Towards a Unified Arabic Government Services Chatbot Based on Ontology

نحو روبوت محادثة عربي موحد للخدمات الحكومية يعتمد على الأنطولوجيا

by

SUFYAN JAMAL SAID AREED

Dissertation submitted in fulfilment

of the requirements for the degree of

MSc INFORMATICS

at

The British University in Dubai

September 2020

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that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.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

The amount of information has risen considerably since the beginning of the electronic era. Meanwhile, the relationships between various types of information have become more and more complex. On the other hand, the growing number of users have encouraged researchers to take an advantage of this information and develop techniques which analyze the experiences of customers to accommodate their requirements and satisfy their needs. Since the birth of the term "e-government", the amount of structured and unstructured information has increased dramatically, which has forced all government entities worldwide to provide many call centers with long working hours that may reach to 24/7 in some entities to answer users' questions related government services, regardless of the time wasted by the clients while waiting to get in touch with the agent. In fact, UAE government is not an exception. As far as Dubai government is concerned, some entities have taken a step forward by introducing their own chatbot technology to respond to customer inquiries. However, the challenges fraught with the chatbot development in general in addition to the challenges associated with the Arabic language in particular, made it extremely difficult to design a unified chatbot that is capable to respond to all services provided by Dubai government especially when dealing with the Arabic language. In this study, a novel approach is proposed to extract an Arabic language knowledge base from a previously built ontology for more than 500 services provided by the Dubai government in order to use the extracted knowledge base into a chatbot application through Artificial Intelligence Markup Language (AIML) files. The current ontology is an enhancement which builds on a previously created ontology. Furthermore, a chatbot response algorithm is proposed to respond to government services queries through a hybrid of three different approaches that were executed in a pipeline fashion based on the query complexity. The feasibility of the proposed algorithm has been proven through executing multiple experimental tests for the same set of questions (414 questions) that were performed on the ontology earlier, then it has been compared with the ontology itself and the formal chatbot that has been designed by Smart Dubai Government (Rashid) chatbot. High Accuracy score has been achieved by the proposed algorithm that reached 96% with Recall of 100% and Precision of 96% as well. These results confirmed that the performance of the proposed algorithm could outperform both a previously developed chatbot-based on ontology and Rashid chatbot as well.

Keywords: Ontology Knowledge Base Extraction, Arabic AIML Files Construction, Arabic Government Services Chatbot.

الملخص

إزداد حجم المعلومات بشكل كبير منذ بداية العصر الإلكتروني. وفي الوقت نفسه، أصبحت العلاقات بين أنواع المعلومات المختلفة أكثر تعقيدًا. من ناحية أخرى ، ادى الإزدياد في عدد المتعاملين الى تشجيع الباحثين على الاستفادة من هذه المعلومات بهدف تطوير تقنيات لتحليل تجربة المتعامل لتلائم متطلباته وتلبي احتياجاته. منذ ولادة مصطلح "الحكومة الإلكترونية" في دبي، وجد العديد من المعلومات المنظمة وغير المنظمة بشكل كبير، مما أجبر جميع الجهات الحكومية على توفير العديد من مراكز الاتصال بساعات عمل طويلة قد تصل إلى 7/24 في بعض الجهات للإجابة على أستفسارات المتعاملين المتعلقة بالخدمات الحكومية، هذا بغض النظر عن الوقت الضائع من قبل العملاء أثناء الانتظار للتواصل مع الوكيل. اتخذت بعض الجهات الحكومية خطوة إلى الأمام من خلال تقديم روبوت محادثة خاص بها للرد على تلك الاستفسارات. ومع ذلك ، فإن التحديات التي تواجه تطوير روبوتات الدردشة بشكل عام بالإضافة إلى التحديات المرتبطة باللغة العربية بشكل خاص، جعلت من الصعب للغاية تصميم روبوت محادثة موحد قادر على الاستجابة لجميع الخدمات التي تقدمها حكومة دبي، وخاصة عند التعامل مع اللغة العربية. في هذه الدراسة، تم اقتراح اسلوب جديد لاستخراج قاعدة المعرفة العربية من الأنطولوجي التي تم بنائها مسبقًا لأكثر من 500 خدمة مقدمة من حكومة دبي لكي يتم استخدامها في تطبيق روبوت محادثة عبر ملفات (AIML). علاوة على ذلك، تم اقتراح خوارزمية استجابة للروبوت للرد على استفسارات الخدمات الحكومية من خلال نظام هجين يجمع ثلاثة أساليب مختلفة سيتم تنفيذها بالتتابع بناءً على مدى صعوبة الاستفسار. ولقد تم إثبات جدوى الخوارزمية المقترحة من خلال تنفيذ اختبارات تجريبية متعددة لنفس مجموعة الأسئلة (414 سؤالًا) التي تم طرحها على الأنطولوجيا في وقت سابق، ثم تمت مقارنة النتائج مع الأنطولوجي نفسها بالإضافة الى مقارنتها مع روبوت محادثة موحد اخر قد تم تصميمه من قبل حكومة دبي الذكية (راشد). تم تحقيق درجة دقة عالية من خلال الخوارزمية المقترحة والتي وصلت إلى 96٪ مع استعادة وصلت الى نسبة 100٪. أكدت هذه النتائج أن الخوارزمية المقترحة يمكن أن تتفوق على كل من الأنطولوجي و روبوت المحادثة راشد أيضًا.

Dedication

This study is wholeheartedly dedicated to my beloved parents who have been the source of strength and inspiration, and who passed away with their dream to see me in the best positions.

To my wife and sons who were very patient while studying and completing this work by offering the required atmosphere to focus more.

To my supervisor, brothers and dear friends who showed a great support and encouragement throughout this journey.

Acknowledgement

First and foremost, I thank Allah Almighty for His guidance and support to complete this dissertation. As I would like to acknowledge everyone who played a role in my academic accomplishments.

I would like to express my sincere gratitude to my advisor Prof. Khaled Shaalan who was always there to provide the required motivation whenever needed.

Besides my advisor, I would like to thank my dear friends Dr. Omar Alqaryouti and Mr. Ali Al Barghothi for their endless support and insightful comments in addition to their encouragement.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

IR	Information Retrieval
QAS	Question Answering System
NLP	Natural Language Processing
POS	Part of Speech
NER	Named Entity Recognition
SW	Semantic Web
DL	Description Logic
AIML	Artificial Intelligence Markup Language
MSA	Modern Standard Arabic
EC	Embodied Chatbot
LC	Linguistic Chatbots
TDS	Textual Dialog Systems
SDS	Spoken Dialog Systems
NN	Neural Network
WSD	Word Sense Disambiguation
SUMO	Suggested Upper Merged Ontology
SSSN	Sentence Similarity based on Semantic Networks
LSA	Latent Semantic Analysis
AWN	Arabic WordNet
SCAF	Semantic Conversational Agent Framework
ASD	Autism Spectrum Disorder
CBT	Cognitive Behavioral Therapy
KDD	Knowledge Discovery in Database
AEDS	Automatically Extract Dataset System

1. CHAPTER 1: INTRODUCTION

1.1. Background

Public users mainly rely on governments' websites to get an answer about an inquiry they have or a service they wish to apply for in a specific entity. Many users may look at search engines (e.g. Google, Bing, Yahoo… etc.) to ask their questions if they don't know the entity's website or service portal. Therefore, the answers usually will be retrieved in a form of a web page or document that consists of many information and might be pointing to some other page that has the required information.

A more advanced approach of Information Retrieval (IR) is by using question answering systems (QAS) (Nakov et al. 2019). The concept of QAS is to pose questions using human language and analyze the question using Natural Language Processing (NLP) by applying linguistic analysis which aims to understand the user's query and relate it to a suitable source of information (Ezzeldin 2012). QAS allows users to acquire the correct answer after obtaining and verifying the possible answers from different sources (Yassine Benajiba, Paolo Rosso, Lahsen Abouenour & Karim Bouzoubaa 2014). Noy and McGuinness (2001) pointed out that NLP has multiple functionalities to manipulate the text e.g. Tokenization, Stemming, Part of Speech (POS) tagging, and Named Entity Recognition (NER), which will help in retrieving the right answer from a huge amount of structured and unstructured data. Albarghothi, Khater and Shaalan (2017) highlighted that the best way to retrieve a correct answer is through a structured data source. However, unstructured data can be treated using NLP techniques in order to achieve accurate answers. Ray and Shaalan (2016) noted various kinds of QAS, like online and offline systems. Such systems have a similar architecture, with some variances in the knowledge base, domain, and IR. There are different systems and approaches to the question-and-answer task, so it is important to have a generic architecture and design. Specific selections systems can distinguish the question-and-answer architecture for representing and handling the general parts of the model. According to Allam and Haggag (2012), there are three main phases to question-and-answer, the first phase is question processing, which handles question analysis, question classification, and question reformulation. The second phase is document processing, which is concerned about information retrieval, paragraph filtering, and paragraph order. The third phase is answer processing that will identify the answer, then extract the answer and finally validate it.

Following the introduction of the Semantic Web (SW) concept for organizing information as a structured data source, SW became a convenient solution for knowledge representing and reasoning (Al-Chalabi, Ray & Shaalan 2016). Among a set of Description Logic (DL) within the real world, ontology represents the main concepts, interactions, and classes. On top of that, ontology is very suitable for the user's arithmetical analysis, but ontology's structures could be understood in a poorly insightful form, e.g. classes, relationships, and properties. This applies often to users who have more expertise in the ontology domain when evaluating conceptual data or updating the data source of the knowledge (Albarghothi 2018). People may therefore wish to obtain more knowledge about natural language in ontology (Gyawali 2011). Jain and Singh (2013) emphasized the need for SW to strengthen the knowledge network by constructing an information structure to enrich the web repository.

Due to a high-tech transition occurring in the world, most governments rely on information and technologies to carry out their day-to-day work and bring their plans into action. Several government researches have been conducted to determine the principles of e-government and to help them create their own model (Gil-Garcia & Martinez-Moyano 2007). To deliver services that meet the needs and expectations of customers, governments have to invest in their infrastructure and Information Technology (IT) in addition to communication (Bekkers 2003). According to Gil-Garcia and Martinez-Moyano (2007), there are seven stages that can help in implementing e-government and identify the benefits on both levels, organizational and technological. These stages can be summarized in initial presence, extended presence, interactive presence, transactional presence, vertical integration, horizontal integration, and finally totally integrated presence.

Moreover, all government entities in the United Arab Emirates (UAE) aims to automate their public services through the implementation of the latest technologies and the development of smart services on mobile devices. This was due to the directives of His Highness Shaikh Mohammad bin Rashid Al Maktoum, Vice-President and Prime Minister of the UAE and Ruler of Dubai to implement the e-government in 2001 (ArabianBusiness 2001), followed with the announcement of launching m-government which also known as the smart government in 2013 (UAE Vice President 2013). Consequently, many Key Performance Indicators (KPIs) took place since then to monitor all government entities and award whoever achieves a higher score in enabling all services online and help their customers in adopting these modern channels. The implementation of "e" and "m" governments played an important role in identifying the customer's happiness and satisfaction by gathering their feedback after applying for any service (GulfNews 2015).

The government's information is spread through various systems and technologies that affect the process of retrieving the information. It can also cause major challenges in policies and regulations and impacts directly on the integration and interoperability (Alazemi, Al-Shehab & Alhakem 2017). In addition, the main purpose of e-government is to enable people to access public services and adequate information through efficient browsing and execution of services (Gil-Garcia & Martinez-Moyano 2007). For that reason, service integration among multiple government entities becomes necessary in order to achieve a totally integrated presence.

Governments are the main knowledge body of services, so these services must be well organized and easily accessible by all citizens (Areed, Salloum & Shaalan 2020). Charalabidis and Metaxiotis (2009) emphasized that the main classifications of knowledge within e-government, public administration in addition to systems and applications will satisfy the needs of citizens through interaction with services. There are three phases for the knowledge management application framework of e-government: the first phase is Publish, where information technology processes are applied to obtain access to e-government information without the need for physical presence in the government departments to avoid long waiting queues. The second phase is Interact, which aims to increase public engagement in government decisions such as meetings with locals and residents and elect representatives in order to address a range of concerns and to create a discussion forum. The third phase is Transact, which seeks to simplify all government processes by automation so that all citizens can easily access them online from any place. For instance, automating tax collection procedures may reduce corruption and boost government trust. In addition, these procedures may increase public and private efficiency.

Chatbots are usually used in conversation related to services and information retrieval. Few chatbots have been developed using systems of complex natural language processing. A simpler design of chatbot is to relay on keywords matching. Many names have been introduced to the chatbot, such as conversational agents, talkbots, interactive agents, and many more (Dias et al. 2019). Androutsopoulou et al. (2019) pointed out that government departments face various challenges to enforce the implementation of Chatbot. Such challenges can be associated with the extraction and representation of the expertise needed to develop their knowledge base while developing Artificial Intelligence (AI) applications. However, the use of Chatbot in e-government would assist the government in collecting information from multiple resources and improve customer satisfaction by enabling instant inquiries response.

1.2. Problem Definition

The services rendered by the government are usually complex and have several criteria to complete the service request. On the other hand, consumers of the services have several questions/ inquiries to be answered by domain experts (Schwarzer et al. 2016). Therefore, customers must study and understand the explanation of the services to acquire some of the details they need on their own. However, Chatbot can play an important role by answering direct e-government-related questions instead of retrieving full-service profiles with much information. It can also replace current call centers that provide multiple customer service 24/7 with a tremendous cost on the government entities. Governments may also make greater use of chatbots to support their employees in addition to customers at the same time.

Moreover, for different English domains, many studies have concentrated on Chatbot, whereas Arabic Chatbot studies are minimal and discuss a particular domain. Studies of Arabic e-government services chatbot were not carried out widely due to difficulties in the Arabic language and lack of information sources. Therefore, this dissertation aims to resolve these problems by building an Arabic unified Chatbot for e-government services based on an enhanced ontology and apply QAS.

1.3. Research Objectives

The main objective of this dissertation is to investigate the feasibility of building a unified Arabic chatbot for e-government services based on a previously built ontology by extracting the knowledge base from an ontology and enhance it, and import it to the chatbot system to become an independent knowledge base. Then, depending on multiple approaches, the best possible answer will be selected as a response to any question that could be raised related to government services by deriving the answers from the chatbot knowledge base. Besides, below other sub-objectives for this research:

1. To review the relevant literature of existing researches and get a strong context before the methodological structure is developed.
2. To review and update the e-government services of the ontology according to the latest service publication on (دبي.امارات).
3. To extract all services' details from the ontology.
4. To extract the terms and terminology from the extracted knowledge base and build the retrieval rules accordingly.

5. To analyze the chatbot knowledge base requirements prior to Artificial Intelligence Markup Language (AIML) files construction in order to design an automatic approach to save huge efforts while constructing AIML files.
6. To employ NLP techniques that can help overcome some of the Arabic language challenges.
7. To enhance the answer retrieval using semantic similarity and select the highest matching score.
8. To perform an evaluation of multiple algorithms for QA compared with available chatbots of government services.

1.4. Research Motivation

Areed et al. (2020) and Alqaryouti et al. (2019) pointed out that the volume of information has significantly increased due to dramatic improvement in internet evolution. In addition, the relationships between information sources have become more complex. This caused an increase in the number of users with different needs. Therefore, an analysis will be required to evaluate consumers' needs and highlight their issues to meet their expectations. Dubai has been improving exponentially in the United Arab Emirates over the past 20 years. The town serves its citizens, comprising both expatriates and locals, with over 500 governmental services. The Government of Dubai has decided to implement smart services and increase citizens' satisfaction; therefore, the need to respond more effectively to customer requests becomes very necessary. This dissertation is mainly motivated by multiple factors, which are listed below:

1. Overcoming the delay of response time in government call centers due to the global pandemic COVID-19, which caused all government's employees to work from home.
2. Facing the increase in information resources, which may lead to information conflict.
3. Missing of harmonization among Data Sources.
4. Attending to the government direction of customer happiness and customer experience.
5. Overcoming the limitations of the Arabic Chatbots because of challenges in the Arabic natural language processing.
6. Transforming data from unstructured into a knowledgeable representation.
7. Obtaining immediate assistance or service-related inquiries.
8. Creating direct relationships with clients and getting their feedback.
9. Contributing to smart services adoption by governments' clients.
10. Offering 24/7 service support, no matter what day, time an inquiry may come.

1.5. Research Questions

The dissertation aims to solve and diagnose the problem defined in section (1.2) adequately through the realization of the objectives of this research represented in section (1.3). Therefore, the questions listed below have emerged from these both concepts (problem and objectives):

1) **Is it possible to build an Arabic Chatbot Knowledge Base for Government Services based on the available Ontology resources?**

   The key purpose of this question is to answer the main objective "the feasibility of building an Arabic chatbot for e-government services based on a previously built ontology". This includes necessary enhancements. It is also linked to the sub-objectives 1, 2, and 3.

2) **Is it possible to build an Arabic Chatbot System that can respond to government services related enquirers?**

   The key purpose of this question is to answer the main objective "the feasibility of building an Arabic chatbot for e-government services". It is also linked to the sub-objectives 4 and 5.

3) **Is it possible to retrieve accurate answer using Chatbot better than the formal Chatbot of Smart Dubai government?**

   The key purpose of this question is to answer the main objective "the feasibility of building an Arabic chatbot for e-government services" that its performance is better than "Rashid" chatbot. It is also linked to the sub-objectives 6, 7, and 8.

1.6. Dissertation Structure

The dissertation is composed of the following five chapters:

- **Chapter One:** introduces the research background of information retrieval through different techniques including, Web Search, Question Answering Systems, Semantic Web, and Chatbots. In addition to defining the problem statement, research motivations and objectives, with their relation to the research questions.
- **Chapter Two:** represents a literature review of chatbots, types, architecture including components and approaches, with highlighting the Arabic language

challenges. In addition to semantic similarity measurement, and a survey on Arabic chatbot and applications.

- **Chapter Three:** represents the proposed methodology of building an Arabic chatbot system including knowledge extraction from the ontology, and knowledge construction for the chatbot, then chatbot response algorithm.
- **Chapter Four:** represents the experimental attempts with results discussion and comparisons.
- **Chapter Five:** represents the answers to the research questions and reveals the dissertation's conclusion and the recommendations for future works.

1.7. Summary

In summary, this chapter briefs the importance of the dissertation and highlighted the necessity of information accessibility for users. Furthermore, multiple approaches of information retrieval have been briefly presented e.g. web search, question answering systems, semantic web, and chatbots with a focus on their importance on enhancing the knowledge and information representation. Moreover, Dubai's improvement and its government achievements were taken as an example of delivering services with efficiency and sustainability. Finally, the problem statement has been defined, with research objectives and motivation, with their relation to the research questions, and dissertation structure was presented as well.

2. CHAPTER 2: LITERATURE REVIEW

2.1. Historical Overview

In 1966, Professor Joseph Weizenbaum created the first chatbot ever, a simple computer program that was written at the *Massachusetts Institute of Technology (M.I.T.)* between 1964 and 1966 (Weizenbaum 1966). ELIZA could simulate conversation using substitution methodology and pattern matching, which answered the questions with few tricks on other questions that gave the impression that the program heard these questions before and answered them. ELIZA was a primitive chatbot and unable to develop real-world knowledge or a thoughtful application of self-confidence. However, this was the first step toward more sophisticated chatbots. In 1972 PARRY was founded by an American psychiatrist named Kenneth Colby at *Stanford University* (Colby 1975), the application mimics a schizophrenia patient, and aims to simulate the illness. It was a natural language program, which illustrated an individual's thought. PARRY operates through a complex network of attributions, assumptions and "emotional responses" caused by adjustments in the weights allocated to verbal inputs. He tried to give more personality through beliefs and emotional classification (accept, reject, neutral) instead of matching the trigger words. PARRY suffered from the inconvenience and the exception of a limited number of non-repeating questions, it was unable to generate answers. In 1988, the developer Rollo Carpenter created Jabberwacky, he tried to amusingly mimic the natural conversation of humans. Jabberwacky's contributed to other technology developments. Some people used it for academic research purposes. Jabberwacky is known to be using an AI technique named "Contextual Pattern Matching" Fryer and Carpenter (2006). Afterward in 1992, *The Creative Labs of MS-DOS* created Dr. Sbaitso chatbot, it was one of the first attempts to integrate Artificial Intelligence within a chatbot. The application had a dialogue with the users as if it were a psychologist. But many of the responses fall under the category of *"Why are you feeling like this?"* instead of having a complex conversation (ZEMČÍK 2019). Later, in 1995, Richard Wallace was the pioneer of building A.L.I.C.E. (Artificial Linguistic Internet Computer Entity), a chatbot with universal language processing that employs a heuristic pattern matching to conduct a conversation. The program was inspired by the classical ELIZA of Joseph Weizenbaum. It was formerly known as Alicebot after it started operating on a machine called Alice. For heuristic communication rules, the program uses an XML schema called AIML (Artificial Intelligence Markup Language). It could simulate a chatting conversation with a human over the internet by opening a dialogue about age, hobbies, and some fascinating facts (Shawar & Atwell 2002). ActiveBuddy created Smarterchild in 2001, which was designed to provide easy access to news, stock statistics, weather, film listings, and schedules for transit. It also had the capabilities to help in doubling the productivity of offices by doubling as a calculator, translator and personal secretary (Wei, Yu & Fong 2018). After that, in 2010 Siri has arrived by Apple, a non-human agent providing services and completing tasks through voice command. It has an "intelligent assistant" that allows Apple device users to speak in their natural language voice as commands to run apps or operate the mobile device (Hoy 2018). Google Now was introduced at Google I/O in 2012. It helps to identify a user's question before they need to ask it, by creating information in real-time about what is relevant for each user (Ehrenbrink, Osman & Möller 2017). A few years later, Alex of Amazon came to the surface, an interactive AI platform that can create a personalized living environment when linking other home appliances to its centralized system (Hoy

2018). Finally, in 2015, Microsoft introduced Cortana, a personal and professional assistant in Microsoft 365 that can help users accomplish more tasks without many efforts, so they can concentrate more on what matters (Hoy 2018).

2.2. The Turing Test

The term "Turing Test" is commonly used to point out to the proposal that was made by the British mathematician Alan Turing in 1950 as a process of answering the issue of how computers can think. It is usually used to refer to certain kinds of behavioral experiments for the mental fortitude, or thought, or intelligence in assumed minded entities (Saygin, Cicekli & Akman 2000). In 1991 an American scientist named Hugh Loebner decided to organize a Turing Test experiment with the *Cambridge Center for Behavioral Studies*. Dr. Loebner guaranteed a $100,000 as a grand prize for the first computer, the computer should give answers that are indistinguishable from an answer given by human, which makes a computer be considered as “thinking” (Mauldin 1994). Each year the most humane computer chatbot receives a bronze medal and cash prize. This inspired the researchers and encouraged experts to start producing more chatbots in order to win the prize. Many chatbots were designed for Loebner prize e.g. CONVERSE (Batacharia et al. 1999), ALICE (Wallace 2009), Mitsuku (Inaba et al. 2014) and many others.

2.3. Chatbot Definition

The learning process is the main focus of AI. The principle of human-language contact with computers refers to AI. A chatbot or conversational agent is an intelligent conversational program that communicates in a natural language with human users and imitates human conversations. In recent years this field has gained greater interest in research and industry (Liu et al. 2018). Computer programs which can communicate with users in human languages are identified as chatbots. Generally, they aim to preserve aimless conversation with users on a range of topics. According to Crockett, O'Shea and Bandar (2011), Conversational Agents is defined as "a computer program that communicates with a user via a natural language dialogue and delivers a type of service by processing user's feedback and providing an appropriate answer". According to Abu Shawar and Atwell (2007), chatbots aim to find out whether they could trick users as being real people. Human-machine interaction is a system that uses the natural language to promote contact between users and computers. Human-machine conversation is a technology, which supports the communication using natural language between users and computers. Natural language technologies are used by chatbots to engage users in a wide range of applications through textual search for information and task-based dialogues (Lester, Branting & Mott 2004). This includes help desk, customer service, guided selling, technical support, and website navigation. The continuous growth of Internet technology, computational linguistics, online apps, and the increase in commercial needs for customer support have led to the creation of commercial conversation agents.

2.4. Types of Chatbot

Based on their interfaces, there are two main types of chatbots. They are Embodied Chatbot, and Linguistic Chatbot.

2.4.1. Embodied Chatbots

The Embodied Chatbot (EC) can be defined as "Computer-generated cartoon-like characters that display in face-to-face dialogue with many abilities as humans, including the capability to create and respond to verbal and nonverbal communication" (Cassell

2000). ECs enhance human characteristic to communicate with people through natural language dialogs to answer questions and perform tasks for the user. Valle (2010) pointed that the ECs structure consists of the following key elements:

- An interface that captures gestures or language inputs into the EC, such as gesture analysis and audio.
- A dialogue manager or engine that determines the behavior of the EC.
- A visual component that manages movement and gestures, such as gesture synthesis and audio.

2.4.2. Linguistic Chatbots

There are two categories for Linguistic Chatbots (LC): Textual Dialog Systems (TDS) and Spoken Dialog Systems (SDS):

- Textual Dialog System (TDS): Is a textual chatbot that provides communication via "User Interface" that has text boxes as an input and output to receive/send text messages and reply accordingly (Hijjawi 2011).
- Spoken Dialogue System (SDS) deal with the converting from voice to text. Typical user can interact with it by speaking directly then having SDS algorithms interpret the voice (O'Shea, Bandar & Crockett 2011). SDS aims to provide efficacious access to services and applications. Several SDS applications have been developed over the past years such as e-mail writing services, schedule calendars, booking of entertainment activities, home appliances control, and product reviews (Demberg & Moore 2006).

TDS entails several difficulties in the structuring of sentences, language grammar and word sense disambiguation, in addition to morphological analysis. In section (2.6), these challenges will be highlighted and explained. Meanwhile, SDS faces some other difficulties as well, such as capturing the voice of the user, separating it from other background noises, transforming the voice into text. Taking into consideration the difference in pronunciation from one user to another and the ambiguity of similarly pronounced words. Also, all difficulties associated with TDS will be faced again in SDS after converting the voice to text. All difficulties faced in both types will be magnified while dealing with Arabic chatbot due to the variety of Arabic dialects and the absence of effective Arabic language analysis systems.

2.5. Chatbots Architecture

The development of conversation agents takes long time with an excessive cost (Razmerita et al. 2004). This needs experience in writing conversations and a clear understanding of the language written form (e.g. English or Arabic). Chatbots main purpose is to mimic human conversations. Therefore, researches have to develop mechanisms based on their designed system architecture to represent and argue the knowledge, gather the required domain knowledge and finally implement the system modules accordingly. This involves many challenges, beginning from capturing and interpreting the statement, clear statement ambiguity by focusing on a particular context, knowledge representing and reasoning in a specific domain, as well as other challenges related to adaptability, responsiveness, and usability of chatbot.

2.5.1. Components

Chatbots consists of three main components (AlHumoud, Al Wazrah & Aldamegh 2018):

- An input and output interface that enables users to interact through.
- A brain or knowledge base that includes the conversation content to keep it on the domain track.
- An engine for the conversation to manage the semantic context.

2.5.2. Dataset Models

There are two dataset models to represent the knowledge source in chatbots, the generative-based model, and the retrieval-based model (AlHumoud, Al Wazrah & Aldamegh 2018). Both models are explained below:

- Generative-Based Model: in this model a set of techniques are used by the chatbot to generate completely new response by using Neural Network (NN) and deep learning, and may utilize predefined responses.
- Retrieval-Based Model: In this model, a group of predefined responses is used by the chatbot to select the right response based on a specified methodology, but if there is no response preset, it may not be applicable.

2.5.3. Approaches

Many approaches have been introduced to build a chatbot (Bradeško & Mladenić 2012), (O'Shea, Bandar & Crockett 2009). The main approaches are described below:

2.5.3.1. Pattern Matching

This is the most commonly used approach in building chatbot. For any chatbot program, there are variants in certain pattern matching algorithms. The approaches to matching patterns can differ in complexity but the basic concept remains the same. The simplest patterns was used by old chatbots such as ELIZA and the PC Therapist. The Alicebot Free

Software Group created AIML between 1995 and 2000 to allow people to incorporate knowledge pattern into chatbots based on A.L.I.C.E free software technology. The syntax of AIML is based on XML and mainly consists of input rules named (categories) with suitable output (Bradeško & Mladenić 2012). The pattern has to fill the entire input and considered as case insensitive. A wildcard (*) binds to one or multiple words can also be used to cover more patterns using a single pattern. In addition, AIML's main strength comes from its ability to call patterns recursively.

2.5.3.2. Natural Language Processing

The textual parsing is the process of taking the original text and transform it into a group of words (lexical parsing) with features, often to decide its grammatical structure. Then, the lexical structure can be verified whether it represents appropriate expression (syntactic parsing). The previous parsers were very basic searching for identifiable keywords in an appropriate order. The sentences "please take the money" and "can you get the money" are examples of parsing, where both sentences are parsed into "take money". With that approach, the chatbot can cover multiple input sentences with a limited set of patterns. The more complex parsers that are used in later chatbots apply the full grammatical parsing of the natural language sentences.

2.5.3.3. Markov Chain Models

The concept of Markov Chain Models is that in any textual dataset, there is a fixed probability for each occurrence of a letter or word. The order of a variable implies the number of consecutive events that the variable takes into account. For instance, if an input text is "excellent", then Markov model of the order 0 predicts that letter "e" will occur with a probability of 3/9. The Markov Chain Models were used in chatbots to build responses that are more likely to be successful and therefore more accurate. In some situations (HeX) these structures were also used as a failback method to produce a nonsense sentence that sounds right.

2.5.3.4. Ontologies

Ontology can be defined as a set of classes that are interconnected in hierarchy. The knowledge-baes is described as a graph containing classes; each class represents the properties and the concepts. Classes with logical relation are often connected, and use those relationships to imply new statements (reasoning) (AlHumoud, Al Wazrah & Aldamegh 2018). OpenCyc[3] is an example of such ontology which was used or at least attempted to be used in chatbots (Douglas 1995).

2.5.3.5. ChatScript

The successor to AIML language is ChatScript. It targets better syntax, making it easier to maintain. It solves the problem of zero word matching and adds a bundle of additional functionalities including, continuations, concepts, variables, fact triples, logical (and/or), and functions. Using these functionalities within the script itself, it is attempting to cover the need for ontologies.

2.5.3.6. Semantic Similarity

Generally, the term similarity is used to represent the level of similarity between two objects. In Al, the similarity-based research has been applied in multiple applications (Feng, Zhou & Martin 2008), such as information extraction, information retrieval, question answering, machine translation, and Conversational Agents (O'Shea, Bandar &

Crockett 2009). Section (2.7) will focus more on semantic similarity and their measurements techniques as well as the related challenges.

2.6. Challenges in Arabic Language

There are many linguistic challenges related to Arabic language. The main challenges are described below:

2.6.1. Many Dialects in Arabic Language

Three primary categories are available in the Arabic language (Ryding 2005): The first category is Classical, which is the original Arabic language that has been used in Quran. It is very rich in grammar and equipped with huge amount of vocabulary, it also contains many diacritical characters to distinguish Arabic words and determine their pronunciation and grammatical meaning that help in recognizing their grammatical cases (e.g. verb /noun). The second category is Modern, which is the official language used by governments and formal correspondences, it rarely uses the diacritical marks, however, the grammatical meaning can be understood from the context of the sentence or paragraph. The third category is colloquial Arabic which is slang language used in many different Arab countries. Arabic dialects may have different phrases and might include other language terms. Therefore, people who live in one Arab country may not understand a dialect spoken in another Arab country. Several Arabic dialects would be without standard grammar. This raises the challenge of creating an Arabic chatbot that can recognize what users from different Arab countries mean to say.

2.6.2. Morphology of Arabic Language

Morphology in linguistics can be defined as the study of the internal word's structure (El Kholy & Habash 2010). Its aims to identify, analyze and describe the morphemes' structure and other units of meaning in one language (Altabba, Al-Zaraee & Shukairy 2010). The root-and-pattern morphology is a major distinguishing feature of Arabic language. The semantic abstraction of a word is the root that may consist of two, three, four or more (rarely) constants from which words are formed by combining patterns of templates. For instance, the word "قرأ" that is pronounced as "QARA" has the broad lexical sense of "reading". Many words can be derived from this word, such as the word "قارئ" means "reader" pronounced as "QARI'". Also "مقروء" means "readable" pronounced as "MAQRU". Another example is "قراءة" means "reading" pronounced as "QERA'AH", while the word "مِقْرَأ" means "desk" pronounced as "MIQRA". Furthermore, Arabic language can have feminine and masculine forms for nouns, as well as singular and plural in addition to dual forms. Adjectives are also morphologically similar to nouns. Altantawy, Habash and Rambow (2011) pointed that Arabic language is morphologically rich and complex. It is characterized by a combination of affixation morphemes (prefixes and suffixes), complex morphological, phonological and orthographic rules and a rich system of features. The author stated that Morphology generally refers to two important aspects: the first aspect is the derivative morphology that influences to word formation. The second aspect is inflection morphology that influences the interaction of words with syntax, but the derivative morphology specifies pre-incipient of transforming a word. For many NLP applications the morphological analysis and generation are essential, such as information retrieval (Aljlayl & Frieder 2002), machine translation (El Kholy & Habash 2010), and conversation agents (O'Shea, Bandar & Crockett 2010). Many works have been conducted on analyzing and generating Arabic morphology in various approaches and depths of linguistic. For example: ISRI (Taghva, Elkhoury & Coombs 2005), Light Stemming (Larkey, Ballesteros & Connell 2007), Qutuf (Altabba, Al-Zaraee & Shukairy 2010), Morphology Analyzer (Altantawy, Habash & Rambow 2011) and Alkhalil Morpho (Boudlal et al. 2010).

2.6.3. Ambiguity in Arabic Language

The same word may have different meanings in the Arabic language. The Arabic language has two types of ambiguities: the first type is called "morphological ambiguity", which is mainly caused by the absence of Arab diacritics. For example the word "عـقـد" means "necklace" while the same word with a slight change in the diacritics can mean "decade" and with additional change in the diacritics, it can become "contract". Morphological ambiguity increases the challenge of developing Arabic conversation systems, because diacritics are typically absent in modern Arabic. Chatbot users now do not usually involve Arabic diacritics in their statements, thus making it more difficult to know the aimed word. The second type of ambiguity is called "word sense ambiguity", which occurs when two words have different meanings with an exact syntactic form (including diacritics). For example, the word "عين" means "eye" and it can mean "water spring". Another example is word "يُسَـلم" means "deliver"; it can also mean "salute". Furthermore, the ambiguity of word sense is another obstacle in developing a semantic chatbot. To distinguish the intended meaning from the word a method to follow will be required. Several disambiguation techniques have been designed to overcome the word sense ambiguity from the context of the sentence, such as: (Li, Szpakowicz & Matwin 1995), (Ide, Erjavec and

Tufis 2002), (Liu, Zhou & Zheng 2007), (Agirre, Lopez De Lacalle & Soroa 2009), (Zouaghi, Merhbene & Zrigui 2011) but these techniques may not be ideal in chatbot. For years, Word Sense Disambiguation (WSD) has been the subject of research in NLP, and recently WSD proved to be significant in many NLP tasks such as parsing and translation. The WSD is considered the main step in language comprehension, including automatic retrieval of information, question answering, text mining and interlocutors (Agirre, Lopez De Lacalle & Soroa 2009).

2.6.4. Grammars of Arabic Language

In terms of word order, Arabic language has a flexible sentence structure. It's possible to write the same sentence in Arabic in three different forms (El Kholy & Habash 2010), below is an example of these forms for the same sentence "Sami read a book" (form is given from right to left):

- "قرأ سامي الكتاب" = (object) + (subject) + (verb)
- "سامي قرأ الكتاب" = (object) + (verb) + (subject)
- "الكتاب قرأه سامي" = (subject) + (verb) + (object)

In contrast, in English language the structure of sentence might be (subject) + (verb) + (object). Hence, this ability of flexibility that is available in the Arabic language increases the level of complexity in developing an Arabic Chatbot in terms of its ability to understand the actual sentences. The analysis of computer semantics in Arabic is much narrower than in other areas of the NLP because of its high level of difficulty and complexity (El Kholy & Habash 2010).

2.6.5. Use of Non-Arabic Words in Arabic Language Dialects

There is a countless number of non-Arabic words that are used while speaking or writing without actual translation to Arabic. For instance, the word "انترنت" means "internet" and it is widely used to express the web network. In addition, the word "باص" which means "bus" that is used to express the transportation vehicle. Another example is the word "دايت" which means "diet" often refers to control the kind of food to be eaten. These words and many other dialect words do not comply with the same morphological analytical and grammatical rules that is certainly adding another challenge in developing an Arabic interlocutor.

2.7. Semantic Similarity Measurement

This section demonstrates the semantic similarity measurement of words or sentences and the different techniques used to calculate them, along with the pros and cons of each technique. In addition, it concentrates on the similarity in Arabic words and sentences with the challenges related to each technique. There are many tools used in measuring the similarity, such as, Suggested Upper Merged Ontology (SUMO), AraMorph, WordNet and many more. Words and sentences similarity evaluation methods are covered in this section as well.

Semantic similarity can be described as measuring the degree to which two words or sentences are similar to each other from a logical point of view. Based on word similarity and corpus statistics, there are two stages to be performed while measuring semantic sentence similarity when text dialogue is exchanged between human and chatbot: The first stage is "Word Similarity" that measures the similarity of all words in the text. The second stage is "Sentence Similarity" that measures the overall similarity of the sentence by calculating the score between each pair of words. However, word similarity is actually a part of sentence similarity; therefore, both are called "Semantic Similarity". These two stages are closely related, and it is difficult to separate them. In various AI applications such as NLP, semantic similarity is widely used. For example, Information Extraction (Hliaoutakis et al. 2006), Automatic Question Answering Systems (Narayanan & Harabagiu 2004), Text Analysis (Malandrakis et al. 2013), Conversational Agents (O'Shea et al. 2012), Automatic Text Summarization (Aliguliyev 2009), and Machine Translation (JEONG 2005).

In English language, the similarity of sentences has been thoroughly studied by several researchers. Mainly, there are two techniques to measure the similarity of the sentences. The first one is Sentence Similarity based on Semantic Networks and corpus statistics (SSSN) which relays on semantic networks such as WordNet, by calculating each word similarity between two sentences, then calculating the semantic similarity of the whole sentence; which can be performed through calculating the similarity between each pair of words. STASIS method is an example for this technique which was developed by (Li et al. 2006). The second technique is Latent Semantic Analysis (LSA) which is a theory and technique for extracting and representing the meaning of contextual-usage of words through statistical computations applied to a wide corpus of text (Landauer, Foltz & Laham 1998). LSA lacks of utilizing semantic networks, knowledge bases, morphologies or dictionaries constructed by human, but it uses only sentences or paragraphs entered as a raw text. The basic concept for the LSA algorithm is that it takes text input and parses it into group of words, which results a matrix of the term-document. The rows of the matrix represent the unique words available in each paragraph, and the columns represent each paragraph. Briefly, rows are words, and columns are paragraphs. Each cell contains the recurrence at which the word in its line occurs in the paragraph denoted by its column. Then each cell recurrence is measured by a function, which expresses both the significance of the word in the particular paragraph, and the extent to which the word type generally carries information in the discourse domain. A sentence in LSA can be expressed in an extremely high-dimensional space that may reach thousands of dimensions (Landauer, Foltz & Laham 1998). This leads to a very sparse sentence vector, which is therefore inefficient in computational terms. Due to high dimensionality and high sparsity the similarity computation could be unacceptable (Li et al. 2006).

O'Shea et al. (2008) conducted a comparison between LSA and STASIS using a dataset that consists of 65 pairs of sentences. A questionnaire was designed and disseminated among a number of participants who were asked to give their rating based on a question of "how close the sentences are in meaning" The rating scale ranged from 0 (minimum) to 4.0 (maximum). The same dataset was then computed via LSA and STASIS. Using the same dataset, the experiment results showed that both techniques have performed well, and the similarity judgements indicated that both algorithms in consistent with human rating. Although STASIS scored (0.816) correlation, LSA could score a higher rank of (0.838) in accordance with human. Even though LSA has the capability of capturing and representing significant components of the meaning that a lexical and passage may have, it lacks some important cognitive abilities of humans such as building and applying knowledge from experience (Landauer, Foltz & Laham 1998). In contrast with LSA, the STASIS technique is entirely designed based on semantic networks (WordNet) to measure the similarity of sentences, where the associations between words and synsets are specified from a human point of view. Researchers found that the STASIS technique would be more appropriate for the development of semantic conversational agents because it measures the similarity of sentences based on a knowledge base that has been built by human experience rather than statistical approach for calculating semantic similarities.

The Arabic language received little attention concerning word and sentence similarity. Almarsoomi et al. (2013) developed an Arabic algorithm that measures word semantic similarity using Arabic WordNet. When dealing with Arabic language chatbots, more challenges can be identified. These challenges can be classified into three main categories: Linguistic Challenges, which are related to the use of Arabic language as discussed in the previous section (2.6). Technical Challenges, which are related to the performance and speed of response. Conceptual Challenges, which are related to the philosophy of using semantic similarity while dealing with Arabic chatbots.

2.7.1. Linguistic Challenges

These challenges have been discussed in section (2.6), and are summarized below:

- Many Dialects in Arabic Language
- Morphology of Arabic Language
- Ambiguity in Arabic Language
- Grammars of Arabic Language
- Use of Non-Arabic Words in Arabic Language Dialects

2.7.2. Technical Challenges

This involves the difficulties of integrating the Arabic chatbot with the available systems, e.g. Arabic WordNet (AWN).

- Tools Limitation: The purpose of AWN is to browsing only; the user cannot apply any changes on the database of the lexical. This limitation prevents the users from adding any new words using the browser of the AWN. On the other hand, the Arabic words database is accessible through XML format. However, the researches must decide the preferable data format based on their needs. Furthermore, another missing functionality in AWN, which results in additional limitation, is that the ontology need to be updated with new entities or relations for some domains. SIGMA has the same limitations as well. Whereas Protégé interface is quite simple and enables the users to create and update the ontology with a graphical presentation but it also does not support "KIF" format that is used by AWN.
- Incompleteness: According to Fellbaum (2020), the total number of Arabic words in AWN database is (23,481), which represents less than 10% of the Arabic stemmed words. Therefore, building an Arabic semantic chatbot, requires AWN to be expanded in order to cover all Arabic stemmed words.

Similarity Measurement: Since AWN browser was designed for browsing purposes only, the similarity measurement functionally is missing. Even though, the source code of AWN browser is available for public use, but there is insufficient documentation for the software that enables researchers to change it.

2.7.3. Conceptual Challenges

The conceptual challenge relates to the principle of similarity itself, which includes:

- The different meaning of similarity: words or sentences are not quite the same always. They may be very similar in certain domains or contexts, but in others, it would be wiser to be more specific. Such information cannot be important in some cases, as some other cases do. For example, the sentence "I lost my mobile" is similar to the sentence "I don't have a mobile". The two sentences have the same meaning that the speaker does not have a mobile. However, this meaning will be different if speaker is explaining this to a security officer.
- Negation: the process of opposing the word by placing a negated word before the original word is called "Negation". The sentence similarity measurement does not consider these negated words, which can completely change the meaning of the sentence. For instance:
    - "اريد الذهاب الى المدرسة" means "I want to go to school".
    - "لا اريد الذهاب الى المدرسة" means "I don't want to go to school".

  These two sentences contain quite similar words but in fact, one of them is actually negating the other.
- Function words: words that have little lexical meaning or have unclear meaning and convey grammatical relationships, or indicate a speaker's attitude or mood, among other words within a sentence. Arabic language like other languages has function words; such as the questions words:
    - "من" means "who"
    - "ما" means "what"

- o "متى" means "when"

These words contains semantic information related to the sentence. However, in the knowledge base or ontology, they cannot be classified as something available in the real world. For instance, the word ("متى" means "when") is used to link between an event with the time when it occurred, but the term itself is not classifiable as an object that always exists. Therefore, a good similarity measurements of sentences need to include function words.

- Type of Sentences: generally, sentences can be categorized into different types; affirmative, informative, negative, and questionable sentences. Each one of them must be identified before similarity measurement. Such as:
    - o "احمد كسر النافذة" means "Ahmed broke a window".

  This sentence is not similar to this sentence:

    - o "من كسر النافذة؟" means "Who did break the window?"

  In addition, sentence similarity cannot extract facts from sentences. Therefore, the following example is showing high similarity:

    - o "أريد اصلاح النافذة المكسورة" means "I want to fix the broken window"
    - o "اين يمكن اصلاح النافذة المكسورة" means "Where to fix the broken window"

  Although all these sentences are not similar but they share the same fact that the window was broken.

- The compound nature of Arabic words: because of the affixes added to the Arabic words, this increases the richness in semantic details. Such affixes contain rich information about singular, dual, plural forms, as well as other sentence information:

- "يدرسون" means "they are studying" in plural masculine
- "يدرسن" means also "they are studying" but in plural feminine.

The word implies a masculine and feminine plural, and in this case, there is a present tense in which the act of studying occurs.

2.8. Evaluation of Chatbots

Like any other program, chatbots have to pass intensive testing and evaluation before launching and deploying them for public use. Chatbots evaluation is the process of conducting multiple test scenarios on the chatbot from various aspects which performed by a group of qualified people from different cultural backgrounds to determine whether the chatbot is suitable for real-world interaction with users, and to identify any weaknesses associated with the chatbot based on the feedback provided by the evaluators.

Evaluations of chatbot were carried out using a number of criteria such as: usability, credibility, efficiency, productivity, effectiveness, performance, speed, error rates, and users' satisfaction. According to Abu Shawar and Atwell (2007), it's not preferable to asses all these criteria together while testing. However, the authors suggested customizing the test by adapting the criteria according to the application domain and users' needs as there are no standard metrics and consistency in the methods of evaluation. Preece, Rogers and Sharp (2006) mentioned that traditional evaluation mainly concentrates on usability criteria in which the chatbot has been designed for. Other approaches focuses more on the subjectivity and objectivity reactions while evaluating, it can also cover the emotional aspects e.g. users' satisfaction which referred to as a user experience.

While researchers may not follow a standard methodology for evaluating chatbots, but the evaluation can be divided into two main categories: subjective and objective:

The subjective evaluation focuses mainly on the requirements of user's satisfaction, such as:

- Ease of Use: to measure the easiness at which the user can get the desired information.
- User Expertise: to determine the extent at which the evaluator understood what he/she was able to say or do at each stage in the conversation.
- Behavioral Anticipation: to determine the degree of which the chatbot is able to meet the user's expectations.
- User Retention: to measure the degree at which the user would use the application rather than speaking to human experts.

The objective evaluation concentrates more on the practical value from using the chatbot. O'Shea, Bandar and Crockett (2010) described a metrics for objective evaluations that includes:

- Length of conversation or dialogue.
- The count of shifting the conversation between user and the chatbot.
- Task completion counts.
- Error counts.
- Correct counts.
- Precision in speech recognition.

Silvervarg and Jönsson (2011) claimed that evaluation of a chatbot is primarily achieved either by sending a questionnaire to users who can report their evaluation of using the chatbot or by observing the resulting dialogue. In general, questionnaires are a notably efficient approach to be applied in order to analyze, since they allow many users with various backgrounds to assess many objectives from various aspects, including functionality, usability and responsiveness. Additionally, they can assess many other criteria that vary from one application to another. They also provide an efficient quantitative measurement of application characteristics. Some questionnaires can be used as a stand-alone assessment method in some certain circumstances

According to Walker et al. (1998), three main limitations are identified in the subjective and objective evaluation:

- Using the reference answers will make it extremely difficult to compare systems that are using various dialog techniques to accomplish the same task; this comparison includes the interpretation of a standard answer for each user utterance.
- Various evaluation metrics can be highly correlated and thus provide contradictory feedback on performance.
- The failure to swap out or merge various metrics to become generalized.

A general framework was introduced by Walker et al. (1998) called PARADISE for the evaluation in order to overcome these limitation and compare the performance of the spoken chatbots. PARADISE was used to assess DARPA communicator (Walker et al. 2001). It uses a range of decision-making approaches to incorporate a number of performance measures such as the completion of user satisfaction task and dialogue cost into a unified evaluation function for performance.

To evaluate rule-based systems independently from other chatbot components, O'Shea, Bandar and Crockett (2011) introduced "Wizard of Oz". The purpose of this wizard is to simulate the interface of the chatbot and runs the rule-based application, so it can allow the user to test and assess the application rules separately. However, this approach was time consuming and was not adopted commercially for application development. An exploratory experiment was conducted by Walker et al. (2001) for a nine different DARPA communicator systems. All of these systems supported the travel planning and used some form of mixed-initiative interaction. The evaluation included subjective and objective assessment based on extracting the objective metrics from the logs and collecting subjective metrics through a survey. Another evaluation methodology was introduced by O'Shea, Bandar and Crockett (2010) for semantic chatbot. The process of the evaluation is split into two phases:

- Phase One: evaluating the capabilities of semantic chatbot for interacting from the user's perspective, this phase is split into two sections:
    - Section A: covers an experiment that uses a number of participants to test the semantic chatbot interaction, which includes the following metrics:
        - Satisfaction: is it trouble-free and pleasing?
        - Usability: is it easy to use?
        - Naturalness: is it human-like?

        - Accuracy: is it able to interact correctly and without misunderstanding?
        - Task Completion: is it completing the task and achieve the goal?
        - Retain User: is it considered for future reuse?
    - Section B: compares the two chatbots, the first one is semantic chatbot developed using the Semantic Conversational Agent Framework (SCAF) and the second one is a text-based chatbot (InfoBot). The goal of the comparison is to address the interactions differences between both chatbots by measuring the user's satisfaction. This was evaluated by analyzing the different aspects of interaction, such as usability and the inherent essence of the dialog.
- Phase Two: evaluating the scripting of the natural language that is used to script the semantic chatbot. The goal of this evaluation is to identify whether or not the natural language scripting is able to easily construct a script in an efficient manner without any script's flaw from the writer's perspective, this phase includes the following metrics:
    - Intuitiveness: is it easy to use?
    - Usefulness: is it beneficial and can contribute in the functionality?
    - Flawlessness: is it errors free?

2.9. Arabic Chatbots Survey

Little work has been conducted in developing Arabic chatbots. This is due to the complexity of the Arabic language and the limited linguistics researches, another factor can be considered is the lack of social acceptance for this kind of applications.

Quran chatbot by Shawar and Atwell (2004b) was one of the earliest researches on the use of Arabic chatbot in the collected survey. It was designed based on the Islamic holy book (Quran). There are 6,236 (Ayah) or verses in Quran, and they are grouped into 114 (Surahs). The user inputs are used as a guide for the chatbot to respond back by presenting the relevant verses. The same authors applied their Java program that was developed by themselves Shawar and Atwell (2003) to transform Quran corpus into Arabic AIML files based on machine learning so they can retrain ALICE chatbot (Shawar & Atwell 2004a).

Shawar and Atwell (2009) also developed a web based chatbot for question and answering (QA). The chatbot corpus was built based on 412 Arabic frequently asked question (FAQs) gathered from five different web pages with their answer. Different health topics were covered in these questions such as dental care, pregnancy, motherhood and fasting in addition to some blood disease such as diabetes and cholesterol. This chatbot was designed without the use of NLP, but based on two AIML files (atomic and default) that was created by a Java program. The atomic file holds the questions in its initial form, with the answers that presented in the corpus. The use of the default file is to assure that any form of question might be raised by the user, a mapping reference will be found that is pointing to the initial question form which is stored in the knowledge base.

A female chatbot named BOTTA was developed by Ali and Habash (2016), the purpose of this chatbot was to entertain users who like to chat in dialect. It supports dialect of Egyptian Arabic as an input and output. As it connects with the maximum possible number of users to maximize the number of Arabic conversation. BOTTA was launched using Pandorabots platform and its knowledge base was AIML files which hold the categories of responses, phrases, topics and a file that was used to map the related words and phrases.

Despite the non-use of text normalization in BOTTA, it was able to resolve 85% of spelling mistakes while typing because of applying orthographic transformation. The conversation length in BOTTA is considered to be long compared to other chatbots because it stores information about the user and respond back accordingly. However, the knowledge base is not being updated with these inputs and responses to be reused in other conversation.

Hijjawi et al. (2014) developed ArabChat, a closed domain chatbot with a web interface that was used by the students of Applied Science University (ASU) in Jordan. The users' interaction with the chatbot was via textual Modern Standard Arabic (MSA). The conversation remains active until one of the parties closes the session. Like BOTTA, ArabChat stores information about the users and respond based on this information, which can create a long conversation. The scripting language and scripting engine are considered the core components of the ArabChat. The engine is divided into multiple subcomponents, which enable handling conversations topics. The knowledge base of ArabChat consists of 1,218 utterances, which are classified into multiple contexts, and each context consists of rules. These rules consists of patterns and correlated textual responses. 174 users tested ArabChat, with an average of seven inputs by each user. The output that could match the expectation was 73%. A mobile based version of ArabChat was developed by Hijjawi, Qattous and Alsheiksalem (2015) which was also used as an advisory for the students of ASU, the app was deployed on Android platform. Despite of many challenges that might be faced by user in some Arab countries such as bandwidth limitation and unstable internet, the application could perform well even with these challenges, as it has shown better users' satisfaction where 96% of the users preferred the mobile version on the web interface. An enhanced version of ArabChat was presented by Hijjawi, Bandar and Crockett (2016), this version came with additional features like Hybrid Rules and Utterance Classification. The Hybrid Rules enabled the chatbot to deal with multi topics utterance. While the purpose of Utterance Classification was to differentiate between a question and a non-question utterance by adding more keywords to the pattern so it can better match keyword. To meet the new features requirement, an improvement had to be implemented at the engine level and the scripting language in addition to the knowledge base. Due to unserious users, the ArabChat showed better results than the enhanced version. However, the results of the manual analysis on the logs of the enhanced version proved to better as it could successfully deal with 82% of utterances.

An intelligent Arabic conversational tutoring system was developed by Alobaidi et al. (2013) and given the name (Abdullah). The aim of this chatbot is to teach children aged between 10 and 12 years old, the Islamic topics through a modern education. It depends on MSA to engage with students by asking a set of questions, then discussing their responses. It also uses Classical Arabic to provide evidences from both Quran and Hadith (Prophet Mohammed sayings and deeds). Abdullah uses visual and auditory effects to interact with students so it can determine the knowledge level of the student and hence drive the conversation accordingly. It is capable to distinguish from the users inputs whether it can be question or an answer. The approach used by the authors to build Abdullah is Pattern Matching and it consists of a knowledge base that contains the topics and scripting language to transfer the tutorial to the students, in addition to the tutorial knowledge base, which is used to determine the student’s level.

Another Arabic intelligent conversational tutoring system was developed by Aljameel et al. (2017) which have been given the name (LANA). The approach used in implementing

LANA was Pattern Matching and short text similarity algorithm. It targets children with Autism Spectrum Disorder (ASD) of an age between 10 to 16 years old. The children should achieve the basic competency of Arabic writing, so they can learn new topics such as science through MSA. Due to some difficulties in the traditional teaching model for ASD children because teachers are unable to meet the needs of every individual, this model of teaching enables children to perform practices independently and improve their skills through kinesthetic in addition to the auditory and visual effects.

2.10. Applications of Chatbots

2.10.1.Education

Artificial intelligence is rapidly transforming education. E-Learning and chatbots are increasingly becoming useful tools for education. A well-developed chatbot will be able to recognize, assess and interpret the educational needs of students in different programs and provide academic advice round-the-clock. There are many universities have already started using chatbots to give information for their students. For instance, Georgia State University, West Texas A&M University, the University of Memphis, and Arizona State University (Cortez 2018).

2.10.2.Healthcare

Smart algorithm powered by text-based or voice-based searches interfaces have multiplied in recent years, they are also taking their place in healthcare. Google search engine is one of the leading platform of information retrieval about healthcare enquiries, such as symptoms, treatments or even procedures. The futurist medical providers hopes to reduce the pressure on the primary care doctors and enable patients to learn and take responsibility toward their health. However, users are get confused and overwhelmed with a diversity of opinions that can come from unreliable sources after each search they make (Mohasses 2019). Healthcare is one of regulated sectors in any government, therefore any information is provided by a registered medical provider will be under the responsibility of that medical provider. But there are some governments such as the government of Rwanda who have taken this responsibility on their shoulders by offering chatbots that can release some pressure on the doctors by handling minor medical cases and forward only major and serious cases to doctors (Bizimungu 2018). Moreover, Depression is considered as a common mental disorder, the World Health Organization (WHO 2020) states that more than 264 million people globally are suffering from depression. A clinical research psychologist Dr. Alison Darcy have founded the first mental health chatbot in 2017 named (Woebot). Based on NLP Woebot delivered a Cognitive Behavioral Therapy (CBT) for adults that suffers from depression symptoms. It was one of the most successful conversational and machine learning interventions for recognizing and assessing the patient's emotional state, tracking patient progress and making appropriate suggestions (Fitzpatrick, Darcy & Vierhile 2017).

2.10.3.Customer Service

One of the main cost drivers in customer-centric organization is customer service. Companies around the world are estimated to spend $1.3 trillion a year to provide support to their customers (Reddy 2017). However, IBM claims that its chatbot, can reduce response time up to 99%, as it can handle more than 40 different cases simultaneously, which will reflect in cost reduction by an average of $15-$200 (human support) to $1 (chatbot support). IKEA was one first organization to provide customer service through chatbots (Brandtzaeg & Følstad 2019). Their chatbot “Anna” was a cartoon character that can show emotional reaction based on a text conversation with customers. The main purpose of Anna is it to help customers discover products and buy them through answering questions related product cost, size, spare parts, schedules of working hours, and place an order. It may also reply to other statements that are not related to fluent responses and try to drive the conversation back to the product's subject. The most interesting thing about Anna is that she presents a summary of what customer needs, and guide the user to the desired page of the product where all related information can be found, so it focuses on both chat and internet browsing. Then provides a menu that a user can choose form to place the order.

2.10.4.Governments / Public Sectors

The future benefits of using Information and Communication Technology (ICT) to provide public services to citizens have been of great interest. There was already a strong belief during the 1990s that IT services would build a modern, better functioning government of the future where governments may become more efficient, provide higher quality and more publicly available (Torres, Pina & Royo 2005). For example, Smart London (2018) introduced a chatbot called (TravelBot), which offers details about the bus arrival and departure times through Facebook Messenger. One of the common features of such a chatbot is collecting information about the passengers and their trip preferences. In addition, the Homeland Security Department in the United States have developed (Emma). A virtual officer powered by artificial intelligence that is able to assist applicant on services about citizenship and immigration then take them to the right page. Emma types her responses in both languages (English/Spanish) depending on the question language

(USCIS 2018). On the other hand, San Diego County Sheriff launched (Coptivity), an AI-based chatbot that can be accessed through a smartphone app and offers immediate response assistance to patrol deputies. Officers may ask Coptivity to immediately determine the registration status of a car and the history of the owner's (criminal and mental health) without the need to wait for a long time (Egland 2019). Another example is (Chip), a chatbot that proved to the City of Los Angeles that it could provide answers to potential recruits from police departments. During its first 24 hours, it could assist more than 180 people. Chip saves the town approximately 70 to 80 minutes of call time every day (Douglas 2018). In the Republic of Latvia, the Register of Enterprises in the Ministry of Justice introduced a chatbot named (UNA) which is available 24/7 on their website and Facebook messenger. It is able to answer frequently asked question related to business registration, merchants, liquidation, and application tracking (Noordt & Misuraca 2019). Moreover, the municipality of Vienna city in Austria have launched (WienBot) based on a research that was made about the most frequent visited pages on their website, where they realized that thousands of hit search occurs monthly to find more information related to the services they provide. WienBot enabled citizens to easily find information instead of searching many pages on the website. It has the ability to answer different questions related to about 350 various topics and expanded the German language by covering local dialect as well (Noordt & Misuraca 2019). The (GovBot) is another government example that was implemented in the City of Bonn in Germany, the purpose of GovBot is to help citizens with their administrative services. Citizens can request for application forms, working hours or they can even book an appointment via the Chatbot. GovBot is currently being used in the administration of City of Bonn, North-Rhein Westphalia, and in the City of

Krefeld through a specialist website (Noordt & Misuraca 2019). From the United Arab Emirates, Dubai Electricity and Water Authority (DEWA) allows people to communicate with the Authority through various technology-enabled mechanisms (internet, web, e-mail, telephone) for specific purposes (account setting up or closure, electronic payment, or suggestions for improvement). DEWA has an artificial intelligence-based virtual assistant called Rammas, which guides customers by responding to their enquiries in real time. Rammas is available on various platforms including Android, iOS, Facebook page, DEWA website, Google Assistant in addition to Alexa from Amazon (Parahoo & Ayyagari 2019). Also, Roads and Transport Authority of Dubai (RTA) announced their (Mahboub) chatbot, a sophisticated AI initiative which provides residents of Dubai with information in both languages (Arabic/English) about transportation, services, balance, in addition to planning for the journey using public transportation (Mohasses 2019).

2.11. Summary

In summary, this chapter gives an overview on chatbots since the beginning with shedding the light on their types, architecture including components and approaches. General challenges of Arabic language have been presented in addition to other challenges related to similarity measurement. In addition, similarity measurement techniques and evaluation methodology have been thoroughly demonstrated. A survey on the Arabic contribution on chatbots have been shown as well. Finally, various application of chatbots have been discussed from multiple sectors.

3. CHAPTER 3: RESEARCH METHODOLOGY

3.1. Introduction

This chapter presents the proposed approach in constructing an Arabic chatbot (Linguistic - Textual Dialog System) using ontology of the government services. The ontology that was used in this study builds on a previously built ontology for Dubai government services designed by Albarghothi (2018). The necessary updates and enhancements were applied. A novel model for constructing the chatbot knowledge base is proposed in this work with an optimum objective that enables the chatbot to answer multiple forms of questions related to Dubai Government services and achieve the highest possible results of accurate responses for a specific query. Thus, the proposed concept relies on extracting the knowledge from the ontology with all services attributes. In addition, it shows how to automatically build the knowledge base through various mapping among the extracted dataset structure and the main components of the chatbot AIML knowledge base. In the proposed methodology, an algorithm of multiple search stages has been designed to maximize the responses accuracy depending on different approaches that will be presented in details through this chapter. The main value of this work is to focus on Arabic language to produce a domain specific textual chatbot. As indicated before, the earlier efforts of producing Arabic chatbot is quite limited. So, the designed the unified chatbot integrates all services of Dubai government with the ability to respond instantly and independently based on the three combined approaches, including: Pattern Matching, NLP and Similarity Measurement.

The proposed work adopted the Knowledge Discovery in Database (KDD) methodology by Azevedo and Santos (2008). KDD applies data mining techniques to uncover useful hidden knowledge within the data (Alkashri et al. 2020), (Siyam, Alqaryouti & Abdallah 2020). This is presented through (Figure 3.1), which combines all phases for constructing the knowledge base resources and implementing the chatbot. In brief, the main phases covered are: Extract Ontology Knowledge Base, Construct Chatbot Knowledge Base, and Develop Chatbot Response Algorithm.

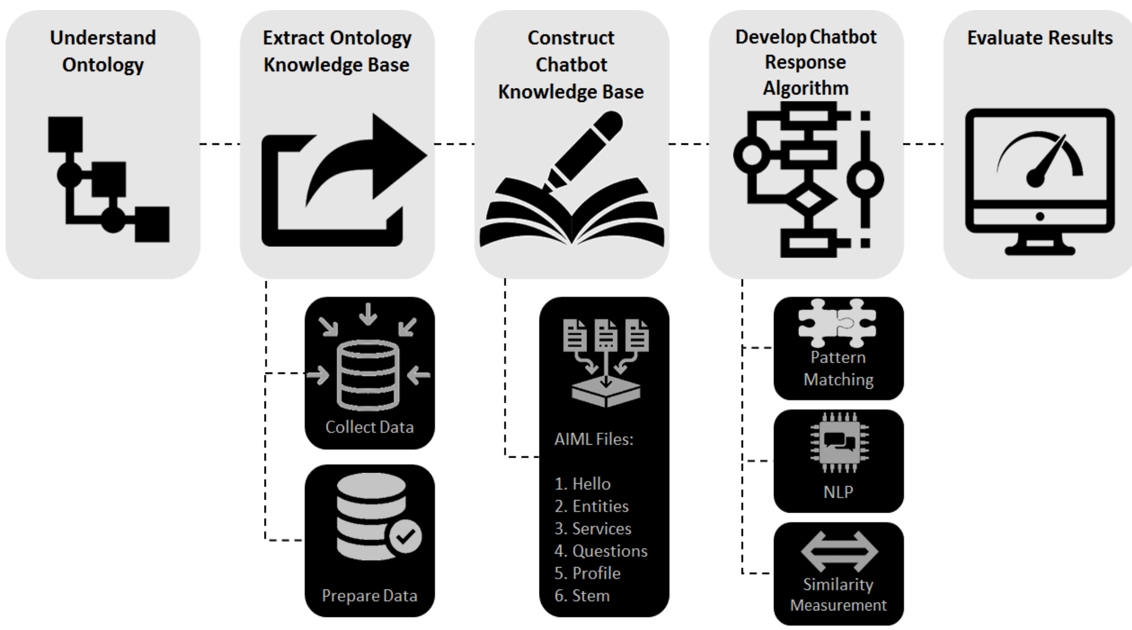

Figure 3.1: Research Methodology

3.2. Extract the Ontology Knowledge Base

3.2.1. Data Collection

As mentioned before, the knowledge base of the chatbot was extracted from a previously built ontology. Therefore, the ontology data source in addition to data extraction and validation will be discussed in this sub-section.

3.2.1.1. Data Source

According to a survey that has been conducted by the United Nations for e-government, the UAE was ranked 1st in the Online Services Index (OSI) regionally and 8th globally, that has been conducted (Gulf News 2020). This research focused on the services of Dubai government. Every government entity in Dubai follows a national standard and has to fulfil its services designing and delivery according to the directives of the Executive Council (Dubai Model) and the Prime Minister's Office (Global Stars Rating). In general, each service contains certain attributes in a structured system: service name, service definition, service requirements, service fees, service procedures, service delivery channels, and service time or contacts channels. The actual data source of the ontology is available on (دبي.الامارات) which presents 504 services distributed among 32 government entities. So, the final dataset is shown in (Table 3.1):

#	Entity Name (EN)	Entity Name (AR)	No. of Services
1	Dubai Municipality	بلدية دبي	30
2	Roads and Transport Authority	هيئة الطرق والمواصلات	56
3	Dubai Police	شرطة دبي	18
4	Dubai Civil Defense	الادارة العامة للدفاع المدني - دبي	12
5	Dubai Courts	محاكم دبي	122
6	Islamic Affairs and Charitable Activities	دائرة الشؤون الإسلامية والعمل الخيري	11
7	Public Prosecution – Dubai	النيابة العامة - دبي	20
8	Awqaf and Minors Affairs Foundation	مؤسسة الأوقاف وشؤون القصّر	4
9	Department of Economic Development	دائرة التنمية الاقتصادية	25
10	Land Department	دائرة الأراضي والأملاك	23
11	Dubai Electricity and Water Authority	هيئة كهرباء و مياه دبي	25
12	Dubai Chamber of Commerce and Industry	غرفة تجارة وصناعة دبي	20
13	Dubai Airports	مطارات دبي	6
14	Dubai Airport Free Zone Authority	المنطقة الحرة بمطار دبي	11

15	Department of Tourism and Commerce Marketing	دائرة السياحة والتسويق التجاري	10
16	Dubai Health Authority	هيئة الصحة بدبي	50
17	General Directorate of Residency and Foreigners Affairs – Dubai	الإدارة العامة للإقامة وشؤون الأجانب - دبي	11
18	Dubai Customs	جمارك دبي	10
19	Mohammed Bin Rashid Housing Establishment	مؤسسة محمد بن راشد للإسكان	9
20	Community Development Authority	هيئة تنمية المجتمع	1
21	Dubai Cares	دبي العطاء	1
22	Zakat Fund	صندوق الزكاة	2
23	Dubai Healthcare City	مدينة دبي الطبية	2
24	Knowledge and Human Development Authority	هيئة المعرفة والتنمية البشرية	4
25	Ministry of Labor	وزارة العمل	2
26	Dubai Culture	هيئة دبي للثقافة والفنون	4
27	Dubai International Academic City	مدينة دبي الأكاديمية العالمية	2
28	Dubai Fishermen Cooperative Society	جمعية دبي التعاونية لصيادي الأسماك	3
29	Commission for Academic Accreditation	الهيئة الوطنية للتقويم والاعتماد الاكاديمي	1
30	Emirates National Development Program	برنامج الإمارات لتطوير الكوادر الوطنية	1
31	Ministry of Social Affairs	وزارة الشؤون الاجتماعية	1
32	National Human Resource Development and Employment Authority	هيئة تنمية وتوظيف الموارد البشرية الوطنية	1
	Services for unknown entities	خدمات جهات غير معرفة	6
		Total	**504**

Table 3.1: Services Distribution among Government Entities

3.2.1.2. Data Extraction

Understanding ontology is the initial step in the proposed methodology to construct a chatbot based on a previously built ontology. Therefore, the representations in the ontology structure was analyzed in order to map the ontology knowledge base components to the chatbot knowledge base elements. (Figure 3.2) shows the structure of the ontology according to the Artificial Intelligence Markup Language (AIML).

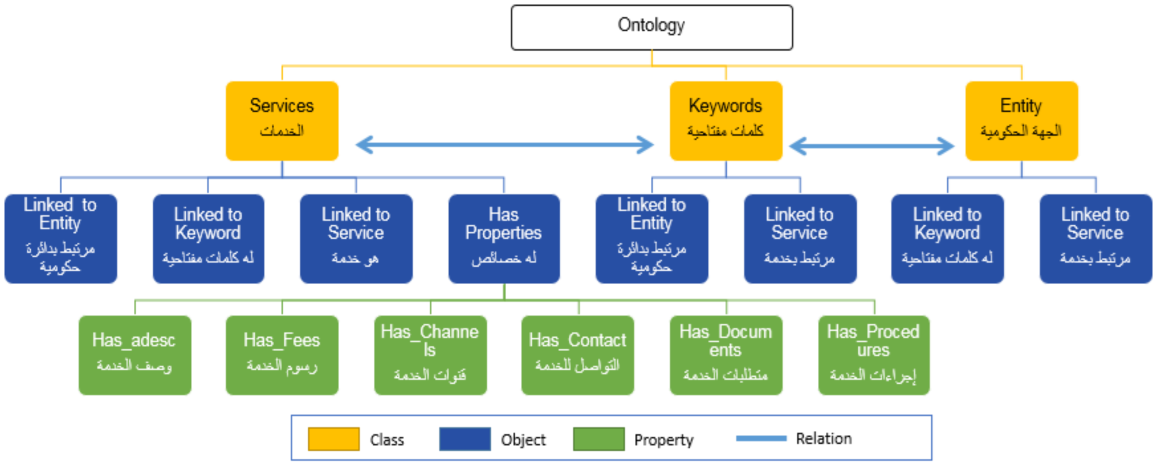

Figure 3.2: Ontology Structure by Albarghothi (2018)

On the other hand, the knowledge base of the chatbot was built based on Artificial Intelligence Markup Language (AIML) files. Through a set of AIML files, the chatbot "personality" is created, where each AIML file consist of a group query-response models that named categories. The <category> mainly consist of <pattern>, which is the "query", in addition to the <template>, which is the response. (Figure 3.3) is showing the knowledge base mapping between the ontology and the chatbot.

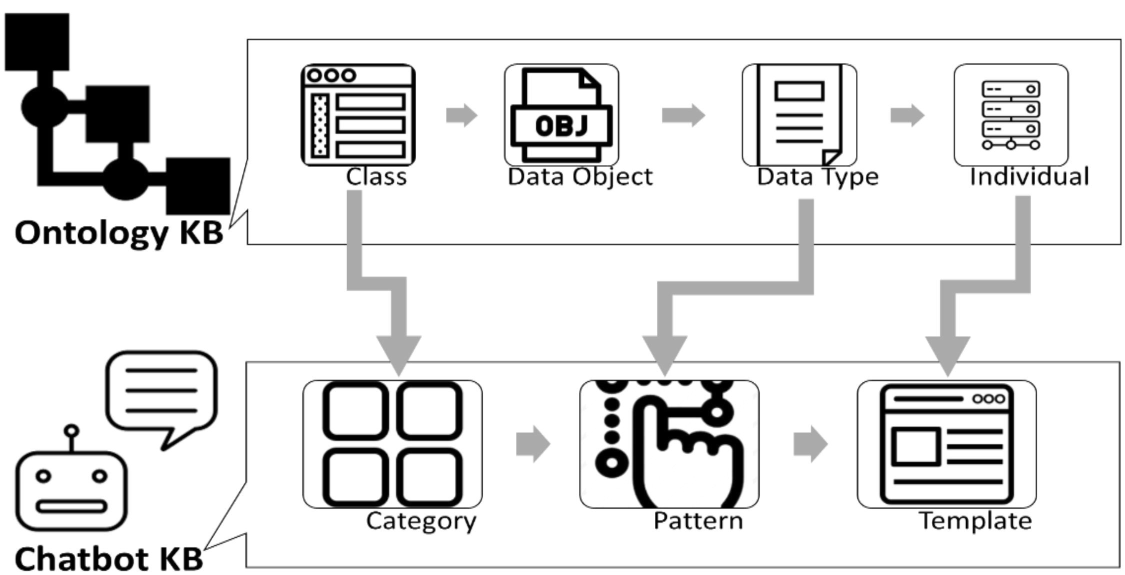

Figure 3.3: Knowledge Base Mapping from AIML to the proposed Chatbot

3.2.1.3. Data Update

The extracted data requires to be updated to make sure that the chatbot's knowledge base is up to date as per the latest information published on the actual data source. Therefore, we relied on the same tool which was developed by Albarghothi, Saber and Shaalan (2018) to Automatically Extract Dataset System (AEDS) to generate a new dataset and compare it with the extracted data from the ontology, and update the service attributes accordingly.

3.2.2. Data Preparation

3.2.2.1. Data Validation

Prior to the construction of the chatbot knowledge base, data validation is required to make sure that the updated data is accurate. Therefore, a team of two domain experts has been appointed to accomplish this process through multiple activities starting from dataset revision, validating services per entity, available attributes per service, and finally approve the service to be included in the knowledge base construction. These activities are detailed below:

- Dataset Revision: Reviewing the structure of the service, and ensure capturing all attributes of the services, with verifying the services count per entity.
- Services Validation: Cross check each service in the dataset is provided by the entity, cross check the accuracy of the service attributes, evaluate the quality of the provided information in each service, identify services that has issues or inaccurate information.
- Attributes Availability: Avoid any mixing of attributes details between services, and ensure that all attributes details belong to the right service.

- Service Acceptance: Either approve the service and accept it or deny the service and reject it by flagging the service by (1) or (0) respectively.

3.2.2.2. Data Quality

Alqaryouti, Siyam and Shaalan (2018) claim that the inter-annotator agreement is usually used to assess the reliability of their validation process and to analyze the degree of agreement between both annotators. It also guarantees that the annotators have a clear and correct understanding of the required validation process and ensure high level of confidence. Therefore, the inter-annotator agreement was used to match the final outcomes of both domain experts that have been hired to finalize the validation process based on kappa statistic, which is mostly used to check the efficiency of the inter-raters (McHugh 2012). The author also stated that the while using two raters, Cohen's kappa can be applied. The results of Cohen's kappa ranges from (-1) to (+1), where the negative value indicates "disagreement", while (0) indicates "no agreement", and (1) means "perfect agreement".

$$k \equiv \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

Where:

- k *is Kappa*
- $Pr(a)$ *is the observed agreement.*
- $Pr(e)$ *is the chance agreement.*

Based on the above measures, the Cohen's Kappa ($\mathcal{K}$) results is 90.06%. (Takala et al. 2010) stated that if the agreement exceeds 81% it will be considered almost perfect. This result shows that domain experts could perceive a proper and common understanding of the services and assures a high level of confidence in the validation process.

3.2.2.3. Data Normalization

It is very important to normalize the data before constructing the knowledge base of a chatbot. However, due to inconsistencies in writing Arabic scripts, there are some challenges in accomplishing this process (Areed et al. 2020). The normalization process focused on the service name attribute only, since this attribute is what would be mentioned by the user while inquiring about a service. Other attributes were not included in this process as they will be displayed as they are without any need to apply any normalization on them. In this process, all diacritics and punctuation marks have been removed from the service name, such as "\", "(", ")", "-", ".". For services that has "/" or "," in their names, the characters were replaced with "او" which means "or" as it was the actual purpose of these characters. Therefore, the number of services that have been normalized is 55 services, which almost represents 11% of the total services that has been included in this work. Some examples of these normalized services are shown in (Table 3.2):

Service Name Before Normalization	Service Name After Normalization
طلب تصوير الكتب و الرسائل الجامعية (فهارس)	طلب تصوير الكتب و الرسائل الجامعيه فهارس
خدمة قوائم المؤسسات المعتمدة (جمعيات/مراكز/ مؤسسات)	خدمة قوائم المؤسسات المعتمده جمعيات او مراكز او مؤسسات

Table 3.2: Service Name Normalization Examples

3.2.2.4. Questions Mapping

To construct a knowledge base for a chatbot using pattern matching approach, dataset needs to be analyzed and query rules need to be identified in order to setup the question structure for each service. This analysis will help in expanding the knowledge base to cover the maximum number of question forms and will reduce the mismatch possibility which will result in no response at the end. As mentioned earlier, each service contains certain attributes: name, definition, requirements, fees, procedures, channels, and time or contact channels. These attributes are the keywords that could be available in the question while enquiring about the service. Therefore, multiple forms of a question for a government service have been mapped using other matching keywords in Arabic as shown in (Table 3.3):

Attribute (EN)	Attribute (AR)	Question Mapping (1)	Question Mapping (2)	Question Mapping (3)
Definition	وصف	ما هو وصف خدمة	ما هو تعريف خدمة	ما هي خدمة
Requirements	الاوراق المطلوبة	ما هي الاوراق المطلوبة خدمة	ما هي الوثائق المطلوبة لخدمة	ما هي المستندات المطلوبة لخدمة
Fees	رسوم	ما هي رسوم خدمة	ما هي تكلفة خدمة	ما هو سعر خدمة
Procedures	اجراءات	ما هي اجراءات خدمة	ما هي خطوات خدمة	كيف اتقدم بطلب خدمة
Channels	قنوات	ما هي قنوات تقديم خدمة	من أين يمكن الحصول على خدمة	اين يمكن تقديم معاملة خدمة
Time / Contact	الاوقات / تواصل	ما هي ارقام التواصل لخدمة	ما هي ساعات العمل لخدمة	كيف يمكن الاستفسار عن خدمة

Table 3.3: Examples of Attributes - Question Mapping for a Government Service

To explain the above table more, if a user makes a query with the following question:

ما هي المستندات اللازمة لخدمة طلب اعتماد الإفراج عن شحنة غذائية مستوردة للبيع بالسوق المحلي؟

The red colored text in above question is mapped to the following attribute:

الاوراق المطلوبة

Where this attribute is linked with a service and defined as a pattern in the knowledge base.

3.3. Construct Chatbot Knowledge Base

As mentioned before, the files format of the chatbot knowledge base is AIML files. The chatbot application stores all AIML files in a tree to generate some sort of repository that manages query-response. When a user types a text as a query, the application searches in the tree for a matching <pattern>, then it responds back with the associated <template>. For queries that has the same answer of a predefined <pattern>, the recursive tag <srai> can be used to refer to them. Such categories can be organized with the use of a few markup tags to generate more complex human-like responses (Wallace 2009). Multiple AIML files have been created and explained in the following sub-sections.

3.3.1. Hello AIML File

The purpose of this file is to cover the initiation of the conversation between the user and the chatbot by welcoming the user and introduce him to the chatbot. The total number of categories prepared in this file is 25, which consist of different Arabic salutes in both dialects and MSA. Some sample of this file is shown in (Figure 3.4) and examples of the chatbot responses are provided in (Figure 3.5) and (Figure 3.6):

```
<category>
<pattern>مرحبا</pattern>
<template>
    <random>
        <li>اهلا وسهلا</li>
        <li>حياكم الله</li>
        <li>يا اهلا وسهلا</li>
        <li>مرحبا بكم</li>
    </random>
</template>
</category>

<category>
<pattern>السلام عليكم</pattern>
<template>
<srai>السلام عليكم ورحمة الله وبركاته</srai>
</template>
</category>
```

Figure 3.4: Hello File Sample

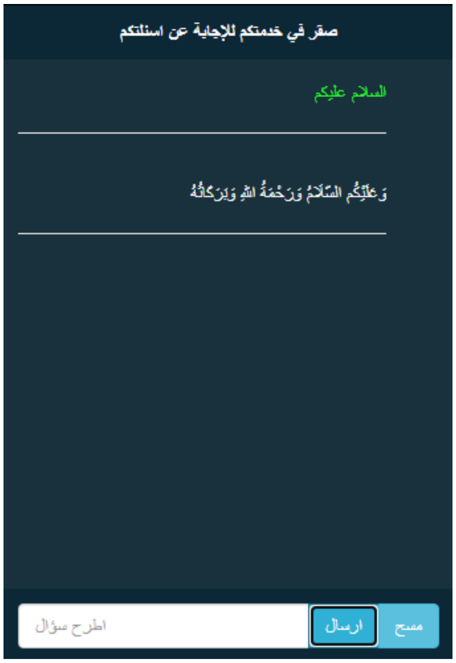

Figure 3.5: Chatbot Response 1

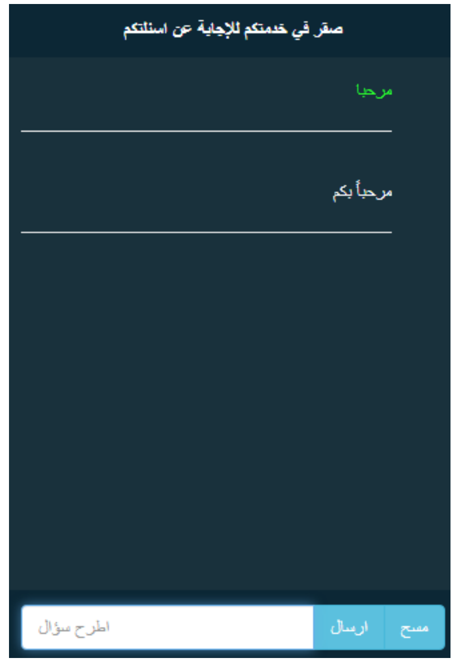

Figure 3.6: Chatbot Response 2

3.3.2. Entities AIML File

The purpose of this file is to capture any input that has not been defined in the knowledge base. It will respond automatically with a sentence that explains the purpose of this chatbot and redirect the user to a list of 32 entities to choose from, and then it segregates the 504 services according to the related entity so the user will be able to know which service to enquire about. Accordingly, the total number of categories is 32 in this AIML file. Samples of this file is shown in (Figure 3.7) and examples of the chatbot responses are provided in (Figure 3.8) and (Figure 3.9):

```xml
<category><pattern>*</pattern>
<template>
        <![CDATA[
            يجب ان يتم الاستفسار عن الخدمة بحسب اسمها الاصلي، يمكنك كتابة رقم الج
            هة الحكومية من القائمة التالية لمعرفة اسماء الخدمات التي تقدمها </br>
                1. بلدية دبي</br>
                2. هيئة الطرق و المواصلات</br>
                3. شرطة دبي</br>
                4. الادارة العامة للدفاع المدني - دبي</br>
                5. محاكم دبي</br>
                6. دائرة الشؤون الإسلامية والعمل الخيري</br>
                7. النيابة العامة - دبي</br>
                8. مؤسسة الأوقاف وشؤون القصّر</br>
                9. دائرة التنمية الاقتصادية</br>
                10. دائرة الأراضي والأملاك</br>
                11. هيئة كهرباء و مياه دبي</br>
                12. غرفة تجارة وصناعة دبي</br>
                13. مطارات دبي</br>
                14. المنطقة الحرة بمطار دبي</br>
                15. دائرة السياحة والتسويق التجاري</br>
                16. هيئة الصحة بدبي</br>
                17. الإدارة العامة للإقامة وشؤون الأجانب - دبي</br>
                18. جمارك دبي</br>
                19. مؤسسة محمد بن راشد للإسكان</br>
                20. هيئة تنمية المجتمع</br>
```

Figure 3.7: Entities AIML File Sample

Figure 3.8: Chatbot Response 1

Figure 3.9: Chatbot Response 2

3.3.3. Services AIML File

This is the main services file, which includes the responses on each query related to any attribute. The final number of services after validation was 504, and there were 6 attributes extracted for each service. Therefore, considering that each attribute has a single form of query, then 3,024 (504*6) categories can be generated and included in this AIML file. Samples of this file is shown in (Figure 3.10) and examples of the chatbot responses are provided in (Figure 3.11) and (Figure 3.12):

```
<category>
    <pattern>وصف تقديم بيان جمركي</pattern>
    <template>
        <![CDATA[تقوم هذه الخدمة بالسماح للمتعاملين أو المسافرين بعمل بيان جمركي و تحصيل التخليص
        اللازم لنقل البضائع داخل و خارج إمارة دبي والتي تنحصر في الأنواع التالية:<br />
            1. تصدير جمركي<br />
            2. استيراد جمركي<br />
            3. عبور جمركي<br />
            4. تحويل جمركي<br />
            5. إدخال مؤقت<br />
        ]]>
    </template>
</category>

<category>
    <pattern>رسوم تقديم بيان جمركي</pattern>
    <template>
        <![CDATA[تختلف الرسوم وفقًا لنوع البيان الجمركي وقناة الشحن:<br />
            1. تصدير جمركي: د.إ 15.00 - د.إ 100.00<br />
            2. استيراد جمركي: د.إ 15.00 - د.إ 100.00<br />
            3. عبور جمركي: د.إ 15.00 - د.إ 80.00<br />
            4. تحويل جمركي: د.إ 80.00<br />
            5. إدخال مؤقت: د.إ 100.00<br />
            سيتم اضافة رسوم درهم المعرفة و الابتكار بقيمة 20 درهم على
            اي خدمة قيمتها 50 درهم إماراتي فما فوق<br />
        ]]>
    </template>
</category>
```

Figure 3.10: Services AIML File

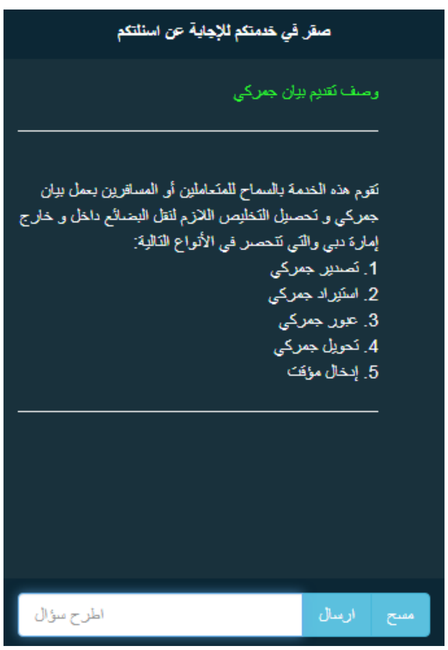

Figure 3.11: Chatbot Response 1

Figure 3.12: Chatbot Response 2

3.3.4. Questions AIML File

Using question mapping technique explained in section (3.2.2.4) and the recursive <srai> tag, the knowledge base has been expanded to cover more possible forms of questions. Therefore, as shown in Table 2, another 55 forms of questions have been added and mapped to the related attribute, which generated 27,720 new categories to be included in this AIML file and enriches the knowledge base as shown in (Figure 3.13). The query shown in the above examples are repeated again using different forms as shown in the below examples. However, the response shown in (Figure 3.14) and (Figure 3.15) are still the same since their queries were mapped to the same query above:

```
<category>
    <pattern>ما هي خدمة تقديم بيان جمركي</pattern>
    <template><srai>وصف تقديم بيان جمركي</srai></template>
</category>

<category>
    <pattern>ما هو سعر تقديم بيان جمركي</pattern>
    <template><srai>رسوم تقديم بيان جمركي</srai></template>
</category>
```

Figure 3.13: Questions AIML File Sample

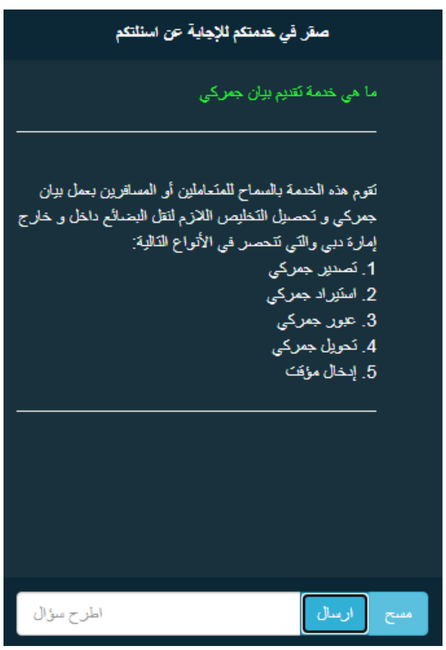

Figure 3.14: Chatbot Response 1

Figure 3.15: Chatbot Response 2

3.3.5. Profile AIML File

If the user wishes to get all information related to a service, the full name of the service must be typed. This action uses the profile AIML file as it holds a complete profile for each service. So obviously, it consists of 504 categories (a category for each service profile). Some samples of this file is shown in (Figure 3.16) and an example of the chatbot response with the complete service profile is provided in (Figure 3.17):

```
<category>
    <pattern>طلب إصدار او تجديد بطاقة الصحة المهنية</pattern>
    <template>
        <![CDATA[<h4 style="color:yellow"><b>أسم الخدمة:</b></h4>
        طلب إصدار او تجديد بطاقة الصحة المهنية
        <h4 style="color:yellow"><b>وصف الخدمة:</b></h4>
        تتيح الخدمة الحصول على بطاقة الصحة المهنية، ويقدم هذا الطلب بشكل إلزامي من قبل العاملين في المؤسسات الصحية وذات العلاقة بالصحة العامة مثل
        المطاعم، والفنادق، والمصانع، ودور الحضانات، وصالونات الحلاقة وغيرها والعاملين في المهن الخطرة والمحولين من إدارة الصحة والسلامة العامة  تقدم
        الخدمة من قبل إدارة خدمات الصحة العامة
        ملاحظات:
        1  يجب إصدار البطاقة خلال 30 يوما من دخول البلد أو إصدار الإقامة
        2  البطاقة سارية المفعول لمدة عام واحد
        3  يتم تجديد البطاقة سنوياً خلال 30 يوم من تاريخ الانتهاء
        4  يتم تسليم البطاقات الصحية من الساعة 12:00 إلى الساعة 2:00 خلال أيام العمل الرسمي
        5  يتم دفع غرامة في حال التأخر عن إصدار تجديد البطاقة   إجراء الفحوصات الطبية في الموعد المحدد
        <h4 style="color:yellow"><b>الأوراق المطلوبة:</b></h4>
        1. صورة من جواز السفر2. صورة شخصية حديثة3. صورة تأشيرة الدخول/ الإقامة4. البطاقة الصحية المهنية القديمة (في حال التجديد)5. إشعار الدفع6.
        ايصال الدفع
        <h4 style="color:yellow"><b>رسوم الخدمة:</b></h4>
        1. رسم إصدار/ تجديد بطاقة الصحة المهنية: 100.0 درهم2. غرامة عدم الحصول على بطاقة الصحة المهنية أو تجديدها أو التأخر في إجراء الفحوصات
        الطبية في الموعد المحدد.: 300.0 درهم3. رسم بدل فاقد لإصدار بطاقة الصحة المهنية خلال 3 شهور من إصدارها: 20.0 درهم4. إصدار بدل فاقد لإيصال
        بطاقة الصحة المهنية: 50.0 درهم
        <h4 style="color:yellow"><b>إجراءات تقديم الخدمة:</b></h4>
        هذه المعلومة غير متوفرة لدي حالياً
        <h4 style="color:yellow"><b>قنوات تقديم الخدمة:</b></h4>
        متاحة إلكترونياً تقليدية غير إلكترونية الأنترنتموظف - التطبيق الهجين للهواتف المحمولة - التطبيق الهجين للأجهزة اللوحية المحمولة - الموقع
        الإلكتروني المخصص للأجهزة المحمولة - https//portal.dm.gov.ae/SCWebUI/DataDetails.aspx?servicecode=3163
        <h4 style="color:yellow"><b>أرقام التواصل:</b></h4>
        هاتف خدمة العملاء: 800900فاكس خدمة العملاء: 04-7033645ساعات عمل خدمة العملاء: 7:30 am - البريد الإلكتروني لخدمة العملاء: smmulla@dm.gov.ae
        2:30pm]]>
    </template>
</category>
```

Figure 3.16: Profile AIML File

Figure 3.17: Chatbot Response

3.3.6. Stem AIML Files

To create this file, a special python program was designed to extract all patterns in the Service AIML file and stem them using Natural Language Possessing techniques, then point them back using the <srai> tag to the same related pattern. The purpose of this file is to lookup for any matching <pattern> after applying NLP processes to the query that could be made by the user in order to search again in case of not finding a match using the first attempt. Some samples of this file are shown in (Figure 3.18):

```xml
<category>
    <pattern>وصف طلب صدر جدد بطق صحة من</pattern>
    <template><srai>وصف طلب إصدار او تجديد بطاقة الصحة المهنية</srai></template>
</category>

<category>
    <pattern>قنو طلب عمد فرج شحن غذي ورد بيع سوق حل</pattern>
    <template><srai>قنوات طلب اعتماد الإفراج عن شحنة غذائية مستوردة للبيع بالسوق المحلي</srai></template>
</category>

<category>
    <pattern>جرء عرف دفع خلف سل</pattern>
    <template><srai>اجراءات استعراض و دفع مخالفات سالك</srai></template>
</category>
```

Figure 3.18: Stem AIML Files

The total number of categories created in this file was matching to the exact number of categories in the Service AIML file, which is 3,024. An example of the chatbot response for this AIML file is presented in section 3.2.4 (Figure 3.19).

3.4. Develop Chatbot Response Algorithm

One of the most important challenges that faces chatbot developers is how to maintain different contexts. Users may type their quires using their own expressions depending on their background and level of education. Therefore, it was important to take into consideration this factor while designing the chatbot. The pattern matching approach is helpful but it will remain limited to the exact match. There is a great possibility for the users not to type their query exactly as defined in the knowledge base. However, applying multiple approaches of query analysis and answer retrieval will improve the chatbot performance significantly and insures maximum response with higher level of accuracy. For that reason another 2 approaches in addition to the patter matching were adopted to be executed sequentially on 3 stages. Below is a detailed explanation of these stages and how they perform:

3.4.1. Stage 1: Pattern Matching (PM)

This is the initial stage, where pattern matching approach is performed depending completely on what have been defined in the knowledge base. In other words, if the query typed by the user matches any <pattern> in the chatbot brain which consists of all AIML files, the <template> of this particular <pattern> will be retrieved and displayed to the user directly as a response. All chatbot responses shown in the examples provided in section (3.3.4) have been retrieved using this approach.

3.4.2. Stage 2: Natural Language Processing (NLP)

This stage will be triggered in case of not finding any match in stage 1 (Pattern Matching). Natural Language Processing will be applied on the user's input as shown below assuming the user provided this input as an example:

"ما هي خدمة طلب إصدار تجديد بطاقة الصحة المهنية؟"

3.4.2.1. Normalization

This process is executed to remove (Hamza - ء) and (Madda - ~) from letters like (Alif – أ إ آ ى), it also converts (Taa Marbotta – ة) to (ه). The outcome for the above example after normalization is:

ما هي خدمه طلب اصدار تجديد بطاقه الصحه المهنيه؟

3.4.2.2. Word Tokenization

Word tokenization is the process of breaking a sentence into individual words where each word needs to be captured for further analysis, such as counting, classifying… etc. The outcome for the above example after tokenization is:

'[ما', 'هي', 'خدمه', 'طلب', 'اصدار', 'تجديد', 'بطاقه', 'الصحه', 'المهنيه]'

3.4.2.3. Find and Remove Stop Word

This process will find all stop words such as "؟", "هي", "ما"… etc. and eliminate them after tokenization. Therefore, the outcome of the above example after removing all stop words is:

['خدمه', 'طلب', 'اصدار', 'تجديد', 'بطاقه', 'الصحه', 'المهنيه']

3.4.2.4. Stem Words

At this process the words are reduced to their word stem by excluding prefixes and suffixes with all feminine pronouns. The outcome for the above example after stemming is:

خدم طلب صدر جدد بطق صحه هن

After applying all the above NLP processes, the chatbot application will look up for a matching <pattern> in the chatbot brain in order to retrieve the related <template> as a response. This is simulated in (Figure 3.19):

Figure 3.19: Chatbot Response based on NLP

3.4.3. Stage 3: Similarity Measurement (SM)

If the application failed in finding a match using the previous stages, then this stage is to be applied to find the best similar match. However, this requires additional information to be fed to the application in order to cover more contexts. A synonyms dataset has been designed to map the keyword in each attribute with the most possible synonyms. Some examples of the dataset are shown in (Table 3.4):

Keywords (EN)	Definition	Requirements	Fees	Procedures	Channels	Time / Contact
Keywords (AR)	**وصف**	**الاوراق المطلوبة**	**رسوم**	**اجراءات**	**قنوات**	**ارقــــام تواصل**
Mapped Synonyms	تعريف	المستندات	تكلفة	خطوات	مواقع	استفسار
	شرح	الوثائق	سعر	عملية	اين	اوقات
	التعريف	المستند	كم	طريقة	اماكن	ساعات
	الشرح	الوثيقة	تسعيرة	خطوة	مكان	مواعيد

Table 3.4: Examples of Keywords - Synonyms Dataset Mapping

The first thing to do at this stage is applying tokenization on the user's input. Assuming the user has provided this query:

"ما هي المستندات اللازمة إلغاء ترخيص العمل الجزئي"

After tokenization, the query will be:

['ما', 'هي', 'المستندات', 'اللازمة', 'إلغاء', 'ترخيص', 'العمل', 'الجزئي']

Then the application looks at any synonyms that matches with the above dataset. If it could find any match, the mapped keyword will be replaced as shown below:

['ما', 'هي', 'الاوراق المطلوبة', 'اللازمة', 'إلغاء', 'ترخيص', 'العمل', 'الجزئي']

Next, it compares the updated input with all patterns in the Service AIML file inside the chatbot brain and measures the sentence similarity with the words similarity and combine them to find the highest 3 similar pattern. The results will be displayed to the user to choose the preferred one for his query as shown in (Figure 3.20). Once the user select any of these three options, it will retrieve the related <template> as a final response as shown in (Figure 3.21).

Figure 3.20: Best 3 Similar Match

Figure 3.21: First Match Selection

3.5. Evaluate Results

There have been numerous attempts to boost the chatbot efficiency and improve the response algorithm. The evaluation of the chatbot was performed according to various settings and criteria. Therefore, standard confusion matrix was adopted to measure the performance and compare it with the ontology built by Albarghothi (2018), and Rashid chatbot, the official chatbot of Smart Dubai Government, based on the same set of quires, i.e. 414 questions, measured by Albarghothi (2018) previously.

The following standard formulas determines the evaluation measures of the performance (Alqaryouti et al. 2018):

1. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
2. $Precision = \frac{TP}{TP+FP}$
3. $Recall = \frac{TP}{TP+FN}$

4. $F - measure = \frac{2 * Precision * Recall}{Precision + Recall}$

(Table 3.5) illustrates the confusion matrix elements for evaluation measures:

	Retrieved	**Not Retrieved**
Relevant	**TP (True Positive)** Number of queries that are correctly responded by the algorithm	**FP (False Positive)** Number of queries that are responded incorrectly by the algorithm
Irrelevant	**FN (False Negative)** Number of queries that are not correct but responded by the algorithm	**TN (True Negative)** Number of queries that are not correct and not responded by the algorithm

Table 3.5: Confusion Matrix

3.6. Summary

In summary, this chapter demonstrated the research methodology in constructing Arabic chatbot for government services based on a previously built ontology. The methodology followed in this work consist of five phases, the initial phase was to understand the ontology. Then, second phase showed the data extraction of the knowledge base and who it has been updated based on the latest source of data. Later, phase three presented how chatbot knowledge base has been constructed. After that, phase four discussed the developed algorithm to enhance the chatbot response. Finally, phase five planed the results evaluation approach to be adopted in estimating the performance.

4. CHAPTER 4: EXPERIMENTS AND RESULTS DISCUSSION

4.1. Introduction

This chapter presents a novel approach for examining the chatbot response algorithm based on the evaluation approach discussed in section (3.5). There were 414 questions to be examined with the chatbot in order to compare the three response stages with Albarghothi (2018) ontology, and Rashid chatbot of Smart Dubai Government. Therefore, 3 experimental tests were executed prior to the comparison. After each attempt, some updates where added to the chatbot application to improve the similarity measurement approach. These updates with their results are explained in the rest of sections within this chapter. However, it was noticed that during these three attempts the results of the PM stage and NLP stage are the same which is in fact logical.

4.2. Experiment Attempt 1

The results of executing the first experiment attempt showed the following results for each stage of the proposed algorithm:

4.2.1. Stage 1: Pattern Matching

Stage 1 result was 25% for the Precision, 100% for the Recall, and 41% for the F-Measure with an Accuracy of 27% only.

	Performance Evaluation			
	Precision	**Recall**	**F-Measure**	**Accuracy**
Stage 1: Pattern Matching	25%	100%	41%	27%

Table 4.1: Experiment Attempt 1 - Stage 1 (PM Results)

4.2.2. Stage 2: Natural Language Processing

Stage 2 result was 20% for the Precision, 100% for the Recall, and 33% for the F-Measure with an Accuracy of 22% only.

	Performance Evaluation			
	Precision	Recall	F-Measure	Accuracy
Stage 2: Natural Language Processing	20%	100%	33%	22%

Table 4.2: Experiment Attempt 1 - Stage 2 (NLP Results)

4.2.3. Stage 3: Similarity Measurement

Stage 3 result was 45% for the Precision, 100% for the Recall, and 62% for the F-Measure with an Accuracy of 46%.

	Performance Evaluation			
	Precision	Recall	F-Measure	Accuracy
Stage 3: Similarity Measurement	45%	100%	62%	46%

Table 4.3: Experiment Attempt 1 - Stage 3 (SM Results)

4.2.4. Combined Stages Results

As explained in section (3.4), the proposed algorithm is a combination of 3 stages (PM, NLP and SM). Therefore, the chatbot responses of the combined 3 stages have shown 90% for the Precision, 100% for the Recall, and 95% for the F-Measure with an Accuracy of 90%.

		Performance Evaluation			
		Precision	Recall	F-Measure	Accuracy
Algorithm Combined Stages	PM	25%	100%	41%	27%
	PM + NLP	45%	100%	62%	47%
	PM + NLP + SM	90%	100%	95%	90%

Table 4.4: Experiment Attempt 1 (Combined Results)

4.3. Experiment Attempt 2

The results of executing the second experiment attempt showed the same results for stage 1 and 2. However, stage 3 results was improved due to changing the comparison file from (Profile AIML File) to (Services AIML File).

4.3.1. Stage 3: Similarity Measurement

At this attempt, stage 3 result was 46% for the Precision, 100% for the Recall, and 63% for the F-Measure with an Accuracy of 48%.

	Performance Evaluation			
	Precision	Recall	F-Measure	Accuracy
Stage 3: Similarity Measurement	46%	100%	63%	48%

Table 4.5: Experiment Attempt 2 - Stage 3 (SM Results)

4.3.2. Combined Stages Results

The changes made in the chatbot application improved the performance to become 91% for the Precision, 100% for the recall, and 95% for the F-Measure with an accuracy of 91%.

		Performance Evaluation			
		Precision	Recall	F-Measure	Accuracy
Algorithm Combined Stages	PM	25%	100%	40%	26%
	PM + NLP	44%	100%	62%	46%
	PM + NLP + SM	91%	100%	95%	91%

Table 4.6: Experiment Attempt 2 (Combined Results)

4.4. Experiment Attempt 3

The results of executing the third experiment attempt showed also the same results for stage 1 and 2. However, the Keywords - Synonyms Dataset Mapping was added to the chatbot application which resulted in a significant improvement.

4.4.1. Stage 3: Similarity Measurement

At this attempt, stage 3 result was 51% for the Precision, 100% for the Recall, and 67% for the F-Measure with an Accuracy of 52%.

	Performance Evaluation			
	Precision	**Recall**	**F-Measure**	**Accuracy**
Stage 3: Similarity Measurement	51%	100%	67%	52%

Table 4.7: Experiment Attempt 3 - Stage 3 (SM Results)

4.4.2. Combined Stages Results

The changes made in the chatbot application improved the performance to become 96% for the Precision, 100% for the recall, and 98% for the F-Measure with an accuracy of 96%.

		Performance Evaluation			
		Precision	**Recall**	**F-Measure**	**Accuracy**
Algorithm Combined Stages	PM	25%	100%	40%	26%
	PM + NLP	44%	100%	62%	46%
	PM + NLP + SM	96%	100%	98%	96%

Table 4.8: Experiment Attempt 3 (Combined Results)

4.5. Results Comparison

The final results of the 3rd experiment have been compared with the results achieved by Albarghothi (2018), and a government services chatbot (Rashid) provided by Smart Dubai

Government (SDG). The comparison with Rashid was performed manually through submitting the 414 questions and evaluate the response of each raised question. (Table 4.9) summaries the results of the 3 models.

		Performance Evaluation			
		Precision	**Recall**	**F-Measure**	**Accuracy**
Comparison	Albarghothi (2018)	95%	94%	94%	90%
	Rashid (SDG)	5%	17%	8%	19%
	Proposed Model	96%	100%	98%	96%

Table 4.9: Results Comparison

4.6. Summary

In summary, this chapter presented the results of the proposed model for building an Arabic chatbot for government services based on a previously built ontology that has been demonstrated in Chapter 3. The updates that have been applied on the chatbot application after each experimental attempt showed a significant improvements as shown in (Figure 4.1) which could outperform the based ontology model of (Albarghothi 2018) as well as Rashid chatbot of SDG.

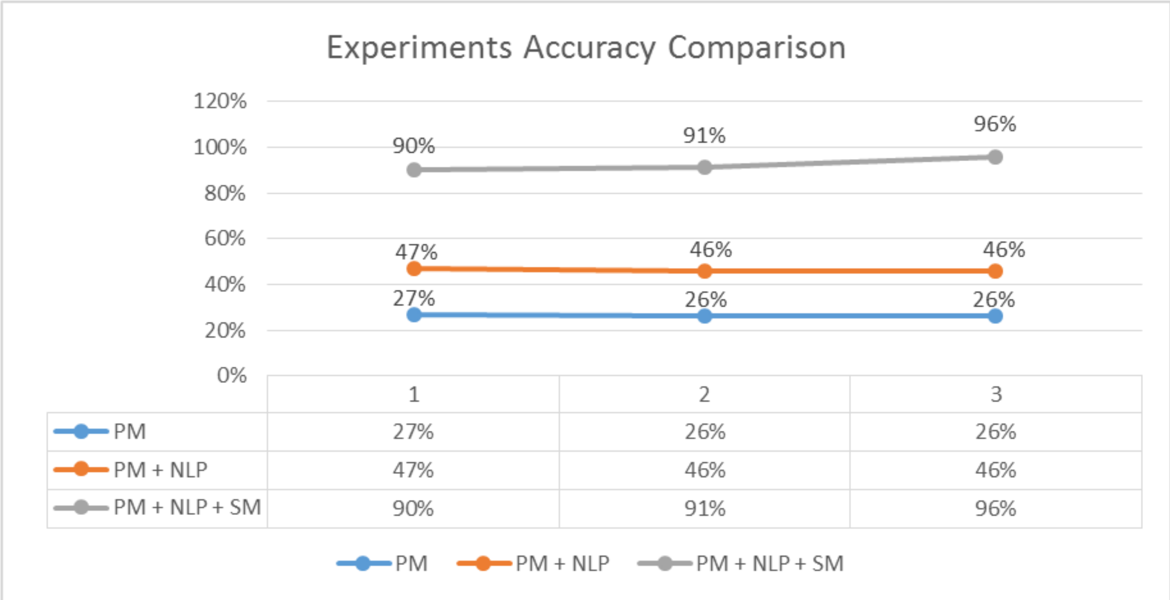

Figure 4.1: Experiments Accuracy Summary

5. CHAPTER 5: CONCLUSION AND FUTURE WORKS

5.1. Introduction

This chapter addresses the answers to the research questions in addition to the conclusion which covers the most important challenges faced throughout this study and the main contribution. Future works are also included to add more value on this work and enhance it further.

5.2. Research Questions Answers

5.2.1. RQ1: Is it possible to build an Arabic Chatbot Knowledge Base for Government Services based on the available Ontology resources?

Yes, as shown in Chapter Three of this dissertation under Sections 3.2, and 3.3. It was possible to extract the knowledge base from a previously built ontology and use it in a different application, which aligns with Al-Zubaide and Issa (2011) for building a chatbot based on ontology through multiple steps and mapping procedure. Moreover, it has been proven that the extracted knowledge base could be used later by the chatbot application in order to respond to government services queries.

5.2.2. RQ2: Is it possible to build an Arabic Chatbot System that can respond to government services related enquirers?

Yes, as shown in Chapter Three of this dissertation under Sections 3.4. It was possible to develop a chatbot response algorithm that has a combination of multiple chatbot approaches to respond to all queries regardless of the query form and without creating a pattern that matches each form of a question. This chatbot was able to unify all government services in Dubai by receiving any service query in Arabic language and respond back accordingly.

5.2.3. RQ3: Is it possible to retrieve accurate answer using Chatbot better than the formal Chatbot of Smart Dubai government?

Yes, as shown in Chapter Three of this dissertation under Sections 3.5. It was possible to outperform the accuracy of the chatbot that has been designed by Smart Dubai Government (Rashid) by achieving 96% compared with 19% accuracy for Rashid which proves a higher capability of the proposed algorithm compered to another chatbot with the same concept.

5.3. Conclusion

The main objective of this dissertation is to investigate the feasibility of building an Arabic chatbot for e-government services based on a previously built ontology. The data source of the previously constructed ontology is (دبـي.امـارات) that contains Dubai governments' services in particular. This work has been carried out through multiple phases. First, it was very necessary to understand the ontology structure. Then, extract the knowledge from the ontology and update it according to the actual data source using AEDS tool, and validate the information manually by the help of domain experts with checking their validation using inter-annotator agreement mechanism. After that, the validated knowledge base has been imported into the chatbot application to become an independent knowledge base using AIML files. Later, a special response algorithm has been designed for the chatbot application to retrieve the most accurate answer and achieve the highest possible score.

Many challenges have been faced throughout this study, the most important one was related to the development of the chatbot application according to the proposed response algorithm which confirms what Razmerita et al. (2004) claimed regarding the difficulty of developing a chatbot. In addition to the challenges associated with Arabic language that have increased this difficulty as well. Multiple experiments have been performed to improve the proposed algorithm, as it has been noticed initially that adopting one approach to build a conversation agent produces poor results, which cannot be released for commercial use. However, adopting a hybrid approach that combines more than one approach resulted in higher accuracy with satisfactory performance. This, of course, requires more time and effort to be developed as it must be designed for a close-domain e.g. government services.

The fundamental contribution of this dissertation is building a unified chatbot that can respond to all services provided by Dubai government using the Arabic language. Moreover, the feasibility of extracting a knowledge base from an ontology and construct an independent knowledge base that can be used by different application have been proved.

5.4. Future Works

This study raised few more ideas that are worth further researching. As a result, there are several possible avenues through which this research might proceed in the future. Such works can be in adopting machine-learning techniques to reduce the possibility of reaching the second or third stage of the proposed algorithm. This also can be done through creating a new pattern in the knowledge base automatically whenever a query is answered through these stages for the same question that have been asked to the chatbot. This will enhance the performance and speed up the application's response time. Another work can be also performed by expanding the knowledge base to cover more services of federal government entities within United Arab Emirates instead of Dubai government only, as well as adding the English language to chatbot with the same approaches in order to make the chatbot application bilingual.

REFERENCES

Abu Shawar, B. & Atwell, E. (2007). Chatbots: are they really useful? *LDV-Forum: Zeitschrift für Computerlinguistik und Sprachtechnologie*, vol. 22(1), pp. 29–49.

Agirre, E., Lopez De Lacalle, O. & Soroa, A. (2009). Knowledge-Based WSD on Specific Domains: Performing Better than Generic Supervised WSD. *IJCAI International Joint Conference on Artificial Intelligence*, pp. 1501–1506.

Al-Chalabi, H., Ray, S. & Shaalan, K. (2016). Semantic based query expansion for Arabic question answering systems. *Proceedings - 1st International Conference on Arabic Computational Linguistics: Advances in Arabic Computational Linguistics, ACLing 2015*, (July), pp. 127–132.

Al-Zubaide, H. & Issa, A. A. (2011). OntBot: Ontology based ChatBot. *2011 4th International Symposium on Innovation in Information and Communication Technology, ISIICT'2011*, pp. 7–12.

Alazemi, N. N., Al-Shehab, A. J. & Alhakem, H. A. (2017). E-government Frameworks based on Semantic Web Services: A Comprehensive Study. *International Journal of Computer Applications*, vol. 158(7), pp. 12–19.

Albarghothi, A. (2018). *An Ontology-based Semantic Web for Arabic Question Answering: The Case of E-Government Services*. *MSc INFORMATICS KNOWLEDGE AND DATA MANAGEMENT The British University in Dubai*.

Albarghothi, A., Khater, F. & Shaalan, K. (2017). Arabic Question Answering Using Ontology. *Procedia Computer Science*. Elsevier B.V., vol. 117, pp. 183–191.

Albarghothi, A., Saber, W. & Shaalan, K. (2018). Automatic Construction of E-Government Services Ontology from Arabic Webpages. *Procedia Computer Science*. Elsevier B.V., vol. 142, pp. 104–113.

AlHumoud, S., Al Wazrah, A. & Aldamegh, W. (2018). Arabic Chatbots: A survey. *International Journal of Advanced Computer Science and Applications*, vol. 9(8), pp. 535–541.

Ali, D. A. & Habash, N. (2016). Botta: An Arabic Dialect Chatbot. *COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: System Demonstrations*, pp. 208–212.

Aliguliyev, R. M. (2009). A new sentence similarity measure and sentence based extractive technique for automatic text summarization. *Expert Systems with Applications*. Elsevier Ltd, vol. 36(4), pp. 7764–7772.

Aljameel, S., O'Shea, J., Crockett, K., Latham, A. & Kaleem, M. (2017). Development of An Arabic Conversational Intelligent Tutoring System for Education of Children with ASD. *International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, vol. 997, pp. 498–516.

Aljlayl, M. & Frieder, O. (2002). On Arabic search: Improving the retrieval effectiveness via a light stemming approach. *International Conference on Information and Knowledge Management, Proceedings*, pp. 340–347.

Alkashri, Z., Alqaryouti, O., Siyam, N. & Shaalan, K. (2020). 'Mining dubai government tweets to analyze citizens' engagement'. , in *Recent Advances in Intelligent Systems and*

Smart Applications, pp. 615–638.

Allam, A. M. N. & Haggag, M. H. (2012). The Question Answering Systems: A Survey. *International Journal of Research and Reviews in Information Sciences (IJRRIS)\t\t\t\t\t*, vol. 2(3) [online].Available at: http://aliallam.net/upload/598575/documents/QA Survey _2012_.pdf.

Almarsoomi, F. A., O'Shea, J. D., Bandar, Z. & Crockett, K. (2013). AWSS: An algorithm for measuring Arabic word semantic similarity. *Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013*, pp. 504–509.

Alobaidi, O. G., Crockett, K. A., O'Shea, J. D. & Jarad, T. M. (2013). Abdullah: An Intelligent Arabic Conversational Tutoring System for Modern Islamic Education. *Lecture Notes in Engineering and Computer Science*, vol. 2 LNECS(October 2016), pp. 762–768.

Alqaryouti, O., Khwileh, H., Farouk, T., Nabhan, A. & Shaalan, K. (2018). Graph-Based Keyword Extraction, pp. 159–172.

Alqaryouti, O., Siyam, N., Monem, A. A. & Shaalan, K. (2019). Aspect-based sentiment analysis using smart government review data. *Applied Computing and Informatics*. The Authors, (xxxx).

Alqaryouti, O., Siyam, N. & Shaalan, K. (2018). A Sentiment Analysis Lexical Resource and Dataset for Government Smart Apps Domain. *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*. Springer International Publishing, pp. 230–240.

Altabba, M., Al-Zaraee, A. & Shukairy, M. A. (2010). An Arabic morphological analyzer and part-of-speech tagger. *A Thesis Presented to the Faculty of Informatics Engineering, Arab International University, Damascus, Syria*, (July 2010) [online].Available at: http://qutuf.com/Files/Report_1.06.pdf.

Altantawy, M., Habash, N. & Rambow, O. (2011). Fast yet rich morphological analysis. *9th International Workshop on Finite State Methods and Natural Language Processing*, pp. 116–124 [online].Available at: http://dl.acm.org/citation.cfm?id=2206310.2206326.

Androutsopoulou, A., Karacapilidis, N., Loukis, E. & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly*. Elsevier, vol. 36(2), pp. 358–367.

ArabianBusiness. (2001). *Dubai's e-government portal launched* [online]. [Accessed 1 June 2020]. Available at: https://www.arabianbusiness.com/dubai-s-e-government-portal-launched-141307.html.

Areed, S., Alqaryouti, O., Siyam, B. & Shaalan, K. (2020). Aspect-Based Sentiment Analysis for Arabic Government Reviews. *Studies in Computational Intelligence*, vol. 874(January), pp. 143–162.

Areed, S., Salloum, S. A. & Shaalan, K. (2020). The Role of Knowledge Management Processes for Enhancing and Supporting Innovative Organizations : A Systematic Review, pp. 143–161.

Azevedo, A. & Santos, M. F. (2008). KDD , SEMMA AND CRISP-DM : A PARALLEL OVERVIEW Ana Azevedo and M . F . Santos. *IADIS European Conference Data Mining*, pp. 182–185 [online].Available at:

http://recipp.ipp.pt/handle/10400.22/136%0Ahttp://recipp.ipp.pt/bitstream/10400.22/136/3/KDD-CRISP-SEMMA.pdf.

Batacharia, B., Levy, D., Catizone, R., Krotov, A. & Wilks, Y. (1999). Machine Conversations. *Machine Conversations*, (December).

Bekkers, V. (2003). E-government and the emergence of virtual organizations in the public sector. *Information Polity*, vol. 8(3–4), pp. 89–101.

Bizimungu, J. (2018). *Babyl's chatbot to enhance digital healthcare platform*. *New Times* [online]. [Accessed 1 July 2020]. Available at: https://www.newtimes.co.rw/section/read/227369.

Boudlal, A., Lakhouaja, A., Mazroui, A., Meziane, A., Ould Abdallahi Ould Bebah, M. & Shoul, M. (2010). Alkhalil Morpho SYS1: A Morphosyntactic Analysis System for Arabic Texts. *international Arab Conference on information technology*, (June), pp. 1–6 [online].Available at: http://www.itpapers.info/acit10/Papers/f653.pdf.

Bradeško, L. & Mladenić, D. (2012). A Survey of Chatbot Systems through a Loebner Prize Competition. *Proceedings of Slovenian Language Technologies Society Eighth Conference of Language Technologies*, vol. 2, pp. 34–37 [online].Available at: https://www.researchgate.net/publication/235664166.

Brandtzaeg, P. B. & Følstad, A. (2019). Chatbots: User changing needs and motivations. *Interactions*, vol. 25(5), pp. 38–43.

Cassell, J. (2000). Embodied Conversational Interface Agents. *Communications of the ACM*, vol. 43(4), pp. 70–78.

Charalabidis, Y. & Metaxiotis, K. (2009). Ontology-based management of e-government knowledge. *Social and Political Implications of Data Mining: Knowledge Management in E-Government*, pp. 221–234.

Colby, K. M. (1975). Artificial Paranoia: A Computer Simulation of Paranoid Process. *Pergamon Press, New York*, vol. 2(1), pp. 1–25.

Cortez, M. B. (2018). *Universities Deploy Chatbots to Aid Students in the Admissions Process and Beyond*. *EdTech* [online]. [Accessed 1 July 2020]. Available at: https://edtechmagazine.com/higher/article/2018/03/universities-deploy-chatbots-aid-students-admissions-process-and-beyond.

Crockett, K., O'Shea, J. D. & Bandar, Z. (2011). Goal Orientated Conversational Agents : Applications to Benefit Society Goal Orientated Conversational Agents : Applications to Benefit Society. *Technologies and Applications*, (June), pp. 16–25.

Demberg, V. & Moore, J. D. (2006). Information Presentation in Spoken Dialogue Systems. *EACL 2006 - 11th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, pp. 65–72.

Dias, J., Kamdi, D., Gharat, N. & Chudhari, P. (2019). Chatbot for Government Examination using AI, vol. 09(4), pp. 58–62.

Douglas, B. L. (1995). CYC: A Large-Scale Investment in Knowledge Infrastructure. *Communications of the ACM*, vol. 38(11), pp. 33–38.

DOUGLAS, T. (2018). *LA Police Making Inroads with 'Chip' the Chatbot*. *Techwire* [online]. [Accessed 1 July 2020]. Available at: https://www.techwire.net/news/la-police-making-inroads-with-chip-the-chatbot.html#:~:text=LA Police Making Inroads with ‘Chip’ the Chatbot,to potential police department recruits.

EGLAND, E. (2019). *Transforming law enforcement with the cloud and AI*. *Microsoft Industry Blogs* [online]. [Accessed 1 July 2020]. Available at: https://cloudblogs.microsoft.com/industry-blog/government/2019/07/31/transforming-law-enforcement-with-the-cloud-and-ai/.

Ehrenbrink, P., Osman, S. & Möller, S. (2017). Google now is for the extraverted, cortana for the introverted: Investigating the influence of personality on IPA preference. *ACM International Conference Proceeding Series*, pp. 257–265.

Ezzeldin, A. M. (2012). A SURVEY OF ARABIC QUESTION ANSWERING: CHALLENGES, TASKS, APPROACHES, TOOLS, AND FUTURE TRENDS, (December), pp. 409–414.

Fellbaum, C. (2020). *Global WordNet Association* [online]. [Accessed 1 July 2020]. Available at: http://globalwordnet.org/resources/arabic-wordnet/.

Feng, J., Zhou, Y. & Martin, T. (2008). Sentence Similarity based on Relevance. *Proceedings of the 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU ’08)*, pp. 832–839 [online].Available at: http://www.gimac.uma.es/ipmu08/proceedings/papers/110-FengZhouMartin.pdf.

Fitzpatrick, K. K., Darcy, A. & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully

Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, vol. 4(2), p. e19.

Fryer, L. & Carpenter, R. (2006). Emerging Techonologies. Bots as language Learning Tools. *Language Learning & Technology*, vol. 10(3), pp. 8–14 [online].Available at: http://llt.msu.edu/pdf/emerging.pdf.

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*, vol. 24(2), pp. 266–290.

Gulf News. (2020). *UAE ranked regional best in UN's 2020 Online Services Index* [online]. [Accessed 1 July 2020]. Available at: https://gulfnews.com/uae/government/uae-ranked-regional-best-in-uns-2020-online-services-index-1.1594408511524.

GulfNews. (2015). *New Dubai meter to gauge happiness* [online]. [Accessed 1 June 2020]. Available at: https://gulfnews.com/uae/government/new-dubai-meter-to-gauge-happiness-1.1496207.

Gyawali, B. (2011). Answering Factoid Questions via Ontologies : A Natural Language Generation Approach. *Department of Intelligent Computer Systems Faculty of Information and Communication Technology University of Malta*, vol. 11(1), pp. 147–173.

Hijjawi, M., Bandar, Z. & Crockett, K. (2016). The Enhanced Arabchat: An Arabic Conversational Agent. *International Journal of Advanced Computer Science and Applications*, vol. 7(2), pp. 332–347.

Hijjawi, M., Bandar, Z., Crockett, K. & Mclean, D. (2014). ArabChat: An Arabic

Conversational Agent. *6th International Conference on Computer Science and Information Technology (CSIT). IEEE, 2014*, pp. 227–237.

Hijjawi, M. H. H. (2011). ArabChat : an Arabic Conversational Agent, (November).

Hijjawi, M., Qattous, H. & Alsheiksalem, O. (2015). Mobile Arabchat: An Arabic Mobile-Based Conversational Agent. *International Journal of Advanced Computer Science and Applications*, vol. 6(10), pp. 111–119.

Hliaoutakis, A., Varelas, G., Voutsakis, E., Petrakis, E. G. M. & Milios, E. (2006). Information Retrieval by Semantic Similarity. *International journal on semantic Web and information systems (IJSWIS)*, pp. 55–73.

Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*. Taylor & Francis, vol. 37(1), pp. 81–88.

Ide, N., Erjavec, T. & Tufis, D. (2002). Sense Discrimination with Parallel Corpora. *Proceedings of the ACL-02 workshop on Word sense disambiguation: recent successes and future directions-Volume 8. Association for Computational Linguistics*, (July), pp. 54–60.

Inaba, M., Iwata, N., Toriumi, F., Hirayama, T., Enokibori, Y., Takahashi, K. & Mase, K. (2014). Constructing a non-task-oriented dialogue agent using statistical response method and gamification. *ICAART 2014 - Proceedings of the 6th International Conference on Agents and Artificial Intelligence*, vol. 1, pp. 14–21.

Jain, V. & Singh, M. (2013). Ontology Development and Query Retrieval using Protégé Tool. *International Journal of Intelligent Systems and Applications*, vol. 5(9), pp. 67–75.

JEONG, B. (2005). *Machine-Learning based Semantic Similarity Measures to Assist*

Discovery and Reuse of Data Exchange XML Schema.

El Kholy, A. & Habash, N. (2010). Techniques for Arabic Morphological Detokenization and Orthographic Denormalization. *the Language Resources and Evaluation Conference (LREC) Workshop on Semitic Languages*, pp. 45–51.

Landauer, T. K., Foltz, P. W. & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, vol. 25(2–3), pp. 259–284.

Larkey, L. S., Ballesteros, L. & Connell, M. E. (2007). Light Stemming for Arabic Information Retrieval. *Arabic Computational Morphology*, pp. 221–243.

Lester, J., Branting, K. & Mott, B. (2004). Conversational Agents. *Practical Handbook of Internet Computing*, pp. 220–240.

Li, X., Szpakowicz, S. & Matwin, S. (1995). A WordNet-based algorithm for word sense disambiguation. *International Joint Conference on Artificial Intelligence*, vol. 14, pp. 1368–1374 [online].Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.47.6728&rep=rep1&type=pdf.

Li, Y., McLean, D., Bandar, Z. A., O’Shea, J. D. & Crockett, K. (2006). Sentence Similarity Based on Semantic Nets and Corpus Statistics. *IEEE Transactions on Knowledge and Data Engineering*, vol. 18(8), pp. 1138–1150.

Liu, B., Xu, Z., Sun, C., Wang, B., Wang, X., Wong, D. F. & Zhang, M. (2018). Content-Oriented User Modeling for Personalized Response Ranking in Chatbots. *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 26(1), pp. 122–133.

Liu, X. Y., Zhou, Y. M. & Zheng, R. S. (2007). Measuring semantic similarity in wordnet. *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, ICMLC 2007*, vol. 6(August), pp. 3431–3435.

Malandrakis, N., Potamianos, A., Iosif, E. & Narayanan, S. (2013). Distributional Semantic Models for Affective Text Analysis. *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21(11), pp. 2379–2392.

Mauldin, M. L. (1994). ChatterBots, tinyMuds, and the turing test entering the loebner prize competition. *Proceedings of the National Conference on Artificial Intelligence*, vol. 1, pp. 16–21.

McHugh, M. L. (2012). Lessons in biostatistics interrater reliability : the kappa statistic. *Biochemica Medica*, vol. 22(3), pp. 276–282 [online].Available at: https://hrcak.srce.hr/89395.

Mohasses, M. (2019). How AI-Chatbots Can Make Dubai Smarter? *Proceedings - 2019 Amity International Conference on Artificial Intelligence, AICAI 2019*. IEEE, pp. 439–446.

Nakov, P., Màrquez, L., Moschitti, A. & Mubarak, H. (2019). *Arabic community question answering*. *Natural Language Engineering*.

Narayanan, S. & Harabagiu, S. (2004). Question Answering based on Semantic Structures. *In COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, pp. 693–701.

Noordt, C. van & Misuraca, G. (2019). New Wine in Old Bottles: Chatbots in Government: Exploring the Transformative Impact of Chatbots in Public Service Delivery. *International*

Conference on Electronic Participation. Springer International Publishing.

Noy, N. F. & McGuinness, D. L. (2001). A Guide to Creating Your First Ontology. *Biomedical Informatics Reseach*, pp. 7–25 [online].Available at: http://bmir.stanford.edu/file_asset/index.php/108/BMIR-2001-0880.pdf.

O'Shea, J., Bandar, Z. & Crockett, K. (2011). 'Systems engineering and conversational agents'. , in *Intelligent Systems Reference Library*, pp. 0–35.

O'Shea, J., Bandar, Z., Crockett, K. & McLean, D. (2008). A Comparative Study of Two Short Text Semantic Similarity Measures. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4953 LNAI, pp. 172–181.

O'Shea, K., Bandar, Z. & Crockett, K. (2009). Towards a new generation of conversational agents based on sentence similarity. *Lecture Notes in Electrical Engineering*, vol. 39 LNEE, pp. 505–514.

O'Shea, K., Bandar, Z. & Crockett, K. (2010). A Conversational Agent Framework using Semantic Analysis. *International Journal of Intelligent Computing Research*, vol. 1(1), pp. 10–19.

O'Shea, K., Crockett, K., Bandar, Z. & O'Shea, J. (2012). An Approach to Conversational Agent Design Using Semantic Sentence Similarity. *Applied Intelligence*, vol. 37, pp. 558–568.

Parahoo, S. K. & Ayyagari, M. (2019). Innovation Ecosystem. *Global Encyclopedia of Public Administration, Public Policy, and Governance*, pp. 1–6.

Preece, J., Rogers, Y. & Sharp, H. (2006). Interaction Design : beyond human - computer interaction. *Interactive Computation: The New Paradigm*, pp. 227–254.

Ray, S. K. & Shaalan, K. (2016). A Review of the State of the Art in Hindi Question Answering Systems, vol. 740(2), pp. 380–392.

Razmerita, L., Nabeth, T., Angehrn, A. & Roda, C. (2004). Inca : an Intelligent Cognitive Agent-Based Framework for Adaptive and Interactive Learning. *Conference on Cognition and Exploratory Learning in Digital Age IADIS CELDA 2004*, (May 2014) [online].Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.4507&rep=rep1&type=pdf.

Reddy, T. (2017). *How chatbots can help reduce customer service costs by 30%*. *IBM* [online]. [Accessed 1 July 2020]. Available at: How chatbots can help reduce customer service costs by 30%25.

RYDING, K. C. (2005). *A Reference Grammar of Modern Standard Arabic*. *Journal of Petrology*.

Saygin, A. P., Cicekli, I. & Akman, V. (2000). Turing Test: 50 Years Later, pp. 23–78.

Schwarzer, M., Düver, J., Ploch, D. & Lommatzsch, A. (2016). An interactive e-government question answering system. *CEUR Workshop Proceedings*, vol. 1670(September), pp. 74–82.

Shawar, B. A. & Atwell, E. (2002). A Comparison Between Alice and Elizabeth Chatbot Systems, p. 23.

Shawar, B. A. & Atwell, E. (2004a). Accessing an Information System by Chatting. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3136, pp. 407–412.

Shawar, B. A. & Atwell, E. (2004b). An Arabic Chatbot Giving Answers from the Qur'an. *Proceedings of TALN04: XI Conference sur le Traitement Automatique des Langues Naturelles Vol. 2*, pp. 197–202 [online].Available at: http://www.comp.leeds.ac.uk/eric/abushawar04taln.pdf.

Shawar, B. A. & Atwell, E. (2009). Arabic question-answering via instance based learning from an FAQ corpus. *Proceedings of the CL2009 International Conference on Corpus Linguistics*, vol. 386(January).

Shawar, B. & Atwell, E. (2003). Using dialogue corpora to train a chatbot. *Proceedings of the Corpus Linguistics 2003*, (October 2014), pp. 681–690.

Silvervarg, A. & Jönsson, A. (2011). Subjective and Objective Evaluation of Conversational Agents in Learning Environments for Young Teenagers. *7th IJCAI Workshop on Knowledge and …*, (November 2014).

Siyam, N., Alqaryouti, O. & Abdallah, S. (2020). Mining government tweets to identify and predict citizens engagement. *Technology in Society*. Elsevier Ltd, vol. 60(November), p. 101211.

Smart London. (2018). *Making our first chat bot*. *Greater London Authority* [online]. [Accessed 1 July 2020]. Available at: https://www.london.gov.uk/city-hall-blog/Making-our-first-chat-bot.

Taghva, K., Elkhoury, R. & Coombs, J. (2005). Arabic stemming without a root dictionary. *International Conference on Information Technology: Coding and Computing, ITCC*, vol. 1, pp. 152–157.

Takala, P., Malo, P., Sinha, A. & Ahlgren, O. (2010). Gold-standard for Topic-specific Sentiment Analysis of Economic Texts. *Proceedings of the Language Resources and Evaluation Conference*, pp. 2152–2157 [online].Available at: http://www.lrec-conf.org/proceedings/lrec2014/pdf/1021_Paper.pdf.

Torres, L., Pina, V. & Royo, S. (2005). E-government and the transformation of public administrations in EU countries: Beyond NPM or just a second wave of reforms? *Online Information Review*, vol. 29(5), pp. 531–553.

UAE Vice President, P. M. and R. of D. H. H. S. M. bin R. A. M. (2013). *Mohammed announces the launch of m-Government* [online]. [Accessed 1 June 2020]. Available at: https://sheikhmohammed.ae/en-us/Pages/NewsDetails.aspx?nid=3154.

USCIS. (2018). *Meet Emma*. *U.S. Citizenship and Immigration Services* [online]. [Accessed 1 July 2020]. Available at: https://www.uscis.gov/tools/meet-emma-our-virtual-assistant.

Valle, K. Del. (2010). The Credibility of Embodied Conversational Agents (ECAs), pp. 1–9.

Walker, M. A., Litman, D. J., Kamm, C. A. & Abella, A. (1998). Evaluating spoken dialogue agents with PARADISE: Two case studies. *Computer Speech and Language*, vol. 12(4), pp. 317–347.

Walker, M., Aberdeen, J., Boland, J., Bratt, E., Garofolo, J., Hirschman, L., Le, A., Lee, S., Narayanan, S., Papineni, K., Pellom, B., Polifroni, J., Potamianos, A., Prabhu, P., Rudnicky, A., Sanders, G., Seneff, S., Stallard, D. & Whittaker, S. (2001). DARPA Communicator Dialog Travel Planning Systems: The June 2000 Data Collection. *EUROSPEECH 2001 - SCANDINAVIA - 7th European Conference on Speech Communication and Technology*, (June), pp. 1371–1374.

Wallace, R. S. (2009). El programa Artificial Linguistic Internet Computer Entity (A.L.I.C.E.). *Parsing the Turing Test*, pp. 181–210 [online].Available at: https://freeshell.de/~chali/programowanie/Anatomy_of_ALICE.pdf.

Wei, C., Yu, Z. & Fong, S. (2018). How to build a Chatbot: Chatbot framework and its capabilities. *ACM International Conference Proceeding Series*, (5), pp. 369–373.

Weizenbaum, J. (1966). ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine. *Communications of the ACM*, vol. 26(1), pp. 23–28.

WHO. (2020). *Depression*. *World Health Organization* [online]. [Accessed 1 July 2020]. Available at: https://www.who.int/news-room/fact-sheets/detail/depression.

Yassine Benajiba, Paolo Rosso, Lahsen Abouenour, O. T. & Karim Bouzoubaa, and L. B. (2014). *Question Answering*. *The Oxford Handbook of Computational Linguistics*.

ZEMČÍK, M. T. (2019). A Brief History of Chatbots. *DEStech Transactions on Computer Science and Engineering*, (aicae), pp. 14–18.

Zouaghi, A., Merhbene, L. & Zrigui, M. (2011). Word sense disambiguation for arabic language using the variants of the lesk algorithm. *Proceedings of the 2011 International Conference on Artificial Intelligence, ICAI 2011*, vol. 2(May 2014), pp. 561–567.

APPENDIX

Sample of Evaluation Questions
ما هي الوثائق المطلوبة لخدمة اعادة طباعة ترخيص مهني
ما هي الأوراق المطلوبة لخدمة إضافة تغيير مسمى المهنيي
ما هي المستندات اللازمة إلغاء ترخيص العمل الجزئي
ما هي خدمة إلغاء ترخيص المهني
ما هي قنوات تقديم الخدمة تجديد تراخيص المنشأة الصحية
من أين يمكن التقديم على خدمة تجديد ترخيص مهني
ما هي ارقام التواصل تجديد ترخيص مهني
كيف يمكن الاستفسار عن خدمة تجديد رخصة مهنيين
ما رسوم خدمة تحميل صور عن إمارة دبي
كم رسوم خدمة طلب تقرير عن المعاملات الجمركية
كم تكلف خدمة طلب مانيفست التصدير
اذكر خدمات هيئة الصحة بدبي
اذكر قنوات الاتصال طلب إصدار تجديد بطاقة الصحة المهنية
اذكر طرق الاتصال بخدمة طلب إصدار شهادة تصدير شهادة صحية لتصدير الأغذية
اين يمكن استخدم الوثيقة تسوية فروقات تخليص البضائع
اين يمكن تقديم معاملة خدمة تصديق الشهادات الطبية
متى يتم العمل في خدمة تصفح المكتبة الالكترونية
ما هي ساعات العمل في تعيين إخطارات الموقع الإلكتروني
كيف يتم خدمة تغيير الموعد
ما هو الرابط الالكتروني الوصول لخدمة تغيير شريك المنشأة
ما هي الوثائق المطلوبة لخدمة تغيير موقع المنشأة
ما هي الأوراق المطلوبة لخدمة تغيير وكيل الخدمة
ما هي المستندات اللازمة تقييد الوكالة التجارية
ما هي خدمة دليل تصديق الشهادات
ما هي قنوات تقديم الخدمة شهادات عدم ممانعة لأعمال توصيل المباني بشبكات الكهرباء المياه الصرف الصحي وتصريف مياه الأمطار الاتصالات والخدمة التلفزيونية
من أين يمكن التقديم على خدمة طلب إعادة تأمين
ما هي ارقام التواصل طلب تحديد جلسة
كيف يمكن الاستفسار عن خدمة طلب تصوير أوراق الدعوى
ما رسوم خدمة طلب استقبال تذكير بانتهاء صلاحية البطاقة الصحية
كم رسوم خدمة طلب اعتماد ترخيص نشاط عقاري
كم تكلف خدمة طلب اعتماد تغيير استشاري مقاول قبل صدور رخصة البناء
اذكر خدمات دائرة السياحة والتسويق التجاري
اذكر قنوات الاتصال طلب التصديق على إعتماد توقيع
اذكر طرق الاتصال بخدمة طلب التصديق على إقرار تنازل عل قضية
اين يمكن استخدم الوثيقة طلب التصديق على إقرار تنازل عن بلاغ
اين يمكن تقديم معاملة خدمة طلب التصديق على توكيل تأسيس منشأة