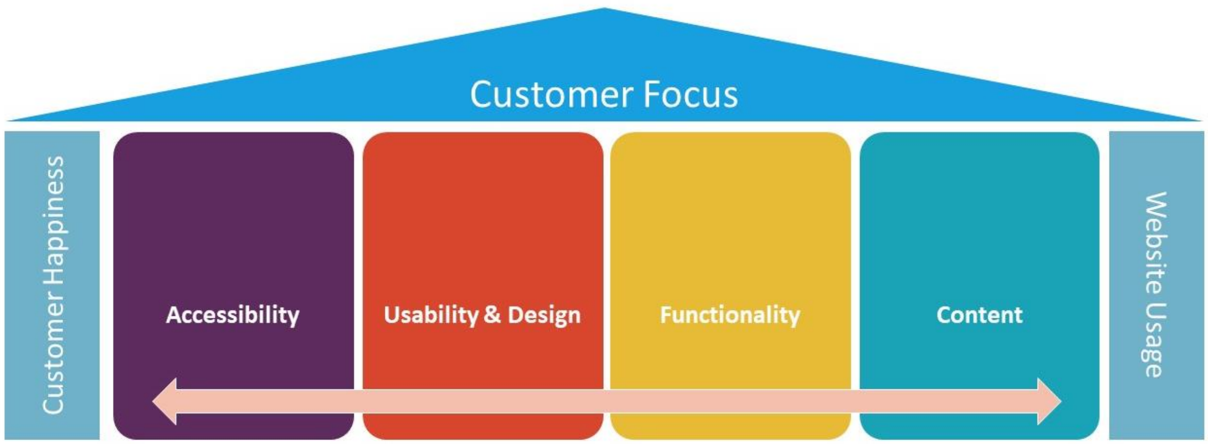

Figure 6: *Smart Dubai Government Excellence Model 4.0*

3. Chapter Three: Building Dataset

As discussed earlier in section 2.4, sentiment resources are essential in order to develop any application for SA. Common publicly available datasets such as products, restaurants and movies reviews, usually fulfil the researchers needs in order to conduct their experiments. However, for specific domains, the needed dataset sources could be difficult to find. This signifies the need to construct domain specific datasets and lexicons which are vital to evaluate SA tasks.

In this chapter, an illustration of the work has been done in order to build a unique dataset that consists of government mobile apps domain aspects and opinion words. Additionally, it explains the approach that was carried out to measure the sentiment scores to opinion words and build the desired lexicons.

3.1. Methodology

A manual annotation approach has been chosen with the help of the developed GARSA tool to facilitate the complete annotation process. Although this manual approach is time consuming and requires dedicated resources, it considered to be more accurate than the automated methods (Liu 2012). The methodology to build the anticipated dataset and lexicons is illustrated in Figure 7. The methodology details are provided in the following sections.

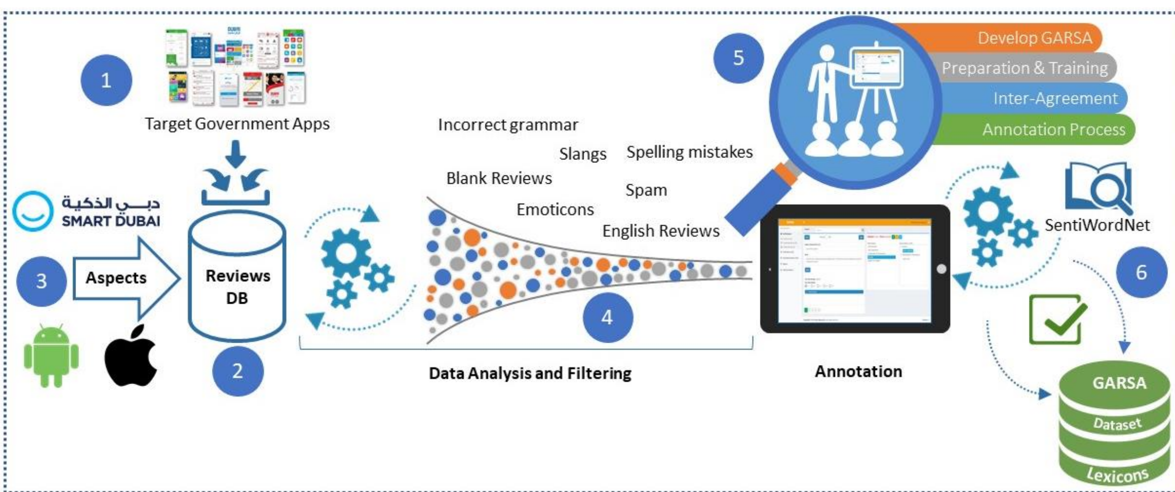

Figure 7: *Methodology for Building Dataset and Lexicons*

3.2. Data Collection

This research scope covered the well-known government apps in UAE mainly in Dubai. There were 30 distinct smart apps with total of 60 apps in both Google Play store and Apple iOS as shown in Table 2. The reviews that have been collected include all reviews posted in the period from the beginning of 2013 to the end of 2016. The reason behind that is as stated in section 1.3, the year of 2013 represents the year when His Highness Sheikh Mohammed bin Rashid announced the launch of the "Smart Government". The collected data consists of a set of reviews each including the app name, store name, app ID, review ID, language, star rating, date, author, subject, and body.

#	App Name	Government Entity
1	AD Services	Abu Dhabi Systems and Information Centre
2	City Guard	Abu Dhabi Systems and Information Centre
3	Darb	Abu Dhabi Department of Transport
4	DEWA	Dubai Electricity and Water Authority
5	Dubai Brokers	Dubai Land Department
6	Dubai Police	Dubai Police
7	Abu Dhabi Taxi	Abu Dhabi Taxi
8	Visit Dubai / Dubai Tourism	Dubai Tourism
9	Dubai Now	Dubai Smart Government
10	EJARI	Dubai Land Department
11	GDRFA Dubai	General Directorate of Residency and Foreign Affairs
12	iDubai	Dubai Municipality
13	Jobs Abu Dhabi	Abu Dhabi Systems and Information Centre
14	KHDA	Knowledge and Human Development Authority
15	Makani	Dubai Municipality
16	UAE MOFA	Ministry of Foreign Affairs

17	mjobs	Dubai Smart Government
18	MOI UAE	Ministry of Interior
19	mPay	Dubai Smart Government
20	RTA Corporate Services	Roads and Transport Authority
21	RTA Drivers and Vehicles	Roads and Transport Authority
22	RTA Dubai	Roads and Transport Authority
23	RTA Public Transport	Roads and Transport Authority
24	RTA Sharekni	Roads and Transport Authority
25	RTA Smart Parking	Roads and Transport Authority
26	DU	DU Telecom
27	RTA Smart Taxi	Roads and Transport Authority
28	Sehhaty	Dubai Health Authority
29	SEWA	Sharjah Electricity & Water Authority
30	Wojhaty	Roads and Transport Authority

Table 2: *List of Government Mobile Apps Used in this Research*

3.3. Data Analysis and Database Design

The collected data consist of 11,912 reviews from both Google Play store and Apple iOS. The collected reviews were stored in Microsoft Excel format. This facilitates the process of the data review and analysis to get insights of the data characteristics. As illustrated in Figure 8, a percentage of 76% of the collected reviews were in English language, 20% in Arabic Language, and 4% in other languages. Moreover, the majority of the reviews were positively rated. For instance, 6438 of the reviews were rated by the users as five stars, 1068 as four stars, 691 as three stars, 693 as two stars, and 3022 as one star. This is consistent with the survey results where people prefer to post positive reviews over negative reviews.

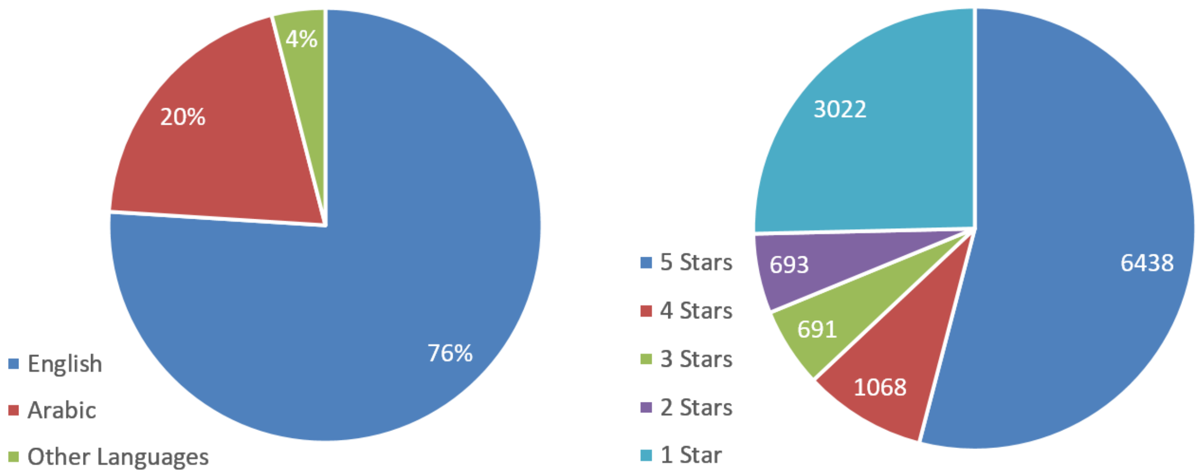

Figure 8: *The Language and Star Rating Statistics of the Collected Reviews*

According to the data analysis, it has been noticed that emoticons are not commonly used in apps reviews. Thus, the emoticons will not be considered in the experiments. Furthermore, it has been observed that slangs were used in a variety of the reviews. Also, there were a lot of spelling mistakes and incorrect grammar which will require special treatments. Interestingly, there were no spam reviews in the whole collected data. 1.62% of the reviews were blank and were excluded for not containing the subject or body fields.

According to the information in Figure 8, it is important to note that the majority of the reviews for government mobile apps in UAE are written in English language. Therefore, for the sake of covering larger segment of the government mobile app users, the experiments in this research will be only limited to reviews written in English language.

In order to build a tool that can assist the reviews annotation process, the collected data have been migrated to Microsoft Access database. Further, the desired tables as well as its relationships were created as shown in Figure 9. The database includes UsersList, Reviews, SentimentAspectAnnotation, and AspectsList tables.

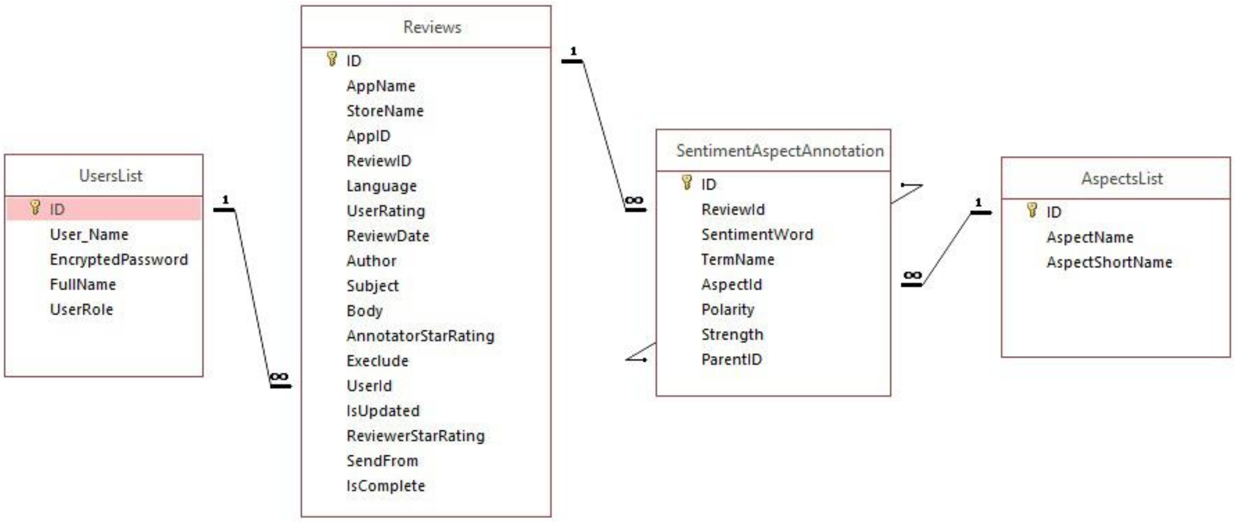

Figure 9: *Entity Relationship Diagram*

The table below (Table 3) explains the details of the database tables such as the important fields and its usage as well as the relationships between the tables:

#	Table Name	Description
1	UsersList	The purpose of this table is to store the users' information in order to manage the application users and to assign the corresponding roles. The predefined roles are: • **Annotator**: this role is used for those users who will read their assigned reviews, tag the candidate opinion words and aspects terms, and assign them to the best matching predefined aspect. • **Reviewer**: the purpose of this role is to have access to verify the reviews that have been annotated by the annotators. The user with this role will be able to override, update and approve the values that are set by the annotators according to the review context. • **Administrator**: the users with this role will be able to access all reviews with full permissions. Also, administrators will act as domain experts who are responsible of conducting the final review and approval on the annotated reviews.
2	Reviews	This is the main table that contains the AS-IS collected reviews a long with the star rating of the users. Various fields have been added to this table in order to fulfil the annotation requirements: These fields are: • **UserID**: this field stores the user who will be assigned to work on this review annotation process. The user details are looked up from the UsersList table. • **AnnotatorStarRating**: this field is used to store the annotator user star rating judgment according to the review context.

		• **IsUpdated**: this field states that the review has been completed by the annotator and ready for verification. The updated reviews will be visible to the users with Reviewer role. • **ReviewerStarRating**: this field is used to store the reviewer user star rating judgment according to the review context. • **Exclude**: this field is used to tag those reviews that need to be excluded from the list for any reason such as blank reviews, reviews in other languages, spam reviews, among others. • **IsComplete**: this field marks the reviews that have been annotated and reviewed as completed and will be locked for update. Only administrator can unlock and update the review annotation.
3	AspectsList	This represents a look up table that stores the information about the government mobile apps domain aspects. These aspects as explained in section 3.4 are User Interface (UI), User Experience (UX), Functionality and Performance (F&P), Security (SEC), and Support and Updates (S&U).
4	SentimentAspectAnnotation	This table stores the annotation details. These details include a reference for the review, the sentiment or opinion word, the aspect term if available, the corresponding aspect, the polarity, and the sentiment strength value. Furthermore, it allows to store the words with spell mistakes and link it with the corrected annotation record through recursive relationship with the same table.

Table 3: *Details of Database Tables*

3.4. Innovative Mobile App Aspects Design Instruments

As per section 2.6, the most widely used smart phones operating systems are iOS (developed by Apple) and Android (developed by Google Inc.) as well as Smart Dubai Government SWEM have defined a set of components and guidelines in order to produce high quality smart applications that meet the customers' expectations and provide an exceptional customer experience. However, the customer experience aspect was not emphasized as a standalone component although it is one of the most important aspects if it is looked at from the customers' perspective rather than the entity's side. On a similar note, Mills (2016) focused on following certain factors to design the desired customer experience according to who are our customers, what they want to do, and why. This can be concluded by answering the question "*What will my customer remember about this experience?*".

Thus, a novel comprehensive model has been designed to serve this study. This model takes into consideration the state-of-the-art guidelines as well as the aspects according to the customers' perspectives and viewpoints. As illustrated in Figure 10, this model consists of five main type of aspects namely; User Interface, User Experience, Functionality and Performance, Security, and Support and Updates. Following is the description of each aspect and the areas it is expected to cover:

- **User Interface**: This aspect conveys a unified design that follows a consistent pattern, clear navigation structure, and unambiguous user interface functionalities. The design elements should be simple, attractive, and easy to use.
- **User Experience**: This aspect is generally concerned about the customers' experience and the level of satisfaction as well as their feelings about the service. This includes how convenient is the app for the users, how appropriate is the structure, and how the users are comfortably completing the service. Also, this aspect includes offering multilingual support as well as avoiding redundant features.
- **Functionality and Performance**: This aspect states that the service should perform as indicated by the government entity and according to the customers' expectations. Likewise, the service should provide responses to the users in a timely manner and according to the expectations.
- **Security**: This aspect asserts that the service should provide secured user login and payment services as required. Additionally, it emphasizes on the user's privacy in a way that assures the user's collected and stored data is protected as per the terms and conditions.
- **Support and Updates**: This aspect states that the service should offer effective channels to engage customers and gather their inputs to understand their pain and gain areas. The app should be reliable, stable and free of errors and should release updates when needed.

Figure 10: *Government Mobile App Aspects for SA*

3.5. GARSA for Reviews Annotation

GARSA is a responsive web tool that has been designed to facilitate the annotation process in an organized way. The proposed tool has been developed using ASP.NET and HTML5 components based on the designed database as explained in section 3.3 and it currently covers the annotation part.

GARSA is a flexible application that has been designed in a way that will allow the annotators to perform the work efficiently with minimum errors. The tool was developed as a web application that can be accessed anytime and anywhere regardless of the location of the users. The application has been hosted with SMARTERASP.NET that supports the required .NET libraries and databases. The table below (Table 4) summarizes the main feature of GARSA:

#	Feature Name	Description
1	Simplified User Workspace	The tool considered the tracking of the users' progress through providing a comprehensive and simplified user workspace. It consists of a personalized user inbox that is classified into three categories (All Reviews, Pending Reviews and Completed Reviews). This allows the users to access and filter all assigned reviews easily with option for administrator to access the users' inboxes. The workspace provides the users with flexible preview of all review details including the review id, subject, body, and the user's star rating. Moreover, it allows the user to set the desired star rating as well as send the review for admin advise in case of any confusion. The workspace also

		provides an option to the annotators to correct the miss-spelled words and add both the corrected word and the miss-spelled one.
2	Responsive Web Application	The application was designed to adjust according to the device used regardless of the orientation. The content will be automatically and dynamically arranged according to the changes of the browser available area for display. It provides optimal arrangement of the application's content to fit on desktops, laptops and any smart devices.
3	Cross Browser Support	The application was designed and tested to run on all browsers. All functionalities were tested across well-knows browsers; namely Internet Explorer, Mozilla Firefox, Google Chrome, and Safari.
4	Search Functionality	The tool provides a single search box that allows the users to search across all reviews through any term within the review or its id. This facilitates the annotation process and helps the users have a deeper look into reviews with certain characteristics.
5	Polarity Assistant	A holistic view of all annotated and approved reviews that helps in deciding on those words which are needed to figure out the polarity based on its appearance in the annotated reviews and the associated star rating. This functionality can be also used as a reference for the annotators to get insights on certain opinion words and aspects and its approved annotations. The polarity assistance allows the users to filter reviews with combination of any of the opinion words, aspect terms, and aspects. Lastly, it offers complete search with a detailed view of all reviews.
6	Simple Workflow	The tool contains a simple workflow that ensures accurate annotation process. As shown in Figure 12, the annotation process comprises of the annotation and verification sub-processes that are covered by the three main roles (Annotator, Reviewer, and Administrator).
7	Automatic Rules Definition	The tool provides intelligent guidance during the annotation process. It considers any approved annotation and automatically creates rules. The rule can be defined as the restriction for the assignment of sentiment word and aspect term to the desired aspect. Thus, when the annotator selects an opinion word and aspect term, the application will check through the created rules. After that, in case of conflict between the user's annotation and the defined rule, the application will prompt and advise the annotator with the correct annotation. This will make sure that all annotators will classify similar cases in the same way whenever it appears. For example, the following approved annotation is considered as a rule: • The sentiment word: "N*ice*" • The aspect term: "D*esign*" • The assigned aspect: "*User Interface*" Therefore, for any new annotation with the above combination of "*nice*" sentiment word and "*design*" aspect term should be annotated as "*User Interface*" aspect. Additionally, the application will create rules that will not allow to assign an approved sentiment words as aspects terms and vice versa.

8	Graphical Reports	The tool provides a set of graphical cloud-based reports that reflects the frequency of the corresponding word (sentiment or aspect term or aspect) and render the representation accordingly. The reports visually aggregate and summarizes the dataset lexicons. The reports include the following: • Positive Lexicon Cloud: displays the aggregated list of positive opinions • Negative Lexicon Cloud: displays the aggregated list of negative opinions • Neutral Lexicon Cloud: displays the aggregated list of neutral opinions • Aspect-Terms-Sentiment Cloud: displays the aggregated list of aspects that consists of the opinion and aspect terms. • Explicit Aspects Cloud: displays the aggregated list of explicit aspect terms.
9	Additional Functionalities	The tool provides additional functionalities that help the users to easily navigate through out the assigned reviews. For faster performance, the navigation consists of two levels of navigation; navigation through the reviews within batches and navigation though the batches. Another convenient feature is the "*Auto Save*" feature that prevents any loss of users' work. In addition to that, the tool provides additional annotation options that ease the selection and the assignment of the sentiment words as well as aspect terms to the corresponding aspects.

Table 4: *GARSA Features Summary*

3.6. Manual Annotation Approach

A team of three "English Language and Literature" students in addition to the domain expert has been formed to perform the complete annotation process. As illustrated in Figure 11, the annotation process includes three main sub-processes which annotate, review, verify annotation, and approve annotation. Each of these phases comprises several activities illustrated in the reviews annotation process flow diagram in Figure 12.

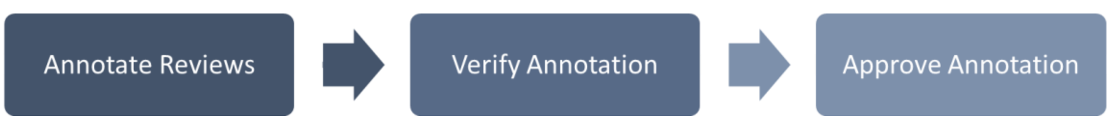

Figure 11: *Summary of the Reviews Annotation Process*

At the beginning, it was required to prepare the team and make sure that all terms and requirements are clear. A user manual was prepared that includes step-by-step details on using the application as well as on the annotation process requirements. The users were first asked to read the manual and play around with the application. Then, a face-to-face training was conducted by the domain expert followed by hands on excersize to ensure that the team members have proper understanding of what needs to be carried out. During the training, it has been decided to assign two of the students the "*Annotator*" role and the third student was assigned the "*Reviewer*" role. And the domain expert was assigned the "*Administrator*" role.

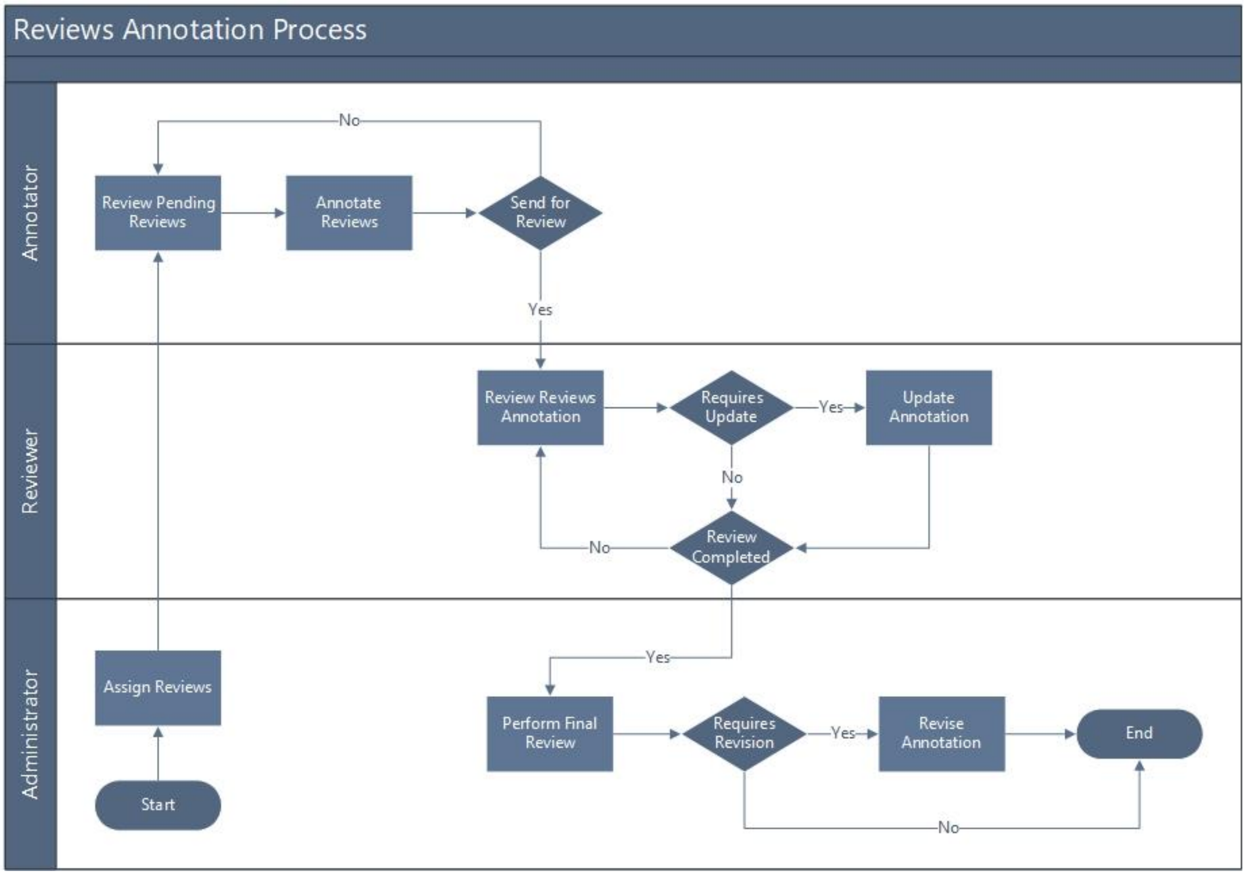

Figure 12: *The Reviews Annotation Process Flow Diagram*

The user manual was built according to the defined standards and definitions that will be used as reference for the team. As shown in Figure 13, the annotators utilized the GARSA functionalities to perform the annotation by reading and understanding each review, selecting the opinion word, aspect term and assign it to the corresponding aspect. Moreover, the annotators assigned a star rating for the review according to their understanding and judgement. Then, the annotated reviews are verified by the reviewer. The reviewer may update the assigned annotations as well as assign star rating for the review similar to the annotators. Additionally, the reviewer will mark the verified reviews as complete so it will be displayed in the domain expert page for final review and approval. The domain expert may conduct the same functionalities similar to the reviewer.

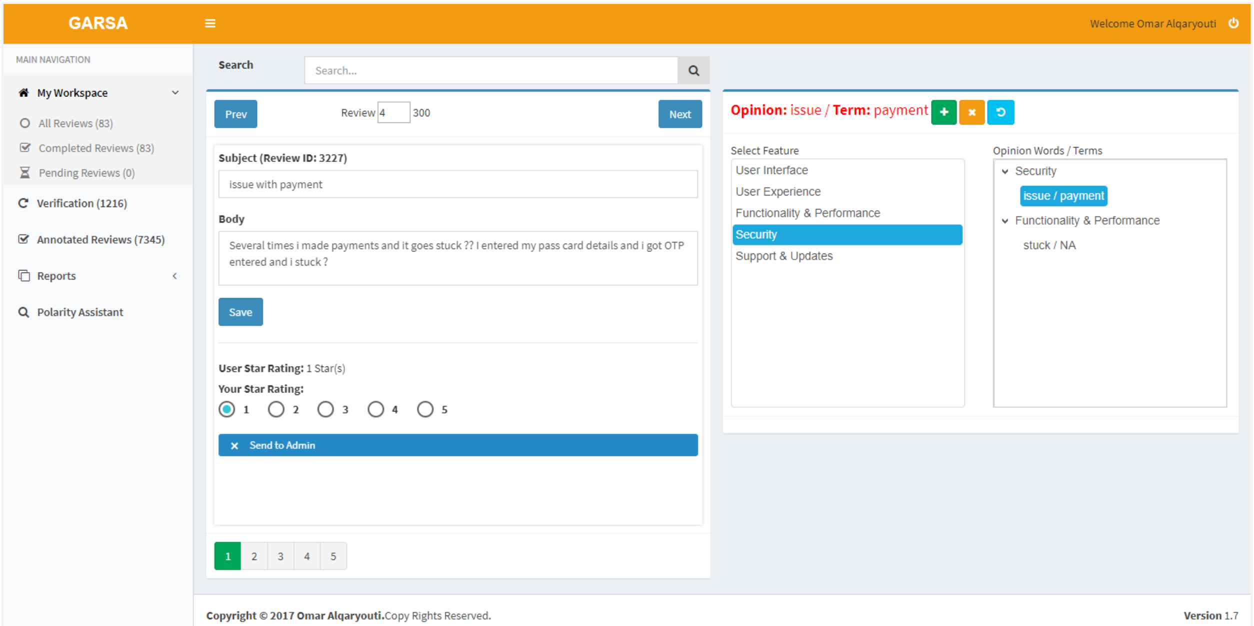

Figure 13: *Screenshot of the Annotation Page in GARSA*

3.7. Inter-Annotator Agreement

As stated in section 3.6, a team of four members has been formed to perform the complete annotation process. Two of these team members are responsible for annotating the reviews. Before the real annotation process started, an agreement analysis was conducted by selecting a random 300 reviews from the database. The purpose of this analysis is to evaluate the reliability of the annotation process and observe the agreement level between the two annotators. Also, it ensures that the annotators have common and proper understanding of the requirements and assures high level of annotators' confidence of the annotation. The agreement measure provides an indication on the disagreement areas which may necessitate further explanation for the annotators.

Similar to Takala et al. (2014), Cohen's Kappa statistics are selected to measure the inter-annotator agreement. The Kappa agreement is only used when there are only two annotators. The Kappa (k) can be defined as per the following formula:

$$k = \frac{P_a - P_c}{1 - P_c}$$

Where:

P_a is the qualified agreement that has been observed for the two annotators.

P_c is the theoretical probability of a random or chance agreement.

Table 5 below illustrates the two annotators agreement and disagreement matrix. The first annotator (A) agreed to annotate the same aspects with the second annotator (B) in 372 occurrences for the same reviews (TT). Also, A agreed with B not to annotate aspects in 748 occurrences for the same reviews (FF). On the other hand, A disagreed with B by annotating 12 aspects that have not been annotated by B (TF). And A disagreed with B by not annotating 13 aspects that were annotated by B (FT). In order to calculate k, P_a and P_c needs to be measured as follows:

$$P_a = \frac{TT + FF}{TT + FF + TF + FT} = \frac{372 + 748}{372 + 748 + 13 + 12} = \frac{1120}{1145} = 0.9782$$

$$P_c = P(T) + P(F)$$

$$P(T) = \frac{TT + TF}{TT + FF + TF + FT} \times \frac{TT + FT}{TT + FF + TF + FT} = \frac{372 + 12}{1145} \times \frac{372 + 13}{1145} = 0.1128$$

$$P(F) = \frac{FT + FF}{TT + FF + TF + FT} \times \frac{TF + FF}{TT + FF + TF + FT} = \frac{13 + 748}{1145} \times \frac{748 + 12}{1145} = 0.4412$$

$$k = \frac{P_a - P_c}{1 - P_c} = \frac{0.9781 - 0.5539}{1 - 0.5581} = \frac{0.4242}{0.4460} = 0.9511$$

		First Annotator (A)	
		Annotated Aspect (T)	**Not Annotated Aspect (F)**
Second Annotator (B)	**Annotated Aspect (T)**	372	13
	Not Annotated Aspect (F)	12	748

Table 5: *Annotators' Agreement and Disagreement Matrix*

The table below (Table 6) provides breakdown details on the agreement and disagreement levels for all aspects along with the Kappa's measures. It is noticed that the "*User Experience*" aspects indicated the lowest k percentage due to the difficulties in indicating the opinion words and aspect terms that represent explicit or implicit aspects. However, the *k* percentage (82.63%) is still indicating high value with almost perfect results according to Takala et al. (2014).

	Aspect	Agreement and Disagreement				
		TT	**TF**	**FT**	**FF**	**K**
Aspects	**User Interface**	45	2	4	174	93.04%
	User Experience	189	4	3	19	82.63%
	Functionality and Performance	39	3	2	181	92.61%
	Security	27	2	2	196	92.09%
	Support and Updates	63	1	2	178	96.84%
	Total	372	12	13	748	95.11%

Table 6: *Annotators Agreement and Disagreement Matrix for All Aspects*

According to the calculated results for Cohen's Kappa measures, all results were greater than 81% for all aspects. As stated by Takala et al. (2014), if the Kappa measure is greater than 81%, then the agreement level is almost perfect as shown in Table 7. Thus, the results confirmed that the annotators perceived a high level of confidence of the requirements understanding and annotation process.

Kappa	**Agreement**
< 0.00 %	Less than chance agreement
1% – 20%	Slight agreement
21% – 20%	Fair agreement
41% – 60%	Moderate agreement
61% – 80%	Substantial agreement
81% – 99%	Almost perfect agreement

Table 7: *Guidelines for Interpreting Kappa Agreement Measures (Takala et al. 2014)*

3.8. Lexicon Generation

Prior to data annotation, all annotated opinion words along with their aspects terms have been accumulated in "*SentimentAspectAnnotation*" database table to build the desired lexicon. First, the redundant opinion words and aspects terms are removed. Then, a process built on python is utilized to go through all opinion words to obtain the polarity score from SentiWordNet. The polarity score ranges from -1.0 for negative words to 1.0 for positive words. SentiWordNet were chosen because it covers more words than any other available sentiment dictionaries. If the opinion word is not found in SentiWordNet, the process utilized WordNet Lemmatizer to remove grammatical markings. For example: flexibility is lemmatized to flexible and buggy is lemmatized to bug. However, the sentiment in SentiWordNet is linked to the word meaning instead of the word itself unlike other lexicon resources such as SO-CAL (Joshi, Bhattacharyya & Ahire 2017). This tolerates a word to have multiple sentiments towards each meaning. Therefore, the developed process has adopted similar approach to Dang, Zhang and Chen (2010) where the following rules were applied to calculate the polarity score:

If the Average Positive Score (Opinion Word) > The Average Negative Score (Opinion Word) Then:

The process will set the average positive score as polarity

If the Average Negative Score (Opinion Word) > the Average Positive Score (Opinion Word) Then:

The process will set the average negative score as polarity

In the emerged dictionary, variety of words were annotated as opinion words although they have zero polarity score in SentiWordNet. For instance, the word "*delay*" has zero average score in SentiWordNet. However, this word is considered as domain-specific opinion and as agreed by annotators, it expresses negative opinion in the reviews. Thus, these kinds of words were reviewed with the help of the GARSA in order to figure out the suitable polarity score.

Basically, the average of the user star ratings on the reviews that contain each of these words were calculated with approximate assumption that 5 stars rating has 1.0 polarity score, 4 stars rating has 0.5 polarity score, 3 stars rating has 0 polarity score, 2 stars rating has -0.5 polarity score and 1-star rating has -1.0 polarity score. For example, if the GARSA showed that the word "*delay*" occurred in 4 reviews where one review has 4 stars rating, two reviews have 3 stars rating, and one review has 1-star rating. Then, the resulted equation will be as follows:

$$Polarity\ Score\ (delay) = \frac{(1\times0.5)+(2\times-0.5)+(1\times-1)}{4} = -0.375$$

On the other hand, the dictionary contained some words with spelling mistakes which were not found in the SentiWordNet as well. These words have been corrected and annotated during dataset annotation phase and the polarity score of the correct word was assigned to them in the dictionary using the above approach. The miss-spelled words were linked to the parent corrected words and inherited its polarity. Finally, the remaining scored opinion words were also reviewed with the help of GARSA in order to verify and confirm the correct scoring with respect to Government Mobile Reviews domain. The resulted dictionary consists of four attributes: Opinion word, aspect term, aspect category and the polarity score.

3.9. Dataset Construction

As explained earlier GARSA was used to annotate 7,345 reviews by manually identifying the opinion words, aspects terms and aspects categories. Then, each opinion word has been assigned with its polarity score by utilizing the generated lexicon. Next, the average of polarity scores associated with opinion words in each aspect category were calculated in order to identify the rating of each aspect in the review. Table 8 shows the criterion of the considered ratings in the dataset based on the resulted polarity score.

Stars	**Description**	**Average polarity score**
1 star	very negative review	<= - 0.6
2 stars	negative review	> - 0.6 and <= – 0.2
3 stars	neutral review	> - 0.2 and <0.2
4 stars	positive review	>=0.2 and <0.6
5 stars	very positive review	>= 0.6

Table 8: *Star Rating Mapping to Average Polarity Score Criterion*

Finally, the dataset is constructed as XML format which contains a tag for each review. Each review tag consists of review subject, review body, user rating and all extracted aspect categories with their start rating, associated opinion words, aspect terms and the polarity score for each opinion word. Figure 14 demonstrates a sample review as part of the final dataset output in XML format.


```xml
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<reviews>
    <review id="3662">
        <reviewbody>one of the worst apps. I would mention that is has very attractive UI and good theme.</reviewbody>
        <reviewStarRating>4</reviewStarRating>
        <aspects>
            <aspect name="User Experience">
                <aspectTerm OpinionWord="worst" term="apps" polarityStrength="-0.604166667"/>
            </aspect>
            <aspect name="User Interface">
                <aspectTerm OpinionWord="attractive" term="UI" polarityStrength="0.354166667"/>
                <aspectTerm OpinionWord="good" term="theme" polarityStrength="0.602678571"/>
            </aspect>
        </aspects>
    </review>
</reviews>
</xml>
```


Figure 14: *Sample Review in XML Format as Part of the Final Dataset*

The resulted dataset contains 7,345 reviews, each review contains an average of nine words. Table 9 and Table 10 demonstrate more statistics about the dataset.

Number of reviews	7345
Number of characters (No Spaces)	307,978
Number of characters (With Spaces)	366,797
Number of Words	67,777
Number of Explicit Aspects Terms	279
Number of Implicit Opinion-Aspects Terms	402
Number of Opinion Words	656
Average Words per Review	9.23

Table 9: *Dataset Content Statistics*

Number of Reviews	7345
Number of Very Positive Reviews	2001
Number of Positive Reviews	2392
Number of Neutral Reviews	1435
Number of Negative Reviews	1015
Number of Very Negative Reviews	502

Table 10: *Dataset Polarity Distribution Statistics*

Figure 15 represents one of the Cloud-based reports from GARSA and illustrates the final produced implicit and explicit aspects according to the constructed dataset and lexicons.

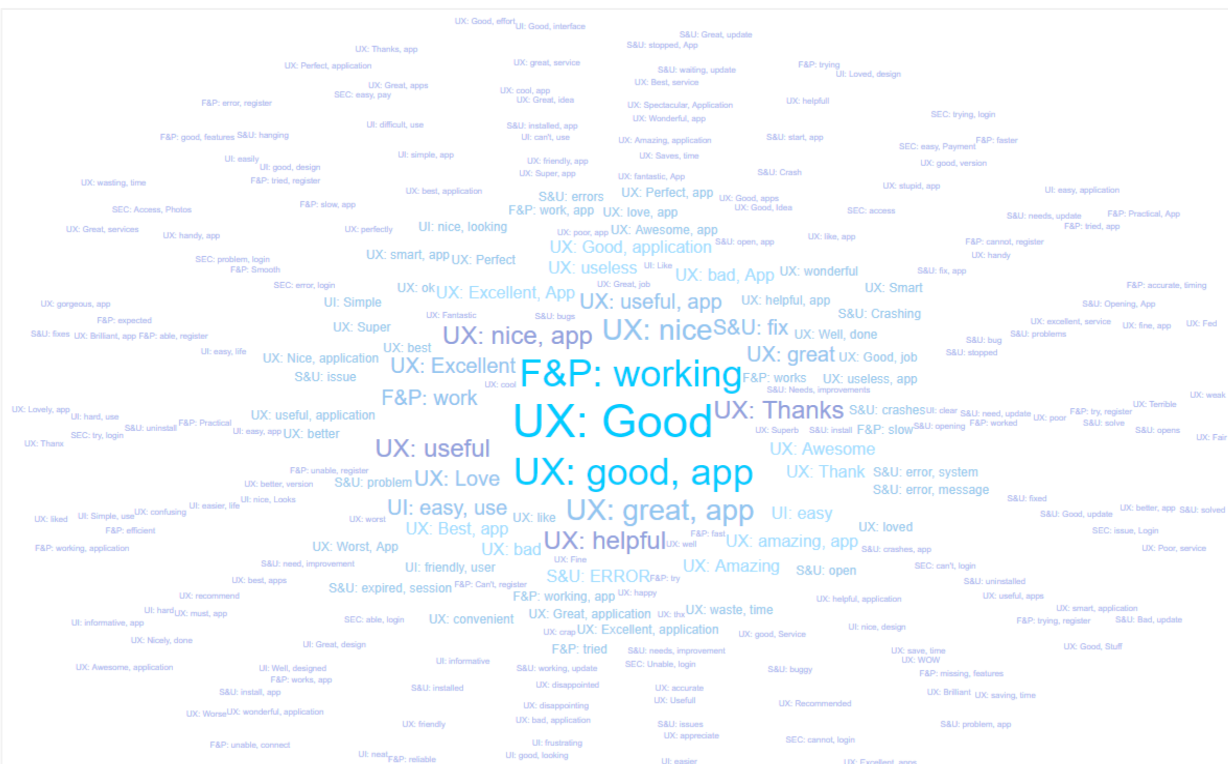

Figure 15: *Dataset Reviews Aspects vs Opinion Cloud*

4. Chapter Four: Integrated Lexicon and Rule-based Model for Aspect Extraction and Opinion Classification

Nowadays, governments are investing huge budgets in order to understand their customers and listen to their voice. It is crucial for government organizations to ensure that they provide the services that meet the needs and expectations of citizens. This can be done through exploring significant data sources that help in having a comprehensive customer centric framework. Online and smart channels are considered as one of the essential channels for providing the government services and obtaining the customers' feedback. Their feedback may include words of praise, defamation or problems encountered when using the service, words that express a unique experience, or proposal to improve this service, and others. It is important that all feedback opinions are understood and categorized so that governments can consider this channel to listen to the customers. Therefore, this can be considered as a factor for future smart services improvements and optimizations that exceeds the peoples' expectations.

This part of dissertation aims to summarize the methodology used to build an integrated lexicon and rule-based SA model to extract government apps reviews aspects and classify the corresponding sentiments. The approach relies on the prepared dataset as well as the various lexicons in chapter three. This model looks into the government mobile apps reviews from different angles in order to have more in-depth understanding of the customers' needs and expectations. Moreover, it aims to establish language processing techniques, rules, and lexicons to address several SA challenges and produce summarized results.

4.1. Methodology

According to section 3.3, the analysis of the collected reviews has revealed the following; short reviews in average, the frequent use of slangs, spelling mistakes, incorrect grammar, no spam reviews, and few occurrences of ineffective emoticons. This led to choosing an integrated lexicon and rule based model. This model utilized the manually generated lexicons in section 3.8 with hybrid rules to handle some of the key challenges in aspect-based SA in particular and SA in general. Shaalan (2010) highlighted the significance of the rule-based over other approaches as it depends on manually created rules and it is easier to integrate domain knowledge with NLP tasks. Figure 16 shows a summary of the followed approach to extract both explicit and implicit aspects in additions to its sentiment classification.

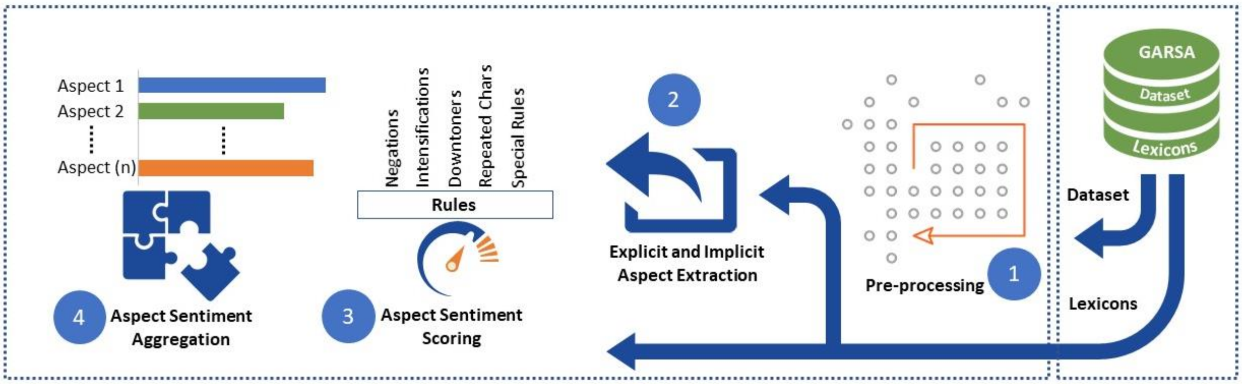

Figure 16: *Methodology for Integrated Lexicon and Rule-based Aspect Extraction*

4.2. Pre-processing

When adopting the rule-based approach with supportive lexicons, it is crucial to perform text pre-processing tasks. The first step in the proposed algorithm is to split the review into sentences based on punctuations that identify a sentence end such as the full-stop, question mark and exclamation mark. This would have an important impact on linking the polarity score with the right aspect term without interfering with irrelevant sentences. In addition, the review subject is added as the first sentence of the review. Next, the sentences are tokenized where for each token, punctuations are removed and all letters are converted to lowercase. However, the pre-processing tasks do not perform repeated characters removal as the aspect sentiment scoring tasks as explained in section 4.4 are responsible to treat it as intensification which affects the polarity score. For example, the proposed approach considers the word “greaaat” as “very great”. Finally, stop words will be marked to be out of any of the following tasks by using a customized list of stop words such as “the”, “an” and “of”. This list has been initiated by studying the domain and the reviews in the dataset.

4.3. Implicit and Explicit Aspects Extraction

One of the main challenges in aspect extraction task is that some aspects are not explicitly mentioned in a review. A sentence with explicit aspect basically contains a term that indicates the aspect category while an implicit aspect would not be expressed by any specific word or term. For example, a review with two sentences "*The app <u>design</u> is <u>attractive</u>. But it is <u>buggy</u>*", where the first part contains an explicit term "design" with an opinion word "*attractive*" that indicates a positive sentiment toward the aspect category "*User Interface*". On the other hand, the second sentence does not contain a term explicitly (Something like: "*The app <u>functionalities</u> are buggy*"). However, it contains an opinion word "*buggy*" which implicitly implies the aspect category "*Functionality and Performance*".

One of the approaches that are widely used in aspect identification is to consider opinion words as a good potential candidate for implicit aspect extraction (Poria et al. 2014). Thus, the designed algorithm first looks for opinion words that directly denote aspect according to the lexicon. Otherwise, if the opinion word cannot determine the aspect category, the algorithm will search for the nearest aspect term in the same sentence with maximum window size of two with more priority to the right side, since the adjective usually occurs before the term. The pair of identified opinion word and aspect term will be looked up in the lexicon in order to determine the aspect category.

In the first sentence of the previous example "*The app <u>design</u> is <u>attractive</u>*", the algorithm will identify "*attractive*" as an opinion word that cannot determine the aspect as per the lexicon, so it will look for the nearest aspect term in the same sentence which is "*design*" and determine the "*User Interface*" as aspect category through the lexicon. Conversely, in the second sentence "*But it is buggy*", as per the lexicon, "*buggy*" implies an opinion toward the aspect category "*Functionality and Performance*" even though the sentence does not contain any explicit aspect term. Similar approach was also adopted by Cruz-Garcia et al. (2014).

4.4. Aspect Sentiment Scoring

The approach that has been followed employs the populated lexicons in section 3.8. Basically, the algorithm navigates through the sentences and once an opinion word is identified, its polarity score will be retrieved through the lexicon and will be linked with the extracted aspect according to the section 4.3. In the experiment, several settings are applied to the algorithm in order to identify opinion words in a sentence in addition to the use of the lexicons. For instance, various rules are adopted to handle negations, intensification, downtoners, repeated characters, and the special case of negation-opinion rules.

As a baseline, the basic lexicon approach is followed. However, the polarity score retrieved from the sentiment lexicon is not final. There are several cases that can affect the polarity such as negation, intensification and downtoners. For each of these modifiers, a lexicon that lists different terms that implies a modification has been constructed. For instance, the term “not” implies a negation and in case it occurs near an opinion word as in the sentence “*The app design is not attractive*”, then the polarity score will be inverted through multiplying it by (-1). This case is called switch negation as it reverses the polarity of the opinion word. However, there is another negation case in which the polarity is not inverted. For example, in the following review: “*The app design is not innovative*”, the negation will not invert the polarity in this case since the strength of the opinion word “*innovative*” is highly positive. The algorithm will perform polarity shifting through multiplying it by certain threshold value (Threshold=0.5) instead of polarity reversal. In shifting negation, the proposed approach identifies the highly positive opinion word score that are greater than (0.75) and highly negative opinion word score that are less than (-0.75). In the previous example, “innovative” polarity score as per the lexicon is (1.0) which is greater than (0.75), so this value will be shifted to (0.375).

Table 11 shows examples for some of the negation words from the negation lexicon.

<table>
<tr><th>Negation Term</th><th>Example</th></tr>
<tr><td>Not</td><td>The app design is not attractive</td></tr>
<tr><td>Without</td><td>Works smooth without bugs. Great app.</td></tr>
<tr><td>Doesn’t</td><td>App doesn't work. It always shows error in login.</td></tr>
<tr><td>Never</td><td rowspan="2">Never able to login. Once you put in login details... That’s it... You can't open it again... Crashes.... Waste of time</td></tr>
<tr><td>Can’t</td></tr>
<tr><td>None</td><td rowspan="2">Old design, none responsive and non-interactive</td></tr>
<tr><td>Non</td></tr>
</table>

Table 11: *Samples that Show Some of the Negation Words*

The proposed algorithm treats the intensifiers and downtoners which can increase or decrease the polarity score of an opinion word when they appear in the sentence. For example, the sentence "*The app design is very attractive*" implies more positive sentiment than the sentence "*The app design is attractive*", so "*very*" considered as intensifier. On the other hand, the sentence "*The app design is quite attractive*" implies less positive sentiment, so the word "*quite*" is considered as downtoner. Hence, the algorithm modifies the polarity score when it locates any of the intensifiers or downtoners before the opinion word according to the lexicon. However, not all intensifiers or downtoners have the same impact or power. For instance, the word "*most*" denotes more intensification than the word "*very*". Similar to Taboada et al. (2011), the proposed approach assigned a multiplication factor to each intensifier and downtoner. For example, the polarity score of the opinion word that follows the intensifier "*very*" is multiplied by (1.25). On the other hand, the polarity score of the opinion word that follows the downtoner "*quite*" is multiplied by (0.75). Table 12 shows several intensifiers and downtoners along with their factor and examples.

Term	**Type**	**Factor**	**Example**
Extremely	Intensifier	2.0	Made me settle very quickly. *Extremely* reliable application
			Extremely slow loading and shows database connection error.
Absolutely	Intensifier	1.75	*Absolutely* useless app. The app is useless, never works properly.
			An *absolutely* useful and friendly application!
Quite	Downtoner	0.75	*Quite* useful app for Dubai residents.
			It's *quite* frustrating as I am unable to use the app
			Great interface. The app is *quite* slow but has good functionalities.
Pretty	Downtoner	0.50	*Pretty* good but some bugs

			Poorly designed app. *<u>Pretty</u>* much useless and probably just a media stunt.
Always	Intensifier	1.5	App doesn't work. It *<u>always</u>* shows error in login.
			New version *<u>Always</u>* crashes.

Table 12: *Sample Intensifiers and Downtoners from the Dataset*

Modifications can appear in different forms and can be embedded in the opinion word itself. For example, the suffix "est" as in "the greatest" is also considered as intensifier and it is equal to the word "most". On the other hand, people like to exaggerate their feelings by repeating letters as in "greaaaat" or "baaaaad" which is somehow equal to "very great" or "very bad". Hence, the algorithm identifies such cases by looking for repeated characters or specific suffixes and treat them as intensifiers. Table 13 provides the details of the complete set of Intensifiers and Downtoners with their Factors.

Term	**Type**	**Factor**	**Term**	**Type**	**Factor**
very	Intensifier	1.25	too	Intensifier	1.25
highly	Intensifier	1.75	absolutely	Intensifier	1.75
really	Downtoner	0.75	completely	Intensifier	1.5
more	Downtoner	0.50	most	Intensifier	1.75
always	Intensifier	1.5	mostly	Downtoner	0.75
many	Intensifier	1.25	pretty	Downtoner	0.75
fully	Intensifier	1.5	somewhat	Downtoner	0.5
so	Intensifier	1.25	slightly	Downtoner	0.75
lot	Intensifier	1.5	extraordinarily	Intensifier	2.0
full	Intensifier	1.5	ever	Intensifier	1.75
extremely	Intensifier	2.0	still	Downtoner	0.75

totally	Intensifier	1.25	just	Downtoner	0.75
little	Downtoner	0.25	barely	Downtoner	0.25
slightly	Downtoner	0.5	hardly	Downtoner	0.25

Table 13: *The Complete set of Intensifiers and Downtoners with their Factors*

4.5. Aspect Sentiment Aggregation

The algorithm targets to determine the star rating for different aspects extracted in the review. The five stars rating is chosen in the experiment where one star expresses a very negative sentiment toward this aspect, two stars express a negative sentiment, three stars express a neutral sentiment, four stars express a positive sentiment and five stars express a very positive sentiment. This can play a crucial role in understanding users' feedback toward specific aspects rather than a general feedback where the government apps owners can be aware of the areas of pains and gains of their customers.

Since the opinion words are extracted along with their final polarity scores and aspects, it would be a straight forward procedure to calculate the average of polarity scores for opinion words that are categorized under each aspect. Consider the below illustrative scenario:

- User Review: "*The app design is very attractive. It is organized and neat. But it is quite buggy and some features are missing*".
- The algorithm identified the following five sentiments as shown in Table 14.

#	Opinion Word	Aspect	Polarity Score	Modified By	Final Polarity
1	Attractive	User Interface	0.69	1.25	0.86
2	Organized	User Interface	0.47	-	0.47
3	Neat	User Interface	0.36	-	0.36
4	Buggy	Functionality and Performance	-0.5	0.75	-0.375
5	Missing	Functionality and Performance	-0.56	-	-0.56

Table 14: *The Extracted Aspects and Corresponding Sentiment Scores*

- The aggregation tasks will group polarities by the aspect categories and calculate the average. Here there are two aspect categories:
    1. User interface: (0.86 + 0.47 + 0.36) / 3 = 0.56
       Which is considered as 4 stars rating in terms of "User Interface".
    2. Functionality and Performance: (-0.375 + -0.56) / 2 = -0.47
       Which is considered as 2 stars rating in terms of "Functionality and Performance".

The final stage of the aspect-based SA is to summarize the full set of reviews in a graphical representation that shows the equivalent star rating of each aspect as shown in Figure 17.

Figure 17: *Aspect-based Sentiment Summarization for the Source Dataset*

4.6. Experiments Results

As explained in the previous sections, several attempts were conducted to evaluate the algorithm improvements on both aspects extraction and opinion classification. The experiments evaluation for aspect extraction and sentiment were carried out according to the several settings and parameters. Confusion matrix is employed to measure the performance. It includes Precision, Recall, F-measure and Accuracy metrics. The confusion matrix elements for both aspect extraction and sentiment classification are defined as illustrated in Table 15 and Table 16.

Actual (Annotated Dataset)	Predicted (Proposed Aspect Extraction Approach)		
		Retrieved	Not retrieved
	Relevant	**TP** (true positive) number of aspects that are correctly extracted	**FP** (false positive) number of aspects that are annotated, but the not extracted by the algorithm
	Irrelevant	**FN** (false negative) number of aspects that are not annotated, but extracted by the algorithm	**TN** (true negative) number of aspects that are not annotated and not extracted by the algorithm

Table 15: *Aspect Extraction Confusion Matrix*

Actual (Annotated Dataset)	Predicted (Proposed Sentiment Classification Approach)		
		Retrieved	Not retrieved
	Relevant	**TP** number of sentiment polarity scores that are calculated correctly by the algorithm	**FP** number of sentiment polarity scores that are incorrectly calculated by the algorithm
	Irrelevant	**FN** number of aspects that does not have sentiment assigned, but calculated by the algorithm	**TN** number of aspects that does not have sentiment assigned and not calculated by the algorithm

Table 16: *Sentiment Classification Confusion Matrix*

The corresponding performance measures are calculated according to the following formulas:

$$Precision = \frac{TP}{(TP + FP)}$$

$$Recall = \frac{TP}{(TP + FN)}$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FP)}$$

The aspect extraction results include a comparison of experiments that were conducted in this dissertation. As Table 17 shows, there is a significant increment in evaluation results when considering implicit aspects as well as the explicit ones. While analysing the algorithm progress in identifying the sentiments toward aspects without taking the implicit aspects into consideration, the algorithm was still able to identify opinion words and attempt to search for surrounding terms to recognize the aspect. Thus, it succeeded in aspect extraction in some cases especially when there is another opinion word with explicit aspect in the same sentence. However, in many cases a misleading aspect term appeared nearby the identified opinion word, where in this case the sentiment has been assigned to false aspect and the correct aspect has not been extracted. This may increase both FP and FN values and as a result it will affect both Precision and Recall.

	Performance Evaluation			
	Precision	**Recall**	**F-Measure**	**Accuracy**
Explicit Aspects	75.49%	70.64%	72.98%	86.83%
Explicit & Implicit Aspects	92.63%	84.03%	88.12%	93.01%

Table 17: *Aspect Extraction Performance Evaluation Results*

In a similar manner, several parameter settings have been employed to measure the progression of the aspects sentiment classification. As stated in section 4.4, the basic lexicon (L) based approach is considered as a baseline for the performance evaluation. Moreover, other rules settings are added to measure the improvements on the performance measures. These rules settings are adopted to handle numerous challenges in SA such as handling negation (N), intensification (I), downtoners (D), repeated characters (R) and special cases of negation-opinion rules (S). As shown in Table 18, the integrated lexicon and the various rules (L+N+I+D+R+S) has achieved the highest performance scores with respect to aspect extraction and opinion classification.

		Performance Evaluation			
		Precision	**Recall**	**F-Measure**	**Accuracy**
Experiment Parameters	**L (Baseline)**	73.66%	60.98%	66.72%	83.17%
	L+N	80.43%	65.84%	72.41%	85.89%
	L+N+I	83.42%	69.43%	75.79%	87.37%
	L+N+I+D	85.21%	72.15%	78.14%	88.57%
	L+N+I+D+R	86.17%	72.95%	79.01%	88.97%
	L+N+I+D+R+S	92.25%	79.59%	85.46%	91.64%

Table 18: *Aspect-based Sentiment Classification Performance Evaluation Results*

5. Chapter Five: Research Questions Answers

This chapter provides the answers to the research questions along with the reference to the corresponding parts in the dissertation.

5.1. Research Question 1

Question: What are the government smart applications aspects that can be used in aspect-based SA?

Answer: According to sections 2.7, 2.8 and 3.4, it was possible to define government smart applications aspects that can be used in aspect-based SA. These aspects were defined based on the most widely used smart phones operating systems which are iOS (developed by Apple) and Android (developed by Google Inc.) as well as Smart Dubai Government SWEM with emphasis on customer experience as an important aspect according to the users' perspective. The proposed model consists of five main aspects namely; User Interface, User Experience, Functionality and Performance, Security, and Support and Updates.

5.2. Research Question 2

Question: How can we produce a domain-specific annotated sentiment dataset and lexicon that targets the government mobile apps?

Answer: Chapter three illustrates the work that has been done in order to produce a unique dataset that consists of government mobile apps domain aspects and opinion words. Additionally, it explains the methodology that was carried out to assign the sentiment scores to opinion words and build the desired lexicons. This was done with the help of the GARSA which is a responsive web tool that has been designed to facilitate the annotation process in a flexible, organized, efficient and tracked way. The approach started with collecting the data, analysing the data, designing the database, defining the domain specific aspects, building GARSA, annotating the reviews, and producing the anticipated dataset and lexicons. The characteristics of the produced dataset are shown in Table 9 and Table 10.

5.3. Research Question 3

Question: How can we improve the performance of aspect extraction and sentiment classification through an integrated lexicon and rule-based model?

Answer: Yes, chapter four explains the details of the employed approach in which it was possible to produce high performance results through an integrated lexicon and rule-based model. The approach confirmed that integrating sentiment and aspects lexicons with various rules settings that handle various challenges in SA such as handling negation, intensification, downtoners, repeated characters and special cases of negation-opinion rules outperformed the lexicon baseline and other rules combinations as stated in Table 17.

6. Chapter Six: Conclusion and Future Prospects

In this chapter, a conclusion of the work that was carried throughout this dissertation work as well as the proposed future work.

6.1. Conclusion

This study aimed to produce a novel domain specific annotated dataset that involves government apps in the UAE as well as its corresponding aspects terms and opinion lexicons. Additionally, it aimed to come up with an integrated aspect-based model for the government apps. This model resulted in a more in-depth understanding of the customers' needs and expectations through looking into the app reviews from different angles in addition to previous models of assessing and evaluating the government apps based on written standards.

In this dissertation, a review of credible resources was conducted to determine the findings of previous studies on SA. These resources were used to develop information about basic terms in SA and its applications. This provides a solid background information on SA that covers the SA basics and levels of sentiment classification. Additionally, the SA approaches and techniques details to fulfil SA tasks were reviewed according to the literature. The SA techniques have been employed by several models that can be classified into two main approaches namely ML approaches for SA (Pang & Lee 2008) and SO approaches for SA (Dang, Zhang & Chen 2010). Moreover, Medhat, Hassan and Korashy (2014) stated a third approach that can be also utilized in SA which is the hybrid approach that combines both ML and SO approaches. The hybrid approach is very common when sentiment lexicons are adopted with other techniques. This dissertation also discusses the most recent work done on aspect-based SA field. According to Asghar et. al (2014) and Ganeshbhai and Shah (2015), aspect-based SA phases comprise aspect or feature extraction, sentiment polarity prediction and classification, and sentiment aggregation. Various techniques for constructing resources for SA that presents significant contributions in improving various tasks were also reviewed. It is vital to discover the suitable resources in order to conduct the desired experiments. For most cases, researchers use known publicly available datasets (Poria et al. 2014; Liu et al. 2013; Xianghua et al. 2013; Mukherjee & Liu 2012; Poria, Cambria & Gelbukh 2016).

In other cases, it could be difficult to find a dataset in specific domain or in specific area. Tan and Wu (2011) confirmed that domain dependent sentiment dictionaries are essential to measure SA tasks performance. Joshi, Bhattacharyya and Ahire (2017) stated that the annotated dataset comprises of two main components; the textual part and its labels. The textual part may consist of manually annotated reviews with its corresponding sentiment orientation called labels.

One of the most important components in aspect-based SA in the government mobile applications is to have a clear definition of the aspects that need to be extracted. This was one of the key contributions of this dissertation. The government smart applications aspects that can be used in aspect-based SA were defined based on the most widely used smart phones operating systems which are iOS (developed by Apple) and Android (developed by Google Inc.) as well as Smart Dubai Government SWEM with emphasis on customer experience as an important aspect according to the users' perspective. The proposed model consists of five main aspects namely; User Interface, User Experience, Functionality and Performance, Security, and Support and Updates.

Another key contribution of this study is the unique government mobile apps domain specific dataset and lexicons that have been produced. The final produced dataset that consists of government mobile apps domain aspects and opinion words. This was done with the help of the GARSA which is a responsive web tool that has been designed to facilitate the annotation process in a flexible, organized, efficient and tracked manner. The approach started with collecting the data, analysing the data, designing the database, defining the domain specific aspects, building GARSA, annotating the reviews, and producing the anticipated dataset and lexicons. It is important to point out that GARSA can be also used for annotation process in other domains.

Aspect-based SA is considered as one of the challenging tasks in SA. It is important that all feedbacks opinions are understood and categorized so that governments can consider this channel to listen to the customers. Therefore, this can be considered as a factor for future smart services improvements and optimizations that exceed the peoples' expectations. In this regard, an integrated lexicon and rule-based approach was employed to extract explicit and implicit aspect as well as sentiment classification for these aspects. It has been noticed according to the analysis of the collected data that the reviews are short in average, slangs are frequently used, contains spelling mistakes and incorrect grammar, spam reviews were not present, and few occurrences of ineffective emoticons. For these reasons, an integrated lexicon and rule based model has been chosen for this study. This model utilized the manually generated lexicons in this study with hybrid rules to handle some of the key challenges in aspect-based SA in particular and SA in general. This approach reported high performance results through an integrated lexicon and rule-based model. The approach confirmed that integrating sentiment and aspects lexicons with various rules settings that handle various challenges in SA such as handling negation, intensification, downtoners, repeated characters and special cases of negation-opinion rules outperformed the lexicon baseline and other rules combinations.

Publics' opinion is very important to increase the usage of the government mobile apps. It is also important to achieve a customer centric government framework. People are keen to look at others' opinions before downloading and using the smart application. Additionally, the reviews represent a major source for apps owners to understand the user's perspectives and hence, provide services that exceed the customers' expectations and improve these services continuously. Government mobile apps owners should take a step forward and think how to gather their customers reviews and feedbacks. One quick and easy way is to prompt for providing a review within the mobile app. Also, a rewarding program could be beneficial to increase the customers reviews towards this particular application.

6.2. Future Prospects

As part of the proposed work, an integrated model of lexicon and rule-based was proposed. This model comprises of lexicons, negation, intensification, downtoners, repeated characters and special cases of negation-opinion rules. The collected reviews contain a substantial number of comparable sentences. This can be considered as additional rules that can be added to the proposed model. Xu et al. (2011) pointed out that users may express their opinions by comparing several entities together. These comparative rules are considered as a significant source for determining the strengths and weaknesses of the government mobile apps and highlighting the future enhancements and improvement areas. Alternatively, it is possible to explore various approaches to extract aspects with its opinion orientation such as various ML algorithms (SVM and Naïve Bayes among others).

There is a possibility to expand the produced dataset to cover more languages with focus on Arabic language. As one of the options, the dataset and produced lexicons can be translated into Arabic as well as other languages. According to the initial collected reviews, 20% of the collected data were written in Arabic language, there is a room to build annotated dataset for this part of the collected reviews. This can be achieved with the help of GARSA. Moreover, a comparison between the Arabic translated dataset and the manually annotated dataset can be conducted.

Another area that can be applied on such dataset, is to identify the improvement areas and innovative ideas through analysing the customer's reviews which may include suggestions, complaints, problems, bugs, comparisons among others. This can be done by following the same manual approach to build a special lexicon for such terms that fulfil this purpose.

Finally, a comprehensive tool can be built to analyse any dataset for any government mobile app reviews in order to evaluate the customer's satisfaction towards this mobile app as well as identifying the weak aspects in it. Such tool will allow the government mobile app owners to visualize the app reviews SA results and provide them with accurate indication of what their customers feel towards the app.

7. References

Agarwal, B., & Mittal, N. (2013, June). Sentiment classification using rough set based hybrid feature selection. In *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis (WASSA'13), NAACL-HLT. Atlanta* (pp. 115-119).

Al Suwaidi, H., Soomro, T. R., & Shaalan, K. (2016). Sentiment Analysis for Emiriti Dialects in Twitter. *Sindh University Research Journal-SURJ (Science Series)*, *48*(4).

Alqaryouti, O., Alqudah, A. & Shaalan, K. (2016). Effective Knowledge Management in Projects: Characteristics and Integration. *Information and Knowledge Management*, vol. 6 (8), pp. 27-33.

Apple. (2017a). "App Store Review Guidelines - Apple Developer". *Developer.apple.com* [online]. [Accessed 3 January 2017]. Available at: https://developer.apple.com/app-store/review/guidelines/

Apple. (2017b). "Themes - Overview - iOS Human Interface Guidelines". *Developer.apple.com* [online]. [Accessed 5 January 2017]. Available at: https://developer.apple.com/ios/human-interface-guidelines/overview/design-principles/

Android. (2017). "Core app quality | Android Developers". *Developer.android.com* [online]. [Accessed 15 January 2017]. Available at: https://developer.android.com/develop/quality-guidelines/core-app-quality.html

Asghar, M. Z., Khan, A., Ahmad, S., & Kundi, F. M. (2014). A review of feature extraction in sentiment analysis. *Journal of Basic and Applied Scientific Research*, *4*(3), 181-186.

Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on* (Vol. 1, pp. 492-499). IEEE.

Barile, S., & Polese, F. (2010). Smart service systems and viable service systems: Applying systems theory to service science. *Service Science*, *2*(1-2), 21-40.

Blanco, E., & Moldovan, D. (2014). Retrieving implicit positive meaning from negated statements. *Natural Language Engineering*, *20*(4), 501-535.

Boldrini, E., Balahur, A., Martínez-Barco, P., & Montoyo, A. (2012). Using EmotiBlog to annotate and analyse subjectivity in the new textual genres. *Data Mining and Knowledge Discovery*, *25*(3), 603-634.

Carenini, G., Cheung, J. C. K., & Pauls, A. (2013). MULTI-DOCUMENT SUMMARIZATION OF EVALUATIVE TEXT. *Computational Intelligence*, *29*(4), 545-576.

Chen, Zhiyuan, Mukherjee, Arjun, & Liu, Bing (2014). Aspect Extraction with Automated Prior Knowledge Learning. *ACL*.

Dang, Y., Zhang, Y., & Chen, H. (2010). A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems*, *25*(4), pp. 46-53.

Deloitte. (2015). "Mobile Consumer Survey 2015 | Deloitte Australia | TMT, mobile devices, social media, internet speed, online shopping". *Deloitte Australia* [online]. [Accessed 8 October 2016]. Available at: http://www2.deloitte.com/au/en/pages/technology-media-and-telecommunications/articles/mobile-consumer-survey-2015.html

Esuli, A., & Sebastiani, F. (2006). SENTIWORDNET: A high-coverage lexical resource for opinion mining. *Institute of Information Science and Technologies (ISTI) of the Italian National Research Council (CNR)*.

Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., & Sadeh, N. (2013, August). Why people hate your app: Making sense of user feedback in a mobile app store. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1276-1284). ACM.

Ganeshbhai, S. Y., & Shah, B. K. (2015, June). Feature based opinion mining: A survey. In *Advance Computing Conference (IACC), 2015 IEEE International* (pp. 919-923). IEEE.

Gartner. (2017). "Gartner Says Worldwide Sales of Smartphones Grew 7 Percent in the Fourth Quarter of 2016". *Gartner.com* [online]. [Accessed 5 March 2017]. Available at: http://www.gartner.com/newsroom/id/3609817

Gil-Garcia, J. R., Helbig, N., & Ojo, A. (2014). Being smart: Emerging technologies and innovation in the public sector. *Government Information Quarterly*, *31*, I1-I8.

Gil-Garcia, J. R., & Martinez-Moyano, I. J. (2007). Understanding the evolution of e-government: The influence of systems of rules on public sector dynamics. *Government Information Quarterly*, *24*(2), 266-290.

GulfNews. (2013a). "Dubai e-Government renamed Dubai Smart-Government". *GulfNews.com* [online]. [Accessed 1 November 2016]. Available at: http://gulfnews.com/business/economy/dubai-e-government-renamed-dubai-smart-government-1.1199985

GulfNews. (2013b). "Hamdan launches smart government programme". *GulfNews.com*[online]. [Accessed 9 October 2016]. Available at: http://gulfnews.com/business/economy/hamdan-launches-smart-government-programme-1.1199546

Harsh, A., & Ichalkaranje, N. (2015). Transforming e-Government to Smart Government: A South Australian Perspective. In *Intelligent Computing, Communication and Devices* (pp. 9-16). Springer India.

Hasbullah, S. S., Maynard, D., Chik, R. Z. W., Mohd, F., & Noor, M. (2016, January). Automated Content Analysis: A Sentiment Analysis on Malaysian Government Social Media. In *Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication* (p. 30). ACM.

Hatzivassiloglou, V., & McKeown, K. R. (1997, July). Predicting the semantic orientation of adjectives. In Proceedings of *the eighth conference on European chapter of the Association for Computational Linguistics* (pp. 174-181). Association for Computational Linguistics.

Himmat, M., & Salim, N. (2014). Survey on product review sentiment classification and analysis challenges. In *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)* (pp. 213-222). Springer Singapore.

Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.

Internet Live Stats. (2016). "Twitter Usage Statistics - Internet Live Stats". *Internetlivestats.com* [online]. [Accessed 8 October 2016]. Available at: http://www.internetlivestats.com/twitter-statistics/

Joshi, A., Bhattacharyya, P., & Ahire, S. (2017). Sentiment Resources: Lexicons and Datasets. In *A Practical Guide to Sentiment Analysis* (pp. 85-106). Springer International Publishing.

KhaleejTimes. (2014). "Dubai Municipality implements happiness index to evaluate services". *Khaleejtimes.com* [online]. [Accessed 8 October 2016]. Available at: http://www.khaleejtimes.com/nation/government/dubai-municipality-implements-happiness-index-to-evaluate-services

Khan, K., Baharudin, B., Khan, A., & Ullah, A. (2014). Mining opinion components from unstructured reviews: A review. *Journal of King Saud University-Computer and Information Sciences*, *26*(3), 258-275.

Kim, S. M., & Hovy, E. (2004, August). Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics* (p. 1367). Association for Computational Linguistics.

Kundi, F. M., Khan, A., Ahmad, S., & Asghar, M. Z. (2014). Lexicon-based sentiment analysis in the social web. *Journal of Basic and Applied Scientific Research*, *4*(6), 238-48.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, vol. *5*(1), pp. 1-167.

Liu, B., Hu, M., & Cheng, J. (2005, May). Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web* (pp. 342-351). ACM.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data* (pp. 415-463). Springer US.

Liu, Q., Gao, Z., Liu, B., & Zhang, Y. (2013). A logic programming approach to aspect extraction in opinion mining. In *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on* (Vol. 1, pp. 276-283). IEEE.

Liu, Q., Gao, Z., Liu, B., & Zhang, Y. (2015, July). Automated Rule Selection for Aspect Extraction in Opinion Mining. In *IJCAI* (pp. 1291-1297).

Liu, Q., Liu, B., Zhang, Y., Kim, D. S., & Gao, Z. (2016, March). Improving opinion aspect extraction using semantic similarity and aspect associations. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (Vol. 1, No. 1, p. 496). Cambridge: Cambridge university press.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), pp. 1093-1113.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, *38*(11), 39-41.

Mills, A. (2016). *Certified Innovation Strategist*. 1st edn. USA:Global Innovation Institute.

Montoyo, A., MartíNez-Barco, P., & Balahur, A. (2012). Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments.

Mukherjee, A., & Liu, B. (2012, July). Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1* (pp. 339-348). Association for Computational Linguistics.

O'reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & strategies*, vol.(1), pp.17.

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

Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, *108*, 42-49.

Poria, S., Cambria, E., Ku, L. W., Gui, C., & Gelbukh, A. (2014, August). A rule-based approach to aspect extraction from product reviews. In *Proceedings of the second workshop on natural language processing for social media (SocialNLP)* (pp. 28-37).

Ray, S. K., & Shaalan, K. (2015, December). Using Freeware Resources to Analyse Sentiments in Social Media. In *Developments of E-Systems Engineering (DeSE), 2015 International Conference on* (pp. 204-209). IEEE.

Robaldo, L., & Di Caro, L. (2013). Opinionmining-ml. *Computer Standards & Interfaces*, *35*(5), 454-469.

Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017). A survey of text mining in social media: Facebook and Twitter perspectives. *Adv. Sci. Technol. Eng. Syst. J*, *2*(1), 127-133.

Samha, A. K., Li, Y., & Zhang, J. (2014). Aspect-based opinion extraction from customer reviews. *arXiv preprint arXiv:1404.1982*.

Sarawagi, S. (2008). Information extraction. *Foundations and Trends® in Databases*, *1*(3), 261-377.

Shaalan, K. (2010). Rule-based approach in Arabic natural language processing. *The International Journal on Information and Communication Technologies (IJICT)*, *3*(3), 11-19.

Sharma, R., Nigam, S., & Jain, R. (2014). Mining of product reviews at aspect level. *arXiv preprint arXiv:1406.3714*.

Siqueira, H., & Barros, F. (2010, October). A feature extraction process for sentiment analysis of opinions on services. In *Proceedings of International Workshop on Web and Text Intelligence* (pp. 404-413).

Smart Dubai Government. (2016). "Smart Websites Excellence Model Version 4.0". *Smart Dubai Government* [online]. [Accessed 3 January 2017]. Available at: http://dsg.gov.ae/en/OurPublications/Pages/StandardsGuidelines.aspx

Sporleder, C., & Lascarides, A. (2008). Using automatically labelled examples to classify rhetorical relations: An assessment. *Natural Language Engineering*, *14*(3), 369-416.

Statista.com. (2016). *Statista - The Statistics Portal for Market Data, Market Research and Market Studies*. [online] Available at: https://www.statista.com/ [Accessed 11 Dec. 2016].

Steinberger, J., Ebrahim, M., Ehrmann, M., Hurriyetoglu, A., Kabadjov, M., Lenkova, P., ... & Zavarella, V. (2012). Creating sentiment dictionaries via triangulation. *Decision Support Systems*, *53*(4), 689-694.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), 267-307.

Takala, P., Malo, P., Sinha, A., & Ahlgren, O. (2014). Gold-standard for Topic-specific Sentiment Analysis of Economic Texts. In *LREC* (Vol. 2014, pp. 2152-2157).

Tan, S., & Wu, Q. (2011). A random walk algorithm for automatic construction of domain-oriented sentiment lexicon. *Expert Systems with Applications*, *38*(10), 12094-12100.

Wang, T., Cai, Y., Leung, H. F., Lau, R. Y., Li, Q., & Min, H. (2014). Product aspect extraction supervised with online domain knowledge. *Knowledge-Based Systems*, vol. *71*, p. 86-100.

Wawre, S. V., & Deshmukh, S. N. (2016). Sentiment Classification using Machine Learning Techniques. *International Journal of Science and Research (IJSR)*, 5(4), pp.819-821.

Weitzel, L., Prati, R. C., & Aguiar, R. F. (2016). The comprehension of figurative language: what is the influence of irony and sarcasm on NLP techniques?. In *Sentiment Analysis and Ontology Engineering* (pp. 49-74). Springer International Publishing.

Xianghua, F., Guo, L., Yanyan, G., & Zhiqiang, W. (2013). Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems*, *37*, pp. 186-195.

Xu, K., Liao, S. S., Li, J., & Song, Y. (2011). Mining comparative opinions from customer reviews for Competitive Intelligence. *Decision support systems*, *50*(4), 743-754.

Zabin, J., & Jefferies, A. (2008). Social media monitoring and analysis: Generating consumer insights from online conversation. *Aberdeen Group Benchmark Report*, *37*(9).

Zephoria Inc. (2016). "The Top 20 Valuable Facebook Statistics". *Zephoria.com* [online]. [Accessed 8 October 2016]. Available at: https://zephoria.com/top-15-valuable-facebook-statistics/